

**T.C.
ARTVİN ÇORUH ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
MATEMATİK VE FEN BİLİMLERİ EĞİTİMİ ANABİLİM DALI**

**TÜRKİYE İSTATİSTİKİ BÖLGE BİRİMLERİ
SINIFLANDIRMASINA GÖRE PISA FEN BAŞARI SKORLARININ
MAKİNE ÖĞRENME YÖNTEMLERİ İLE KARŞILAŞTIRILMASI VE
İNCELENMESİ**

YÜKSEK LİSANS TEZİ

Büşra YILDIZ

**Danışman
Doç. Dr. Cafer Mert YEŞİLKANAT**

ARTVİN-2025

TEZ BEYANNAMESİ

Artvin Çoruh Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğine göre hazırlamış olduğum “Türkiye İstatistiki Bölge Birimleri Sınıflandırmasına Göre Pısa Fen Başarı Skorlarının Makine Öğrenme Yöntemleri İle Karşılaştırılması Ve İncelenmesi” adlı tezin tamamen kendi çalışmam olduğunu, her alıntıya kaynak gösterdiğimi ve çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı taahhüt ederim. Aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi de beyan ederim. Tezimin kâğıt ve elektronik kopyalarının Artvin Çoruh Üniversitesi Lisansüstü Eğitim Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım.

Lisansüstü Eğitim-Öğretim yönetmeliğinin ilgili maddeleri uyarınca gereğinin yapılmasını arz ederim.

- ☐ Tezimin tamamı her yerden erişime açılabilir.
- ☐ Tezim sadece Artvin Çoruh Üniversitesi yerleşkelerinden erişime açılabilir.
- ☐ Tezimin ... ay süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin tamamı her yerden erişime açılabilir.

...../...../2025

Büşra YILDIZ

İmza

T.C.
ARTVİN ÇORUH ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

TEZ KABUL TUTANAĞI

Doç. Dr. Cafer Mert YEŞİLKANAT danışmanlığında, **Büşra YILDIZ** tarafından hazırlanan bu çalışma **10/01/2025** tarihinde aşağıdaki jüri tarafından **Matematik ve Fen Bilimleri Eğitimi** Anabilim Dalı'nda **oybirliği** ile Yüksek Lisans tezi olarak kabul edilmiştir.

Başkan : Doç. Dr. Ümmü Gülsüm DURUKAN

İmza:.....

Jüri Üyesi : Prof. Dr. Sibel AÇIŞLI ÇELİK

İmza:.....

Jüri Üyesi : Doç. Dr. Cafer Mert YEŞİLKANAT

İmza:.....

Yukarıdaki imzalar adı geçen öğretim üyelerine aittir.

.../.../....

Prof. Dr. Mustafa Çağatay KORKMAZ

Enstitü Müdürü

İÇİNDEKİLER

TEZ BEYANNAMESİ	II
TEZ KABUL TUTANAĞI	III
İÇİNDEKİLER.....	IV
TABLO LİSTESİ.....	VIII
ŞEKİL LİSTESİ	XI
KISALTMALAR.....	XVII
ÖZET	XVIII
SUMMARY	XIX
ÖNSÖZ.....	XX

BİRİNCİ BÖLÜM

GİRİŞ

1.1	Araştırmanın Amacı	4
1.2	Araştırmanın Önemi	4
1.3	Araştırma Problemi	5
1.4	Araştırmanın Özgünlüğü	7

İKİNCİ BÖLÜM

KURAMSAL VE KAVRAMSAL ÇERÇEVE

2.1	PISA 2015 Üzerine Yapılan Çalışmalar.....	8
2.1.1	Dünyadaki Çalışmalar	8
2.1.2	Türkiye’deki Çalışmalar.....	11
2.2	Eğitim Çalışmalarında Makine Öğrenmesi	17
2.2.1	Dünyadaki Çalışmalar	17

ÜÇÜNCÜ BÖLÜM

MATERYAL VE YÖNTEMLER

3.1	PISA Uygulaması	21
3.2	Araştırmanın Örneklemi ve Evreni	23
3.2.1	PISA 2015 Uygulamasında Türkiye Örneklem Yapısı	23
3.3	Çalışmanın Metodolojik İşlem Adımları.....	27
3.4	Makine Öğrenme Algoritmaları	29
3.4.1	Aşırı Gradyan Artırma (Extreme Gradient Boosting, XGBoost).....	29
3.4.2	Destek Vektör Regresyonu (Support VectorRegression, SVR).....	32

3.4.3	Bayes Düzenlemeli Yapay Sinir Ağları (Bayesian Regulation Artificial Neural Network, BRNN).....	33
3.4.4	Rastgele Orman (Random Forest, RF)	35
3.5	Model Değerlendirme Metrikleri	36
3.6	Değişkenlerin Önem Parametrelerinin Tespiti: SHAP Yöntemi.....	37
3.7	Değişkenlerin Benzerliklerinin Tespiti: Jaccard İndeksi.....	39
3.8	Yazılım/Donanım Kaynakları ve Programlama	40

DÖRDÜNCÜ BÖLÜM

BULGULAR VE TARTIŞMA

4.1	TR1 Bölgesi PISA-2015 Bulguları.....	42
4.1.1	TR1 Bölgesi Hakkında Temel Bilgiler.....	42
4.1.2	TR1 Bölgesinin Veri Yapısı	42
4.1.3	TR1 Bölgesinde Makine Öğrenmesi	43
4.1.4	TR1 Bölgesi İçin Değişkenlerin Önem Seviyesi.....	50
4.2	TR2 Bölgesi PISA-2015 Bulguları.....	54
4.2.1	TR2 Bölgesi Hakkında Temel Bilgiler.....	54
4.2.2	TR2 Bölgesinin Veri Yapısı	54
4.2.3	TR2 Bölgesinde Makine Öğrenme Modeli	55
4.2.4	TR2 Bölgesi İçin Değişkenlerin Önem Seviyeleri	62
4.3	TR3 Bölgesi PISA-2015 Bulguları.....	66
4.3.1	TR3 Bölgesi Hakkında Temel Bilgiler.....	66
4.3.2	TR3 Bölgesinin Veri Yapısı	66
4.3.3	TR3 Bölgesinde Makine Öğrenme Modeli	67
4.3.4	TR3 Bölgesi İçin Değişkenlerin Önem Seviyeleri	74
4.4	TR4 Bölgesi PISA-2015 Bulguları.....	78
4.4.1	TR4 Bölgesi Hakkında Temel Bilgiler.....	78
4.4.2	TR4 Bölgesinin Veri Yapısı	78
4.4.3	TR4 Bölgesinde Makine Öğrenme Modeli	79
4.4.4	TR4 Bölgesi İçin Değişkenlerin Önem Seviyeleri	86
4.5	TR5 Bölgesi PISA-2015 Bulguları.....	90
4.5.1	TR5 Bölgesi Hakkında Temel Bilgiler.....	90
4.5.2	TR5 Bölgesinin Veri Yapısı	90
4.5.3	TR5 Bölgesinde Makine Öğrenme Modeli	91

4.5.4	TR5 Bölgesi İçin Değişkenlerin Önem Seviyeleri	98
4.6	TR6 Bölgesi PISA-2015 Bulguları.....	102
4.6.1	TR6 Bölgesi Hakkında Temel Bilgiler.....	102
4.6.2	TR6 Bölgesinin Veri Yapısı	102
4.6.3	TR6 Bölgesinde Makine Öğrenme Modeli	103
4.6.4	TR6 Bölgesi İçin Değişkenlerin Önem Seviyeleri	110
4.7	TR7 Bölgesi PISA-2015 Bulguları.....	114
4.7.1	TR7 Bölgesi Hakkında Temel Bilgiler.....	114
4.7.2	TR7 Bölgesinin Veri Yapısı	114
4.7.3	TR7 Bölgesinde Makine Öğrenme Modeli	115
4.7.4	TR7 Bölgesi İçin Değişkenlerin Önem Seviyeleri	121
4.8	TR8 Bölgesi PISA-2015 Bulguları.....	125
4.8.1	TR8 Bölgesi Hakkında Temel Bilgiler.....	125
4.8.2	TR8 Bölgesinin Veri Yapısı	125
4.8.3	TR8 Bölgesinde Makine Öğrenme Modeli	126
4.8.4	TR8 Bölgesi İçin Değişkenlerin Önem Seviyeleri	133
4.9	TR9 Bölgesi PISA-2015 Bulguları.....	137
4.9.1	TR9 Bölgesi Hakkında Temel Bilgiler.....	137
4.9.2	TR9 Bölgesinin Veri Yapısı	137
4.9.3	TR9 Bölgesinde Makine Öğrenme Modeli	138
4.9.4	TR9 Bölgesi İçin Değişkenlerin Önem Seviyeleri	145
4.10	TRA Bölgesi PISA-2015 Bulguları.....	149
4.10.1	TRA Bölgesi Hakkında Temel Bilgiler.....	149
4.10.2	TRA Bölgesinin Veri Yapısı	149
4.10.3	TRA Bölgesinde Makine Öğrenme Modeli	150
4.10.4	TRA Bölgesi İçin Değişkenlerin Önem Seviyeleri	157
4.11	TRB Bölgesi PISA-2015 Bulguları.....	161
4.11.1	TRB Bölgesi Hakkında Temel Bilgiler	161
4.11.2	TRB Bölgesinin Veri Yapısı	161
4.11.3	TRB Bölgesinde Makine Öğrenme Modeli.....	162
4.11.4	TRB Bölgesi İçin Değişkenlerin Önem Seviyeleri	169
4.12	TRC Bölgesi PISA-2015 Bulguları.....	173
4.12.1	TRC Bölgesi Hakkında Temel Bilgiler	173

4.12.2	TRC Bölgesinin Veri Yapısı	173
4.12.3	TRC Bölgesinde Makine Öğrenme Modeli.....	174
4.12.4	TRC Bölgesi İçin Değişkenlerin Önem Seviyeleri	181
4.13	Önem Parametresi Yüksek Değişkenlerin Karşılaştırılması	185

BEŞİNCİ BÖLÜM

SONUÇ VE ÖNERİLER

5.1	Makine Öğrenmesi Algoritmalarının Performansı.....	190
5.2	Önem Seviyesi Yüksek Değişkenler ve Bölgesel Etkiler	191
5.3	Bölgeler Arası Farklılıklar ve Değişkenlerin Etkisi	193
5.4	Öneriler.....	193
KAYNAKLAR.....		197

EKLER

EK-1:	Algoritmalarda hiperparametre ayarları yapılan özelliklerin anlamları	208
EK-2:	Algoritmalarda kullanılan 99 değişkenin anlamları.....	209
EK-3:	TR1 Bölgesi ve TR3 Bölgesi 1. Grup Benzerlik Seviyesi Yüksek Değişkenler. ...	212
EK-4:	TR1 Bölgesi ve TR4 Bölgesi 1. Grup Benzerlik Seviyesi Yüksek Değişkenler. ...	213
EK-5:	TR1 Bölgesi ve TR6 Bölgesi 1. Grup Benzerlik Seviyesi Yüksek Değişkenler	214
EK-6:	TR2 Bölgesi ve TR6 Bölgesi 1. Grup Benzerlik Seviyesi Yüksek Değişkenler	215
EK-7:	TR1 Bölgesi ve TR6 Bölgesi 2. Grup Benzerlik Seviyesi Orta Seviyedeki Değişkenler.....	216
EK-8	TR7 Bölgesi ve TR8 Bölgesi 2. Grup Benzerlik Seviyesi Orta Seviyedeki Değişkenler.....	217
EK-9:	TR6 Bölgesi ve TR8 Bölgesi 3. Grup Benzerlik Seviyesi Yüksek Seviyedeki Değişkenler.....	218
EK-10	TR1 Bölgesi ve TR8 Bölgesi 3. Grup Benzerlik Seviyesi Yüksek Seviyedeki Değişkenler.....	219
ÖZGEÇMİŞ.....		220

TABLO LİSTESİ

Sayfa No

Tablo 1: PISA uygulamalarındaki temel ve ağırlıklı branşlar (kalın olarak gösterilmiştir)	22
Tablo 2: İBBS'ye göre PISA 2015 Türkiye örneklemindeki öğrenci sayıları ve yüzdeleri	24
Tablo 3: PISA 2015 Türkiye örnekleminin okul türlerine göre dağılımı	25
Tablo 4: Yıllara Göre Fen Okuryazarlığı Ortalama Puanları	27
Tablo 5: TR1 (İstanbul) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri (Bu tablo için ayrıntılar EK-1 de sunulmuştur)	44
Tablo 6: TR1 (İstanbul) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri	45
Tablo 7: TR1 (İstanbul) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin ortalama SHAP değerlerine göre önem seviyelerinin sıralanması.	51
Tablo 8: TR2 (Batı Marmara) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri	56
Tablo 9: TR2 (Batı Marmara) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri	57
Tablo 10: TR2 (Batı Marmara) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.	63
Tablo 11: TR3(Ege) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri	68
Tablo 12: TR3 (Ege) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri	69
Tablo 13: TR3 (Ege) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.	75
Tablo 14: TR4 (Doğu Marmara) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans	

metrikleri ve optimize edilmiş Hiperparametre değerleri (Bu tablo için ayrıntılar EK-1 de sunulmuştur).....	80
Tablo 15: TR4 (Doğu Marmara) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri	81
Tablo 16: TR4 (Doğu Marmara) verileri için oluşturulan En yüksek performanslı algoritmaya (RF) göre değişkenlerin önem seviyelerinin sıralanması.....	87
Tablo 17: TR5 (Batı Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri	92
Tablo 18: TR5 (Batı Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri	93
Tablo 19: TR5 (Batı Anadolu) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.....	99
Tablo 20: TR6 (Akdeniz) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri	104
Tablo 21: TR6 (Akdeniz) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri.....	105
Tablo 22: TR6 (Akdeniz) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.....	111
Tablo 23: TR7 (Orta Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri	116
Tablo 24: TR7 (Orta Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri	117
Tablo 25: TR7 (Orta Anadolu) verileri için oluşturulan En yüksek performanslı algoritmaya (SVR) göre değişkenlerin önem seviyelerinin sıralanması.....	122
Tablo 26: TR8 (Batı Karadeniz) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri	127
Tablo 27: TR8 (Batı Karadeniz) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri	128

Tablo 28: TR8 (Batı Karadeniz) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.....	134
Tablo 29: TR9 (Doğu Karadeniz) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri	139
Tablo 30: TR9 (Doğu Karadeniz) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri	140
Tablo 31: TR9 (Doğu Karadeniz) verileri için oluşturulan En yüksek performanslı algoritmaya (SVR) göre değişkenlerin önem seviyelerinin sıralanması.	146
Tablo 32: TRA (Kuzey Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri	151
Tablo 33: TRA (Kuzey Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri	152
Tablo 34: TRA (Kuzeydoğu Anadolu) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.....	158
Tablo 35: TRB (Ortadoğu Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri	163
Tablo 36: TRB (Ortadoğu Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri	164
Tablo 37: TRB (Ortadoğu Anadolu) verileri için oluşturulan En yüksek performanslı algoritmaya (SVR) göre değişkenlerin önem seviyelerinin sıralanması.	170
Tablo 38: TRC (Güneydoğu Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri	175
Tablo 39: TRC (Güneydoğu Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri	176
Tablo 40: TRC (Güneydoğu Anadolu) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.....	182

ŞEKİL LİSTESİ

Sayfa No

Şekil 1: İstatistikî Bölge Birimleri Sınıflaması, Düzey 1 haritası (MEB, 2016)	23
Şekil 2: 2015 PISA'ya katılan öğrencilerin sınıf düzeylerine göre dağılımı (%).....	25
Şekil 3: 2015 PISA'ya katılan öğrencilerin cinsiyet dağılımı	26
Şekil 4: Türkiye'nin TR1 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir).....	43
Şekil 5: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)	48
Şekil 6: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.....	49
Şekil 7: TR1 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.....	52
Şekil 8: TR1 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.....	53
Şekil 9: Türkiye'nin TR2 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir).....	55
Şekil 10: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)	60
Şekil 11: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.....	61
Şekil 12: TR2 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.....	64
Şekil 13: TR2 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.....	65

Şekil 14: Türkiye'nin TR3 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenini olasılık yoğunluklarını göstermektedir)	67
Şekil 15: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)	72
Şekil 16: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları	73
Şekil 17: TR3 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması	76
Şekil 18: TR3 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı	77
Şekil 19: Türkiye'nin TR4 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenini olasılık yoğunluklarını göstermektedir)	79
Şekil 20: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)	84
Şekil 21: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları	85
Şekil 22: TR4 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması	88
Şekil 23: TR4 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı	89
Şekil 24: Türkiye'nin TR5 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenini olasılık yoğunluklarını göstermektedir)	91
Şekil 25: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama	

diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim).....	96
Şekil 26: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.....	97
Şekil 27: TR5 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.....	100
Şekil 28: TR5 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.....	101
Şekil 29: Türkiye'nin TR6 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)	103
Şekil 30: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim).....	108
Şekil 31: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.....	109
Şekil 32: TR6 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.....	112
Şekil 33: TR6 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.....	113
Şekil 34: Türkiye'nin TR7 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)	115
Şekil 35: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim).....	120
Şekil 36: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.....	121

Şekil 37: TR7 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.....	123
Şekil 38: TR7 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.....	124
Şekil 39: Türkiye'nin TR8 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenini olasılık yoğunluklarını göstermektedir)	126
Şekil 40: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)	131
Şekil 41: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.....	132
Şekil 42: TR8 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.....	135
Şekil 43: TR8 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.....	136
Şekil 44: Türkiye'nin TR8 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenini olasılık yoğunluklarını göstermektedir)	138
Şekil 45: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)	143
Şekil 46: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.....	144
Şekil 47: TR9 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.....	147
Şekil 48: TR9 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.....	148

Şekil 49: Türkiye'nin TRA bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenini olasılık yoğunluklarını göstermektedir)	150
Şekil 50: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)	155
Şekil 51: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları	156
Şekil 52: TRA bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması	159
Şekil 53: TRA bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı	160
Şekil 54: Türkiye'nin TRB bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenini olasılık yoğunluklarını göstermektedir)	162
Şekil 55: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)	167
Şekil 56: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları	168
Şekil 57: TRB bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması	171
Şekil 58: TRB bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı	172
Şekil 59: Türkiye'nin TRC bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenini olasılık yoğunluklarını göstermektedir)	174
Şekil 60: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama	

diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)	179
Şekil 61: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.....	180
Şekil 62: TRC bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.....	183
Şekil 63: TRC bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.....	184
Şekil 64: Jaccard indeksi değerlerine karşılık gelen bezer değişken sayılarının gruplandırılması.....	186
Şekil 65: Bölgeler arasındaki jaccard benzerlik matrisi	189

KISALTMALAR

BİT	:	Bilgi İletişim Teknolojileri
BRNN	:	Bayesçi Düzenlemeli Yapay Sinir Ağları
MEB	:	Millî Eğitim Bakanlığı
OECD	:	Ekonomik Kalkınma ve İş birliği Örgütü
PISA	:	Programme for International Student Assessment
RF	:	Random Forest (Rasgele Orman)
SHAP	:	SHapley Additive exPlanations
SVR	:	Support Vector Regression (Destek Vektör Regresyon)
XGBoost	:	Extreme Gradient Boosting
YSA	:	Yapay Sinir Ağları

ÖZET

TÜRKİYE İSTATİSTİKİ BÖLGE BİRİMLERİ SINIFLANDIRMASINA GÖRE PISA FEN BAŞARI SKORLARININ MAKİNE ÖĞRENME YÖNTEMLERİ İLE KARŞILAŞTIRILMASI VE İNCELENMESİ

Bu çalışma, Türkiye'nin istatistiki bölgelerine göre PISA 2015 fen bilimleri başarı skorlarının farklı makine öğrenmesi yöntemleri kullanılarak tahmin edilmesini, başarıya etki eden değişkenlerin analiz edilmesini ve bölgeler arasındaki başarı farklılıklarının incelenmesini amaçlanmıştır. Araştırma kapsamında Türkiye'nin 12 istatistiki bölgesi ele alınmış ve her bölgenin fen bilimleri başarı skorlarının tahmin edilmesi için Rastgele Orman (Random Forest), Destek Vektör Regresyonu (Support Vector Regression, SVR), Bayes Düzenlemeli Yapay Sinir Ağları (BRNN) ve Aşırı Gradyan Artırma (XGBoost) gibi dört farklı makine öğrenmesi algoritması kullanılmıştır. Modellerin performansı, test verileri üzerindeki tahmin doğruluğu esas alınarak karşılaştırılmıştır. Ayrıca, SHAP (SHapley Additive exPlanations) yöntemi ile değişkenlerin önem sıralamaları belirlenmiş ve bölgeler arası farklılıklar Jaccard benzerlik indeksi kullanılarak analiz edilmiştir. Çalışmanın bulguları, XGBoost ve SVR modelinin, diğer algoritmalara kıyasla test verilerinde en yüksek tahmin performansını gösterdiğini ortaya koymuştur. Buna karşın BRNN ve Random Forest modelleri ise orta düzeyde tahmin doğruluğu sağlamıştır. Bu bulgular, makine öğrenmesi algoritmalarının bölgesel eğitim verilerinin analizinde etkin bir araç olarak kullanılabileceğini göstermektedir. SHAP yöntemi ile her bölge için ayrı ayrı gerçekleştirilen değişken önem sıralamalarına göre, fen bilimleri başarı skorlarını en güçlü etkileyen değişkenlerin genel olarak ANXTEST (Kaygı Testi Skoru), ST013Q01TA (Evde bulunan kitap sayısı), ISCEDO (Ailedeki ebeveynlerin en yüksek eğitim seviyesi), BSMJ (Bilime yönelik meslek hedefi) ve ST092Q02TA (Evdeki eğitim materyalleri) olduğu tespit edilmiştir. Bölgeler arasındaki benzerlik analizlerinde, Türkiye'nin istatistiki bölgeleri arasında orta düzeyde benzerlik gözlenmiştir. Ancak TR1 (İstanbul) ve TR6 (Akdeniz) gibi bölgeler arasında fen bilimleri başarı skorlarında belirgin farklılıklar ortaya çıkmıştır. Ayrıca TR2 (Batı Marmara) ve TR5 (Batı Anadolu) bölgelerinde, değişkenlerin önem sıralamalarında farklılıklar tespit edilmiştir.

Anahtar Kelimeler: PISA 2015, Makine Öğrenmesi, Fen Bilimleri Başarı Skoru, SHAP Yöntemi, Jaccard İndeksi, Bölgesel Eğitim Analizi.

SUMMARY

COMPARISON AND ANALYSIS OF TÜRKİYE'S PISA 2015 SCIENCE ACHIEVEMENT SCORES BASED ON THE STATISTICAL REGIONAL UNITS CLASSIFICATION USING MACHINE LEARNING METHODS

This study aims to predict the PISA 2015 science achievement scores of Türkiye's statistical regions using various machine learning methods, analyze the variables affecting achievement, and examine the performance disparities between regions. The research considers Türkiye's 12 statistical regions, and four machine learning algorithms Random Forest (RF), Support Vector Regression (SVR), Bayesian Regularized Neural Networks (BRNN), and Extreme Gradient Boosting (XGBoost) were employed to predict the science achievement scores for each region. The performance of these models was compared based on their prediction accuracy on test data. Additionally, the importance rankings of the variables were determined using the SHAP (SHapley Additive exPlanations) method, while regional differences were analyzed using the Jaccard similarity index. The findings of the study revealed that the XGBoost and SVR models exhibited the highest prediction performance on the test data compared to other algorithms, while the BRNN and Random Forest models achieved moderate prediction accuracy. These results demonstrate that machine learning algorithms can serve as effective tools for analyzing regional educational data. According to the variable importance rankings performed separately for each region using the SHAP method, the most influential variables on science achievement scores were generally identified as ANXTEST (Test Anxiety Score), ST013Q01TA (Number of Books at Home), ISCEDO (Highest Parental Education Level), BSMJ (Science-Related Career Aspiration), and ST092Q02TA (Educational Resources at Home). The similarity analysis among regions indicated a moderate level of similarity across Türkiye's statistical regions. However, significant disparities in science achievement scores were observed between regions such as TR1 (Istanbul) and TR6 (Mediterranean). Furthermore, differences in the importance rankings of variables were detected in regions like TR2 (Western Marmara) and TR5 (Western Anatolia).

Keywords : PISA 2015, Machine Learning, Science Achievement Scores, SHAP Method, Jaccard Index, Regional Educational Analysis

ÖNSÖZ

“Makine Öğrenme Yönetimleri kullanarak 2015 PISA Türkiye İstatistik bölgelerinde Fen başarı skorlarını etkiyenden değişkenler” konusunda yapılan bu çalışma; Artvin Çoruh Üniversitesi Lisansüstü Eğitim Enstitüsü Matematik ve Fen Bilimleri Eğitimi Anabilim Dalında yüksek lisans tezi olarak hazırlanmıştır.

Araştırma sürecinde çeşitli yönlendirmelerde bulunan, bilgi ve tecrübesi ile karşılaştığım zorlukları aşmamda yardımcı olan değerli danışmanım Doç. Dr. Cafer Mert YEŞİLKANAT’a, literatür taramasında yardımcı olan eşim Uzm. Aygün YILDIZ’a teşekkürlerimi sunarım.

Yüksek lisans ders döneminde bilimsel bakış açısı ve eleştirel düşünce yapısı kazandıran değerli hocam Prof. Dr. Sibel AÇIŞLI ÇELİK'e teşekkürlerimi sunarım.

Eğitim hayatımın boyunca desteğini esirgemeyen annem Emel TAŞ’a, babam Mehmet TAŞ’a ve hayatımın her alanında motivasyon kaynağım olan ablam Pelin DERELİ’ye teşekkürlerimi sunarım.

Araştırmanın bilimsel ve teknik açıdan uygulayıcılara faydalı olmasını dilerim.

Büşra YILDIZ

Artvin – 2024

BİRİNCİ BÖLÜM

GİRİŞ

Bilgi akışının ve bilgi artış ivmesinin giderek hız kazandığı 21. yüzyılda, bilginin toplanması ve bu bilgi birikiminin aktarılması, yeni teknolojilerin üretilmesine yol açmaktadır. Bilim ve teknolojinin bu denli hızlı gelişmesi de öğrencilerin bu hıza adapte olmasını zorlaştırmakta, aynı zamanda başarı oranlarının korunması veya ilerlemesi açısından yeni teknolojik yöntemlerin doğmasını mecbur kılmaktadır. Bu nedenlerden, eğitim sistemi dinamiklerinin sürekli olarak güncellenmesi, yeni teknoloji ve yöntemlere göre senkronize edilmesi, bu iyileştirmelerin de öğrenci ve eğitimcilerce benimsenmesi, eğitim kalitesinin artırılması noktasında oldukça önemlidir (Güllüpınar, Kuzu, Dursun, Kurt ve Gültekin, 2013).

Öğrencilerin bireysel performansları ve akademik başarıları gibi değişkenleri mercek altına alan birçok bilimsel çalışma bulunmaktadır (Güzel, 2018; Berberoğlu, Çalışkan ve Karşlı, 2019). Bu çalışmaların amaçları genel olarak öğrenci başarılarını tespit etmek ve eksiklikleri saptayıp ulusal ve uluslararası düzeyde eğitim yöntemlerinin iyileştirmesine ve geliştirilmesine odaklanmaktadır. Bunlara ek olarak gerçekleştirilen çalışmalar incelendiğinde, öğrencilerin öğrenmesi gereken bilgileri kullanarak günlük hayatlarına yansıtabilmeleri ve bu problemler karşısında analitik yeteneklerini ortaya koyup sorunlara hızlı çözümler üretebilmelerinin amaçlandığı görülmüştür (Aydın ve Yılmaz, 2010).

Ülkelerin bu amaçlar doğrultusunda sistemleştirdikleri eğitim felsefelerinin, öğretim programlarının, yakın ve uzak hedeflerinin birbirlerinden farklı olmasının ortak sebepleri bulunmaktadır. Benzer şekilde aynı kıtada bulunan komşu ülkelerde dahi öğrenci ve öğretmen profillerinin farklı eğitim sistemlerine tabi tutulmasının nedeni, ülkedeki akademik başarıyı ideal düzeye getirmek ve istenen donanıma sahip bireyler yetiştirmeyi hedeflemektir (Özerbaş ve Safi, 2022). Öğrenci yetiştirme konusunda ülkeler her ne kadar bağımsız görünüyor olsalar da öğrenci başarıları ve eğitim çıktılarının objektif bir biçimde ölçülmesi konusunda uluslararası hakem niteliğinde olan ve bilimsel olarak geçerlik ve güvenilirliğini ispatlamış organizasyonlar mevcuttur. Bu hedefleri gerçekleştirmek ve uluslararası eğitim düzeylerini, öğrenci çıktılarını ve akademik başarı oranlarını ölçüp karşılaştırmak üzere çalışmalar yürüten programların başında Uluslararası Öğrenci Değerlendirme Programı (PISA) gelmektedir (Eraslan, 2009).

OECD (Ekonomik İş Birliği ve Kalkınma Teşkilatı) tarafından üç yılda bir yürütülen PISA projesi; okuma becerileri, matematik ve fen bilimleri konularında temel becerileri esas alarak eğitim-öğretim süreçlerinin sonunda öğrencilerin topluma gerçek anlamda kazandırılması noktasında bu öğrenim girdi ve çıktılarını hangi seviyede edindiklerini değerlendirmektedir (MEB, 2022). PISA, üç yıllık dönemler halinde matematik, fen bilimleri ve okuma konuları ile bağlantılı bir şekilde bilgi ve becerileri kullanma, uygulama ve her periyotta farklı becerilerin saptanmaları konusunda çalışmalar yürütmekte olup her PISA döngüsünde temel alanlardan biri “ağırlıklı branş” olarak kararlaştırılmaktadır (Demir ve Yıldırım, 2016). PISA 2015 araştırmasında da ağırlıklı alanın fen bilimleri olarak belirlenmiş olması, ulaşılan skorların fen bilimleri alanı özelinde içerisindeki öğrenci performans ve başarılarına ilişkin detaylı bir veri deposu olarak dikkat çekmektedir (Kahraman ve Çelik, 2017). Bu doğrultuda; fen başarı skorlarının ölçülmesi, değerlendirilmesi ve karşılaştırmalı olarak çeşitli yöntemler bağlamında analiz edilmesi açısından PISA 2015 bulguları, önemli bir veri kaynağı olarak ortaya çıkmaktadır.

PISA sonuçlarına göre öğrencilerin başarı skorlarına etki eden değişkenler literatür bakımından incelendiğinde, kullanılan yöntemler ve başarı skorları üzerindeki etkileri analiz edilen faktörlerin çeşitlilik gösterdiği göze çarpmaktadır. Nitekim PISA 2009 verileri ile yapılan bazı çalışmalarda; öğrencilerin, fen ve matematik alanlarına ait bilgilerin transferini etkileyen faktörler incelendiğinde, analiz sonuçlarına göre fen başarısının diğer branşlardaki başarı skorlarının anlamlı bir tanımlayıcısı olduğu ve aynı zamanda sosyo-ekonomik ve bölge bazındaki değişkenlerin diğer beceriler üzerinde önemli bir etkisi olduğu bulgularına ulaşılmıştır (Shelley ve Yıldırım, 2013).

Öğrenci başarılarını ve performans düzeylerini ölçen yöntemler olduğu gibi bu verileri karşılaştırıp analiz eden birçok yöntem de bulunmaktadır. Bunlar arasında araştırmacılara en etkili analiz sonuçları sunan güncel yöntemlerden biri de makine öğrenme yöntemleridir. Makine öğrenmesi, toplanan verileri temelinde istatistiksel ve matematiksel metotlar kullanarak değişkenler arasındaki ilişki ağlarını tespit eden ve bu tespitler doğrultusunda tahmin ve çıkarım işlemlerinde bir sistemler bütünüdür. Makine öğrenmesi metotlarının ana hedeflerinden biri mümkün olabilecek en düşük hata düzeylerinde ve gerçek sonuca en yakın seviyede tahminler gerçekleştirmektir (Soydan, Gacar ve Akay, 2020). Makine öğrenmesi, farklı istatistiksel metotları da bir arada irdelemeyi sağlayan disiplinler arası bir çalışma sahasıdır. Özellikle eğitim verilerini analiz etme ve modelleme bakımından önemli bir

alıřma alanını kapsamaktadır (Nizam ve Akin, 2014). Bu nedenle makine ğrenme yntemleri aracılıęı ile oluřturulan alıřmalar, halihazırdaki eęitim tekniklerinin sistematize olmuř problemlerini analiz etmelerinde arařtırmacılara byk kolaylıklar saęlamıřtır (Medhat, Hassan ve Korashy, 2014).

Makine ğrenme yntemleri kullanılarak yapılan ve ęrencilerin bařarı skorları arasındaki belirleyicileri analiz eden alıřmaların en popler temalarından biri, farklı lkelerin PISA sonularını karřılařtırmalı olarak analiz etmektir. Popler dięer bir tema ise ęrencilerin akademik bařarılarının tahmin edilmesinde ne ıkan saptamaktır. Bu alıřmalardaki temel amacın, hangi ęrenci zelliklerinin test puanlarıyla, hangi okul zelliklerinin okul dzeyinde llen okul katma deęeriyle iliřkili olduęunu ortaya koymak olduęu gzlemlenmiřtir (Erdoęan ve Tařtan, 2024). Bu alıřmalardaki bir bařka ama da incelenen yardımcı deęiřkenler ve bařarı puanları arasındaki iliřkileri keřfetmek ve sonuları etkileyen faktrlerin arasındaki etkileřimleri modellemektir. Bu alıřmaların biroęunda; rastgele orman (Random Forest, RF), torbalama (bagging, bootstrap aggregating) ve arttırma algortimaları (boosting algorithm), Destek vektr makineleri (Support Vector Machines, SVM), regresyon ve sınıflandırma aęaları (CART) gibi makine ğrenme yntemlerinin kullanıldıęı grlmektedir (Uęuz, řahin ve Yılmaz, 2021; Demircioęlu, 2023; řehribanoęlu, 2024; Aksu ve Gzeller, 2016; Masci, Johnes ve Agasisti, 2018; Acıslı-Celik ve Yesilkanat, 2023).

Makine ğrenmesi, bir bilgisayarın verilen veriler temelinde eřitli problemlere olasılıksal zmler retmesini, bu zmleri sınıflandırarak gruplandırmasını ve bu gruplamalar zerinden anlamlı ıkarımlar yapmasını saęlayan bir algoritmik sretir. Bu yntem, verilerden ğrenme, sınıflandırma, kmeleme ve tahmin yapma gibi temel bileřenleri ierir (Onan ve Korukoęlu, 2016). Bu kapsamda deęerlendirildięinde makine ğrenmesi yntemlerinin avantajlarından faydalanmak amacıyla, son yıllarda PISA gibi byk veri setlerine sahip sınavlara katılan ęrencilerin bařarı performanslarını inceleyen alıřmaların sayısında nemli bir artıř gzlemlenmektedir (Anil, 2009; Oktay Gzeller ve řeker, 2016; Saarela, Yener, Zaki ve Krkkinen, 2015; Acar ve ęretmen, 2012; Aksu ve Doęan, 2018; Koyuncu ve Gelbal, 2020; Hu, Peng ve Ma, 2022; Lezhnina ve Kismihk, 2022; Immekus, Jeong ve Yoo, 2022; Bezek Gre, řevgin ve Kayri, 2023; Kurt, Yksel ve Dinergk, 2023; Kasap, Doęan ve Koak, 2021; Gllpınar ve dięerleri, 2013; ztrk, 2018; Acıslı-Celik ve Yesilkanat, 2023; Hacıoęlu, 2024; Gzeller ve řeker, 2016)

1.1 Araştırmanın Amacı

Bu araştırmanın temel amacı; PISA'nın 2015'te gerçekleştirdiği sınavdaki veriler temel alınarak, Türkiye istatistiki bölgelerine göre fen bilimleri öğrenci başarı skorlarının farklı makine öğrenmesi yöntemleri ile hangi oranlarda tahmin edilebilir olduğunun belirlenmesi ve incelenmesidir. Bu kapsamda araştırma çerçevesinde ele alınan temel amaçlar aşağıda görünen biçimde sıralanabilir:

1. Türkiye'nin her bir istatistiki bölgesi için, öğrencilerin PISA-2015 fen başarı skorlarını farklı makine öğrenme algoritmalarını [Rasgele Orman (Random Forest, RF), Destek Vektör Regresyon (Support Vector Regression, SVR), Bayes Düzenlemeli Yapay Sinir Ağları (Bayesian regularized artificial neural network, BRNN), Aşırı Gradyan Arttırma (eXtreme Gradient Boosting, XGBoost)] kullanarak tahmin etmek ve sonuçlarını karşılaştırmak,

2. İstatistiki bölgelere ait fen başarısını etkileyen faktörleri incelemek ve bu faktörlerin bölgeler arasındaki önem sıralamalarını analiz etmek ve karşılaştırmak,

3. Farklı makine öğrenmesi algoritmalarının tahmin performansını değerlendirerek her bir bölge için en doğru tahmini yapan modeli tespit etmek,

4. Elde edilen sonuçlar doğrultusunda eğitim politikalarına ve bölgesel eğitim stratejilerine yönelik öneriler geliştirmek,

5. Fen bilimleri alanında öğrenci başarılarını artırmaya yönelik bölgesel farklılıkları ortaya koyarak, bu farklılıkların giderilmesi noktasında tespitlerde bulunmak.

1.2 Araştırmanın Önemi

Uluslararası düzeyde çalışmalarda bulunan en büyük eğitim araştırmalarından biri, OECD tarafından düzenlenen PISA'dır. PISA; öğrencilerin matematik, fen ve okuma becerileri alanlarındaki bilgi ve becerilerinin değerlendirildiği, üç yılda bir gerçekleştirilen ve OECD'ye üye devletler ile diğer katılımcı devletlerdeki 15 yaş grubu öğrencilerin günümüz toplumlarında yerlerini alabilmeleri için gereken temel akademik kazanımlara ne ölçüde sahip olduklarını değerlendiren bir projedir (OECD, 2014). Bu projenin, ülkelerin eğitim sistemlerinin gidişatları hakkında tahminlerde bulunma özelliği taşıdığını düşünmeye yol açan nedenlerden biri de bu denli geniş kapsamlı bir amaca hizmet etmesidir. Özellikle Türkiye gibi gelişme sürecini devam ettiren ülkelerdeki eğitim stratejilerinin yol haritalarının çizilmesi konusunda, PISA analizlerinden faydalanmanın anlamlı sonuçlar

doğurabileceği düşünülmektedir (Aydın ve Yılmaz, 2010b; Kılıç, Çene ve Demir, 2012; Gür, Çelik ve Özoglu, 2012).

Ülkelerin eğitim stratejilerinin güncel kalabilmesi ve modern çağa ayak uydurabilmesi ülkelerin refah seviyelerinin giderek artış göstermesi hedefinde günümüzde bir koşul haline gelmiştir. Bu koşulların baskın parametrelerinden birinin fen bilimleri başarı skorları olması, PISA gibi projelerin ortaya koyduğu analizlerden açıkça görülmektedir. Bu analizler paralelinde tespit edilen başarılı eğitim sistemi stratejilerinin, ülkelerin gelişmişlik seviyelerini de belirtmesi, PISA'daki fen bilimleri başarı skorlarının önemini farklı bir açıdan vurgulamaktadır.

Bu araştırma 2015 PISA sınavlarında Türkiye'nin istatistiki bölgeler bazında elde ettiği sonuçları, makine öğrenmesi gibi güncel ve güçlü yaklaşımlarla analiz ederek öğrencilerin fen bilimleri başarı skorlarını derinlemesine incelemeyi amaçlamaktadır. Bu araştırmadan elde edilecek bulguların Türkiye'nin eğitim politikası stratejilerinin geliştirilmesine önemli katkılar sağlayacağı ve eğitim sistemimizin güçlendirilmesi ile uluslararası düzeyde rekabet edebilirliğinin artırılması yönünde kritik bir adım teşkil edeceği düşünülmektedir. Dolayısı ile bu çalışma, sadece mevcut durumu ortaya koymakla kalmayıp aynı zamanda geleceğe yönelik stratejik planlamalar için de sağlam bir temel oluşturması açısından büyük bir öneme sahiptir. Buna bağlı olarak bu araştırmanın; eğitim-öğretim stratejileri geliştiricilerine, eğitim felsefecilerine, eğitim sosyologlarına ve bilumum eğitim girdisi uyarılama kanallarına daha detaylı bir perspektif yansıtacağı, bu edinimler sayesinde de bahsi geçen eğitim düzenleyicilerinin yeni felsefeler oluşturma konusunda anlamlı ve istatistiki argümanlar ile uygulamalar yapabileceği öngörülmektedir.

1.3 Araştırma Problemi

Teknolojinin ilerlemesiyle birlikte verilerin dijital ortamda analiz edilmesi, muhtelif gerekçelere dayandırılarak oldukça popüler bir duruma gelmiştir. Zira analiz uygulamalarında önemli olan maksimum hızda maksimum verime ulaşmaktır. Karşılaştırmalı analiz süreçlerinde veri analizlerine yardımcı olan en önemli uygulamalardan biri makine öğrenme yöntemleridir (Wang, Fu, He, Hao ve Wu, 2020; Awad ve Khanna, 2015). Teknolojik imkânların yaygınlaşması sürecinde de makine öğrenme yöntemleri; halihazırda bulunan eğitim yöntem ve stratejilerini irdelemek, analiz etmek, yeni ve daha

ideal teknikler geliřtirmek hedefleriyle iřleme sokulan en önemli yaklařımlardan biri olma niteliğine eriřmiřtir.

Birçok alıřmada, arařtırmacıların gerekleřtirdikleri makine ğrenme uygulamaları ile ğrencilerin ğrenme edinimleri ile ilgili daha iyi fikirlere sahip oldukları ve mevcut veri setlerini daha iyi ğrenebildikleri gözlemlenmiřtir. ünkü makine ğrenme yöntemlerinin yalnızca alıřma sahiplerinin uygulamalarına kolaylık sağlamadıėı, aynı zamanda pratik stratejiler ile yararlı bilgileri etkileřime sokarak ğrencileri de olumlu açılardan etkilediėi görölmüřtür(Onan ve Korukoėlu, 2016).

Makine ğrenme algoritmalarının hızlı veriler elde etme, verileri test etme, hipotez oluřturma ve ğrenci performanslarında bireyin geliřimi için tekrar eden bir araç řeklinde yansıma göstermesi gibi özellikleri, eğitim-ğretim konuları ile ilgili mevcut literatür ve verilerin detaylı bir incelemesi sonrasında ortaya çıkmıřtır (Romero ve Ventura, 2007). İncelenen sonuçlara göre, PISA gibi projelerin analiz süreçlerinde alıřılagelmiř yöntem ve stratejilerin hız anlamında verim sağlamamaları, makine ğrenme yöntemlerinin hızlı ve anlamlı veriler ortaya koyması noktasındaki gerekliliėi öne ıkarmıřtır.

Bu bağlamda bu alıřmanın ana arařtırma problemleri ařaėıdaki gibi sıralanabilir:

1. Farklı makine ğrenmesi yöntemleri kullanılarak PISA 2015 sınavına giren ğrencilerin fen okuryazarlıėı başarı puanları Türkiye'nin istatistiki bölgelerine göre ne derece başarılı bir řekilde tahmin edilebilir?

2. Tahmin sonuçları bölgeler bazında karřılařtırıldıėında, fen okuryazarlıėı başarı puanlarında hangi bölgeler arasında anlamlı farklılıklar veya benzerlikler bulunmaktadır?

3. alıřmada incelenen tahmin edici deėiřkenler fen okuryazarlıėı başarı puanlarının tahminini bölgeler bazında nasıl ve hangi oranlarda etkilemektedir?

4. Bölgeler arasında tahmin edici deėiřkenlerin etkileri açısından hangi benzerlikler ve farklılıklar mevcuttur? Hangi deėiřkenler belirli bölgelerde daha etkili veya önemsizdir?

5. Makine ğrenmesi yöntemleri ile elde edilen tahmin sonuçları ve deėiřken etkileri, Türkiye'nin eğitim politikalarının geliřtirilmesi için hangi ıkarımları ve önerileri sunmaktadır?

6. Elde edilen bulgular ışığında, Türkiye'nin istatistiki bölgelerindeki eğitim stratejilerinin geliřtirilmesi ve bölgesel eřitsizliklerin azaltılması için hangi adımlar atılabilir?

1.4 Araştırmanın Özgünlüğü

Bu araştırma; PISA 2015'teki veriler ışığında, Türkiye istatistiki bölgelerine göre fen bilimleri başarı skorlarının makine öğrenme algoritmaları kullanılarak tahmin edildiği, sonuçların karşılaştırıldığı, istatistiki bölgeler bakımından başarı skorlarını etkileyen değişkenlerin önem sıralarının tespit edildiği ilk araştırma olması açısından özgünlük değeri yüksektir. Ayrıca daha önceki bir başka çalışmada (Hacıoğlu, 2024), PISA 2015 Türkiye geneli fen alt alanı için tespit edilmiş olan en önemli 99 değişken, Türkiye istatistiki bölgelerine göre ilk kez analiz edilmiştir. Bu durum araştırmanın özgünlük seviyesini artırmaktadır.



İKİNCİ BÖLÜM

KURAMSAL VE KAVRAMSAL ÇERÇEVE

Makine öğreniminin kullanım alanları, geçmişte nispeten yeni bir döneme dayanmasına rağmen, bilgisayar sistemleri teknolojilerindeki birikimli ve sürekli gelişmelerle paralel olarak hızla genişlemiştir. Bu doğrultuda, günümüzde makine öğrenme algoritmalarının pek çok alanda yaygın bir şekilde kullanıldığı gözlemlenmektedir. Makine öğrenmesi algoritmalarının kullanım alanları arasında eğitim, uzay bilimleri, doğa bilimleri, finans ve tıp gibi disiplinler öne çıkmaktadır. Bu alanlarda, makine öğrenmesi, sorunların çözümüne yenilikçi yaklaşımlar sunmakta ve geleneksel yöntemlerin ötesinde etkili çözümler geliştirilmesine olanak tanımaktadır. Bu tezin kuramsal çerçevesini kapsayan literatür taraması ise başta PISA 2015 ile ilgili fen bilimleri bazındaki makine öğrenmesi çalışmalarını, bu konu bağlamında gerçekleştirilen dünyadaki ve Türkiye'deki çalışmaları, aynı zamanda makine öğrenme algoritmalarının eğitim-öğretim alanındaki kullanım biçimlerini ve bu alanın irdelendiği Türkiye ve dünyadaki bilimsel çalışmaları oluşturmaktadır. Ulaşılan çalışmaların amaçladıkları hedefler ve uygulamaların sonuçları aşağıdaki başlıklarda belirtilmiştir:

2.1 PISA 2015 Üzerine Yapılan Çalışmalar

PISA 2015, öğrencilerin bilgisayar tabanlı simülasyonlar aracılığıyla bilimsel araştırmaları analiz etme ve tasarlama becerilerinin ilk kez değerlendirilmesi gibi yenilikçi özellikleri nedeniyle, literatürdeki araştırmacıların ve eğitim politikası düzenleyicilerinin yoğun ilgisini çekmiştir. Bu nedenle, PISA 2015 ile ilgili olarak dünya genelinde çok sayıda akademik çalışma gerçekleştirilmiştir. PISA 2015 sınavı üzerine yapılan araştırmalar incelendiğinde, sınava katılan öğrencilerin fen başarı skorları, matematik başarı skorları ve okuma becerileri skorlarına etki eden değişkenlerin genellikle geleneksel istatistiksel yöntemlerle analiz edildiği görülmektedir. Aşağıda, bu çalışmalardan hem dünya genelinde hem de Türkiye bağlamında öne çıkan bazı örnekler sunulmaktadır.

2.1.1 Dünyadaki Çalışmalar

Lau ve Lam, (2017), çalışmalarında PISA 2015 sınavında en yüksek başarı gösteren 10 ülkeyi bilimsel performansları ve öğretim uygulamaları açısından kapsamlı bir şekilde analiz etmiştir. Bu ülkeler arasında Singapur, Japonya, Estonya, Taipei, Finlandiya, Makao, Kanada, Hong Kong, Çin ve Kore yer almaktadır. Araştırmacılar, bu ülkelerin fen bilimleri

başarı skorlarını ve öğretim uygulamalarını tanımlayarak karşılaştırmalı bir değerlendirme yapmıştır. PISA 2015'te geliştirilen sorgulamaya dayalı öğretim yapısı, faktör analizi yoluyla iki yeni yapıya dönüştürülmüş ve ardından bu yapılar ile öğretim uygulamaları arasındaki ilişkiler, “Hiyerarşik Doğrusal Modelleme (HLM)” yöntemi kullanılarak incelenmiştir. Çalışma, öğretim uygulamaları üzerindeki kültürel etkileri tartışırken, Batı eğitim felsefesi ile Konfüçyüs kültürü arasındaki etkileşim sonucunda didaktik ve yapılandırmacı pedagojinin nasıl harmanlandığını ortaya koymuştur. Sonuçlar, uyarlamalı öğretim, öğretmen rehberliğinde gerçekleştirilen öğretim ve etkileşimli uygulamaların, tüm ülkelerde başarı skorları ile pozitif bir ilişkiye sahip olduğunu göstermiştir. Bununla birlikte, algılanan geri bildirimlerin tüm ülkelerde negatif bir etki yarattığı tespit edilmiştir. Bu bulgular, kültürel farklılıkların ve öğretim yöntemlerinin başarı üzerindeki etkisini vurgulamak açısından önemli katkılar sunmaktadır.

Hwang, Choi, Bae ve Shin (2018), araştırmalarında, sosyoekonomik statü ile fen bilimleri veya matematik okuryazarlığı arasındaki ilişkilerin öğrenci merkezli öğretim tarafından yumuşatılma durumlarını incelemişlerdir. Uluslararası bir karşılaştırma için çok çeşitli başarı düzeylerini ve eğitimde eşitliği kapsayan on ülke seçilmiştir. PISA 2012 ve PISA 2015'ten öğrenci başarısı, eşitlik ve öğrenci merkezli öğretim verilerinin sıklığına doğrusal bir regresyon analizi uygulanmıştır. Çalışmanın sonunda; öğrenci merkezli öğretimin daha sık sunuldukça, düşük ve yüksek sosyoekonomik statü arasındaki matematik ve fen bilimleri okuryazarlığının farkının genellikle daralmakta olduğu tespit edilmiştir. Çoğu ülkede, öğrenci merkezli eğitim daha sık verildiğinden, öğrencilerin matematik ve fen bilimleri başarı skorlarının tüm sosyoekonomik statülerde düştüğü gözlemlenmiştir.

Meng, Qui ve Boyd-Wilson (2018), çalışmalarında, PISA 2015 baz alınarak bilgi ve iletişim teknolojileri (BİT) katılımının faktör yapısını ve ölçüm değişmezliğini; BİT katılımı ile Çin ve Almanya'daki öğrencilerin fen, matematik ve okuma başarı skorları arasındaki ilişkiyi araştırmışlardır. Regresyon modeli ile yapılan çalışmada; Çin'de BİT'e ilgi ile öğrencilerin başarı skorları arasında önemli ölçüde olumlu bir korelasyon tespit edilirken, Almanya'da önemli ölçüde olumsuz bir korelasyon görülmüştür. Tüm başarı skorları, Çin'de algılanan BİT yeterlilik skorları ile olumsuz ve önemli ölçüde ilişkili olduğu görülmüştür. Almanya'da fen bilimleri ve matematik başarı skorları BİT yeterlilik skorları ile tahmin edilememiştir. Araştırma sonucunda, Çin ve Almanya'da BİT kullanımında algılanan

özerklik ve BİT kullanımında öğrencilerin başarı skorlarının tahmin etmede sosyal ilişki açısından benzer örüntülere rastlanmıştır.

She (2019), PISA 2015 öğrenci anketlerinden seçilen değişkenlere dayanarak öğrencilerin fen başarı skorlarına göre potansiyel gizli kümelerini belirlemeyi amaçlamışlardır. Tayvan'dan 7.973 öğrenciyi içeren çalışmada, Kümeleme Analizi ve Karar Ağaçları Algoritması kullanılmıştır. Sonuçlar, fen bilimleri okuryazarlığı ve epistemolojiye dayalı dört küme belirlemiş ve fen başarı skorlarının yüksek olduğu öğrencilerin epistemik inançlar, başarı motivasyonu, bilim öz yeterliği ve öğretmen desteği gibi faktörlerle pozitif ilişkili olduğunu göstermiştir. Regresyon analizi, bilim hakkında epistemik inançların en güçlü yordayıcı olduğunu, bunu öğrenme süresi ve başarı motivasyonunun izlediğini ortaya koymuştur. Karar ağacı sonuçları ise epistemik inançların, öğrenme süresi ve öz yeterliğin öğrencileri ayırt etmede en önemli değişkenler olduğunu göstermiştir (She, 2019).

Habenicht ve Goldhammer (2020), PISA 2015'e eklenen bilgi ve iletişim teknolojileri (BİT) okuryazarlığının yeni bir boyutu olan "BİT Katılımı"nın yapısal özelliklerini incelemiş ve cinsiyet farklılıkları ile BİT kullanım ölçekleri ile başarı skorları arasındaki ilişkileri araştırmıştır. İsviçre ve Almanya verilerine dayanan çalışmada, Yapısal Eşitlik Modellemesi ile BİT katılımının dört boyutlu yapısı doğrulanmıştır. Bulgular, eğlence amaçlı BİT kullanımının en yüksek varyansı açıkladığını, kız öğrencilerin BİT katılım ölçeklerinde (BİT ilgisi hariç) daha düşük değerlere sahip olduğunu ve sosyal etkileşim alt ölçeği ile okuma başarısı arasında küçük bir negatif korelasyon bulunduğunu göstermiştir. Elde edilen sonuçlar, dört alt ölçeğin yapısal geçerliliğini desteklemiştir.

You, Park ve Delgado (2021), PISA 2015 verilerini kullanarak ABD okullarının fen bilimleri okuryazarlığıyla ilişkili özelliklerini incelemişlerdir. Çok değişkenli üç düzeyli modelleme yöntemiyle yapılan çalışmada, fen bilimleri başarı skorlarındaki toplam varyansın %21'inin okullar arasındaki farklılıklardan kaynaklandığı belirlenmiştir. Öğrenci düzeyinde sınıf, eğlence, motivasyon ve sosyo-ekonomik statünün her iki fen bilimleri okuryazarlığı ölçüsüyle anlamlı ve pozitif ilişkili olduğu görülmüştür. Erkeklerin fen içerik bilgisinde üstün olduğu tespit edilmiştir. Okul düzeyinde ekonomik/sosyal/kültürel statü en önemli faktör olarak öne çıkarken, okul büyüklüğü, öğretim deneyimi ve bilime özgü kaynakların başarıya anlamlı katkıda bulunmadığı belirlenmiştir. Mesleki gelişim katılımı

ve okul iklimi ise yalnızca ortalama ve yüksek performans gösteren öğrencilerle anlamlı ilişki göstermiştir.

Putra ve Agusnita (2021), çalışmalarında biyoloji konularındaki farklı soru formatlarının öğrencilerin fen bilimleri yeterliliklerini nasıl etkilediğini incelemeyi amaçlamışlardır. Araştırma, Endonezya'nın Padang şehrinde gerçekleştirilmiş olup örneklem grubu 44 onuncu sınıf öğrencisinden oluşmaktadır. Veri toplama sürecinde basit rastgele örnekleme yöntemi kullanılmış, sorular ise PISA 2015 verilerinden temin edilmiştir. Araştırma bulgularına göre, öğrencilerin fen bilimleri başarı skorları soru formatlarına göre farklılık göstermiştir. Nesnel sorularda başarı oranı %50,66, bileşik sorularda %35,91 ve deneme formatındaki sorularda %42,72 olarak ölçülmüştür. Genel olarak, soru formatlarına dayalı fen bilimleri okuryazarlık yeterliliği puanı ortalaması %43,9 olarak hesaplanmıştır. Sonuçlar, öğrencilerin nesnel soruları diğer formatlara kıyasla daha kolay cevaplayabildiklerini ortaya koymuştur. Bu durum, fen bilimleri okuryazarlık yeterliliklerinin değerlendirilmesinde soru formatlarının önemli bir rol oynadığını ve nesnel soruların öğrencilerin bilgi ve becerilerini test etmede daha etkin olabileceğini göstermektedir.

2.1.2 Türkiye'deki Çalışmalar

Buldu ve Olgan (2018), çalışmalarında, PISA'nın Ekonomik, Sosyal ve Kültürel İndeksi (ESCS) değişkenleri ile Türkiye'nin PISA 2015 sınavına katılan 15 yaş grubu öğrencilerinin fen bilimleri başarı skorları arasındaki farklılıkları analiz etmeyi amaçlamışlardır. Bu bağlamda, öğrencilerin fen bilimleri başarı skorları, bireysel ve aileye ait değişkenler açısından tek yönlü varyans analizi (ANOVA) yöntemi kullanılarak incelenmiştir. Araştırmanın bulguları, 2-3 ve 4 yaşlarında okul öncesi eğitime başlayan çocukların, PISA fen bilimleri okuryazarlık becerileri açısından 5-6 yaşlarında veya 1 yaşından önce okul öncesi eğitime başlayan çocuklara göre daha yüksek başarı skorlarına sahip olduğunu göstermiştir. Ayrıca, anne ve baba eğitim düzeyindeki artışın, öğrencilerin fen bilimleri okuryazarlık başarı skorları üzerinde anlamlı bir fark yarattığı tespit edilmiştir. Bu sonuçlar, erken yaşta başlayan okul öncesi eğitimin ve ebeveynlerin eğitim düzeyinin, öğrencilerin fen bilimleri okuryazarlığı üzerinde önemli bir etkisi olduğunu ortaya koymakta ve eğitim politikalarının geliştirilmesine yönelik değerli bilgiler sunmaktadır.

Yıldırım ve Karaman (2018), çalışmalarında ailenin yoksulluk-zenginlik düzeyi ile eğitim arasındaki ilişkiyi çeşitli değişkenler açısından incelemeyi hedeflemişlerdir. Ayrıca, PISA 2015 öğrenci anketinde yer alan “ailenin zenginliği (family wealth)” değişkeninin, Türk öğrencilerin fen bilimleri, matematik ve okuma becerileri başarı skorlarını ne ölçüde tahmin edebildiğini belirlemeyi amaçlamışlardır. Araştırmada, veri analizi yöntemi olarak "Basit Doğrusal Regresyon Analizi" kullanılmıştır. Çalışma bulguları, fen bilimleri, matematik ve okuma becerileri başarı skorları arasında pozitif ve anlamlı bir korelasyon olduğunu ortaya koymuştur. Bunun yanı sıra, ailenin zenginliği değişkeninin, öğrencilerin akademik başarılarını anlamlı bir şekilde tahmin ettiği tespit edilmiştir. Bu sonuçlar, aile ekonomik durumunun öğrencilerin akademik performansı üzerinde güçlü bir etkisi olduğunu ve eğitim politikalarının bu faktörü göz önünde bulundurarak şekillendirilmesi gerektiğini göstermiştir.

Demirez (2018), çalışmasında PISA 2015 Türkiye, Singapur ve Almanya verilerini kullanarak bazı değişkenlerin fen bilimleri başarı skorlarına etkisinin bahsi geçen ülkelere göre değişim gösterme durumlarını incelemiştir. Aynı zamanda belirlenen değişkenlerin etkilerinin “düşük”, “orta” ve “yüksek” başarı skorlarına erişen öğrencilerdeki farklılaşma durumları araştırılmıştır. Araştırmada “Kantil Regresyon (QR) Analizi” ve “Sıradan En Küçük Kareler (OLS)” regresyon yöntemi kullanılmıştır. Araştırma sonuçlarında; Türkiye, Singapur ve Almanya’da, belirlenen değişkenlerin bazılarının fen bilimleri başarı skorlarını tahmin edebildiği, bazılarının ise tahmin etmede yetersiz kaldığı tespit edilmiştir.

Üstün, Özdemir, Cansız ve Cansız (2019), çalışmalarında PISA 2015 verilerini kullanarak Türk öğrencilerin fen başarı skorlarını etkileyen öğrenci ve okul düzeyindeki değişkenleri belirlemeyi amaçlamışlardır. Bu doğrultuda, öğrencilerin fen başarı skorlarını tahmin eden değişkenlerden oluşan hiyerarşik bir model geliştirmeyi hedeflemişlerdir. Modelin oluşturulmasında "Hiyerarşik Doğrusal Modelleme (HLM)" analizi kullanılmıştır. Araştırma sonucunda, öğrencilerin fen başarı skorlarını tahmin eden bir model oluşturulmuş ve bu modelde öğrenci düzeyinde 9, okul düzeyinde ise 4 anlamlı değişken tespit edilmiştir. Çalışma bulgularına göre, öğrenci düzeyinde etkili değişkenin "öğrencilerin haftalık fen bilimleri dersi süresi" olduğu, okul düzeyinde ise en güçlü tahmin edici değişkenin "okulun fen bilimlerine özgü kaynakları" olduğu belirlenmiştir. Bu sonuçlar, öğrenci başarılarını artırmak için hem bireysel düzeyde ders sürelerinin optimizasyonuna hem de okulların fen bilimlerine yönelik kaynaklarının iyileştirilmesine odaklanılması gerektiğini göstermiştir.

Berberoğlu, Çalışkan ve Karşlı (2019), çalışmalarında PISA 2015 verilerini kullanarak, okul ortamına ilişkin çeşitli değişkenlerin fen bilimleri başarı skorlarındaki varyansı ne ölçüde açıkladığını incelemişlerdir. Bu amaç doğrultusunda analiz süreçlerini "Hiyerarşik Lineer Regresyon" yöntemi ile gerçekleştirmişlerdir. Analiz sonuçları, fen bilimleri başarı skorları varyansının %48'inin okullar arası farklılıklarla, %52'sinin ise öğrenciler arasındaki farklılıklarla açıklandığını ortaya koymuştur. Çalışmada raporlanan regresyon analizi bulgularının, PISA 2015 raporlarında sunulan sonuçlarla uyumlu olduğu belirtilmiştir. Araştırma sonuçları, eğitim sisteminde yapılması gereken reformların holistik bir yaklaşımla ele alınması ve farklı boyutlar çerçevesinde değerlendirilmesi gerektiğini vurgulamaktadır. Bu durum, eğitim politikalarının geliştirilmesinde bütüncül bir bakış açısının önemini açıkça ortaya koymaktadır.

Batı, Yetişir ve Güneş (2019), çalışmalarında PISA 2015 verilerini kullanarak, "epistemolojik inanç", "motivasyon" ve "öz yeterlik" gibi duyuşsal özelliklerin öğrencilerin fen bilimleri başarı skorlarını ne ölçüde açıkladığını ve bu açıklama düzeyinin farklı başarı performansına sahip ülkeler arasında farklılık gösterip göstermediğini incelemişlerdir. Araştırma kapsamında, PISA'nın tanımladığı üç başarı diliminden her birine rastgele yöntemle ikişer ülke seçilmiş ve bu ülkelerde sınava katılan tüm öğrenciler analiz sürecine dahil edilmiştir. Veri analizleri, PISA verileri için geliştirilmiş olan "IDB Analyzer" programı kullanılarak gerçekleştirilmiş ve değişkenlerin fen bilimleri başarı skorlarını tahmin etme düzeylerini belirlemek için "Doğrusal Regresyon Analizi" uygulanmıştır. Araştırma sonuçlarına göre, tüm başarı dilimlerinde duyuşsal özelliklerin öğrencilerin fen bilimleri başarı skorlarını yaklaşık %30 oranında tahmin ettiği tespit edilmiştir. Bu bulgu ile, öğrencilerin fen bilimlerindeki başarılarının yalnızca bilişsel faktörlerle değil, aynı zamanda duyuşsal faktörlerle de yakından ilişkili olduğunu ortaya çıkartılmıştır.

Yitik (2019), araştırmasında PISA 2015'e katılan öğrencilerin bireysel, ailesel ve okula ait değişkenler ile fen bilimleri başarı skorları arasındaki ilişkiyi incelemeyi amaçlamıştır. Bu doğrultuda, araştırmada ilişkisel tarama yöntemi kullanılmıştır. Çalışmada uygulanan "Hiyerarşik Çoklu Doğrusal Regresyon" analizi sonuçlarına göre; "evde konuşulan dil", "okul öncesi eğitime başlama yaşı", "test kaygısı", "motivasyon", "fen öz yeterliği", "epistemolojik inançlar", "fen etkinlikleri", "fen öğrenmekten zevk alma", "evdeki kültürel varlıklar", "ev eğitim kaynakları", "ekonomik, sosyal ve kültürel durum indeksi", "fen sınıflarındaki disiplin koşulları", "sorgulamaya dayalı fen öğretim ve öğrenim

alıştırmaları”, “öğretmen yönlendirmeli fen öğretimi” ve “öğretmen adaleti” değişkenlerinin fen bilimleri başarı skorları üzerinde anlamlı tahmin ediciler olduğu sonucuna ulaşılmıştır.

Uluğ (2019), araştırmasında PISA 2015’e katılan öğrencilerin fen bilimleri başarı skorları ile okuma becerileri başarı skorları arasındaki farkı açıklamakta kullanılan bazı değişkenlerin etkilerini incelemeyi amaçlamıştır. Araştırmada yöntem olarak ilişkisel tarama modeli tercih edilmiştir. Ulaşılan verilerin analizleri “Kantil Regresyon” tekniği yardımı ile gerçekleştirilmiştir. Araştırmanın bulgularına göre, “haftalık fen ders saati” ve “haftalık okul dışı çalışma süreleri” değişkenlerinin istatistiki açıdan anlamsız olduğu sonuçlarına varılmıştır. “Matematik okuryazarlığı”, “iş birliği ve ekip çalışması”, “öğretmen adilliği”, “bilim etkinlikleri”, “geniş bilim konularına ilgi”, “okula ait olma duygusu”, “öğrenci tutumları motivasyon eğilimleri” ve “sınıf tekrarı” değişkenlerinin anlamlı oldukları saptanmıştır.

Zeybekoğlu (2019), çalışmasında Türk öğrencilerin PISA 2015 öğrenci anketine vermiş oldukları cevaplar ile fen başarı skorlarını açıklayan değişkenleri incelemeyi amaçlamıştır. PISA 2015 sınavında, 72 ülkedeki 29 milyon öğrenciyi içeren bir evreni temsil eden 540 bin öğrenci çalışma kapsamına alınmıştır. Bu çalışmada tarama modellerinden biri olan ilişkisel tarama deseni kullanılmıştır. Türk öğrencilerin fen başarı skorlarını açıklayan değişkenler PISA öğrenci anketinden seçilen maddeler ile belirlenmiştir. Bu araştırmada, verilerin analizini yapmak adına veri madenciliği karar ağaçları algoritmalarından biri olan “Otomatik Ki-Kare Etkileşim Belirleme (CHAID)” yöntemi kullanılmıştır. Analizler sonucunda Türk öğrencilerin fen başarı skorlarını en iyi açıklayan değişkenin “evde bulunan kitap sayısı” olduğu sonucuna varılmıştır.

Gür (2019), çalışmasında PISA 2015 sınavında öğrencilerin fen bilimleri başarı skorlarını etkileyen duyuşsal değişkenlerden oluşan “Yol Modeli”ni test etmeyi ve bu modelin bölgeler üzerindeki değişmezliğini incelemeyi hedeflemiştir. Bu çalışmada yöntem olarak ilişkisel tarama modeli kullanılmıştır. Ayrıca ölçme aracında yer alan ve fen başarı skorları ile ilgili duyuşsal özelliklerin araştırıldığı değişkenlerden; “geniş fen konularına ilgi”, “fen öğrenmekten zevk alma”, “fen öz yeterliği”, “epistemolojik inanç” ve “fen aktiviteleri” değişkenleri yol modeline dahil edilmiştir. Bu değişkenlerin fen bilimleri başarı skorlarını, Türkiye örneklemini ve bölgeler üzerinde anlamlı bir biçimde tahmin etme seviyelerini tespit etmek amacıyla “Çoklu Doğrusal Regresyon Analizi” kullanılmıştır. Fen

bilimleri başarı skorlarını tahmin eden değişkenler arasında Türkiye örnekleme ve bölgeler bazındaki en önemli değişkenin “epistemolojik inanç” olduğu sonucuna varılmıştır.

Ulukan (2019), araştırmasında Türkiye, Almanya, Kanada, İspanya, Fransa ve İtalya'ya ait PISA 2003 ve PISA 2015 verilerini kullanarak, söz konusu ülkelerin gelişmişlik seviyeleri ile öğrencilerin fen bilimleri başarı skorları arasındaki ilişkileri karşılaştırmalı olarak incelemeyi amaçlamıştır. Bu karşılaştırmalarda, öğrencilerin yeteneklerine göre aynı veya farklı sınıf içerisinde sınıflandırılması, öğrencilerin gün boyu okul içinde ve dışında internet kullanım süresi, öğrencilerin ilk bilgisayar edindikleri yaş, öğrencilerin okul değiştirme durumları ve öğrencilerin anaokuluna gitme durumları gibi değişkenler üzerinde durulmuştur. Araştırmadan elde edilen bulgular doğrultusunda, Türkiye'nin bölgeler bazında PISA sonuçları ile gelişmişlik seviyeleri ve PISA sonuçlarının Türkiye ekonomisi üzerindeki etkisinin modellenebileceği yeni çalışmaların yapılması gerektiği yönünde önerilerde bulunulmuştur (Ulukan, 2019).

Yüksel (2019), çalışmasında PISA 2015 sonuçlarını kullanarak Türkiye ve Finlandiya verilerinde okul özellikleri ile öğrencilerin okuma becerisi seviyeleri arasındaki ilişkiyi incelemiştir. Bu kapsamda, Türkiye ve Finlandiya'nın PISA 2015 verilerine göre "okul büyüklüğü", "okulun bulunduğu yerleşim yeri", "öğretmen başına düşen öğrenci sayısı", "kız öğrenci oranı", "öğretmen eksikliği", "eğitim kaynaklarının kalitesi" ve "internete bağlı bilgisayar oranı" gibi değişkenler, ilişkisel tarama modeli ile ele alınmıştır. Çalışmanın sonucunda, tahmin edici değişkenler ile okuma beceri skorları arasındaki ilişkiler analiz edilmiştir. Finlandiya'da, PISA 2015 verilerine göre okuma becerileri, kız öğrenci oranı değişkeninde anlamlı bulunurken; "okul büyüklüğü", "yerleşim yeri", "eğitim kaynaklarının kalitesi", "internete bağlı bilgisayar oranı", "öğretmen başına düşen öğrenci sayısı" ve "öğretmen eksikliği" değişkenlerinde anlamlı bir ilişki tespit edilmemiştir. Türkiye'de ise "eğitim kaynaklarının kalitesi", "internete bağlı bilgisayar oranı", "öğretmen başına düşen öğrenci sayısı", "öğretmen eksikliği" ve kız öğrenci oranı değişkenleri anlamlı bulunmuş; "okul büyüklüğü" ve "yerleşim yeri" değişkenlerinde ise anlamlı bir ilişki saptanmamıştır.

Karasu (2019), çalışmasında ilişkisel araştırma yöntemini kullanarak, PISA 2012 ve PISA 2015 sınavlarında Türkiye'nin "Ortadoğu Anadolu" bölgesinde eğitim gören öğrencilerin fen bilimleri başarı skorlarını etkileyen çeşitli faktörleri incelemiştir. Ayrıca, söz konusu sınavlara katılan öğrenciler üzerinden, bölgedeki öğrencilerin 2012-2015 yılları

arasındaki fen bilimleri başarı skorlarını "cinsiyet", "anne eğitim durumu", "baba eğitim durumu", "evinde kendine ait oda olma durumu", "evinde bilgisayar olma durumu" ve "evdeki kitap sayısı" değişkenleri temelinde karşılaştırmalı olarak analiz etmiştir. Verilerin analizinde, tek yönlü varyans analizini gerçekleştirmek için ANOVA istatistiksel yöntemi kullanılmıştır. Karasu, çalışmanın sonuçlarında, belirtilen değişkenlerin fen bilimleri başarı skorları üzerinde anlamlı bir farklılık yaratıp yaratmadığını belirlemiştir.

Orhan (2020), araştırmasında PISA 2015 verilerini kullanarak, fen ile ilişkili öğrenci ve okul özelliklerinin Türkiye'deki 15 yaşındaki öğrencilerin fen bilimleri başarı skorlarına etkisini incelemeyi amaçlamıştır. Bu doğrultuda, öğrencilerin birleşik fen puanları, fen bilimleri puanları ve fen yeterlilik puanları ayrı ayrı değerlendirilmiştir. Çalışmada veriler, “Hiyerarşik Lineer Modelleme (HLM)” yöntemi ile analiz edilmiştir. Araştırma sonucunda, okul dışında fen bilimleri çalışmaya ayrılan zaman, fen aktiviteleri ve çevresel optimizmin, fen bilimleri başarı skorlarının tüm boyutlarını olumsuz yönde etkilediği belirlenmiştir. Buna karşılık, fen bilimleri öğrenmeye ayrılan zaman, epistemolojik inançlar, çevre farkındalığı ve erişmeyi hedefledikleri mesleki statü değişkenlerinin, fen bilimleri başarı skorlarının tüm boyutlarını olumlu yönde etkilediği tespit edilmiştir. Araçsal güdülenmenin ise öğrencilerin fen bilimleri başarı skorları üzerinde anlamlı bir etkisinin olmadığı gözlemlenmiştir. Bu bulguların yanı sıra, diğer değişkenlerin fen bilimleri başarı skorları üzerindeki tüm olumlu ve olumsuz etkileri de detaylı olarak ortaya konulmuştur.

Zaman ve diğerleri (2022), araştırmalarında fen bilimleri başarı skorlarındaki değişimleri PISA verilerine dayanarak betimsel araştırma yöntemi ile incelemiştir. Çalışma, ikincil verilerin kullanıldığı aşamalı bir araştırma sürecinden oluşmaktadır. Verilerin analizi, “SPSS” programı ile “ANOVA” kullanılarak gerçekleştirilmiştir. Çalışmanın bulguları, fen bilimleri başarı skorları ile Türk öğrencilerin annelerinin eğitim düzeyi, Türk öğrencilerin babalarının eğitim düzeyi, sahip olunan cep telefonu sayısı, evdeki bilgisayar sayısı, evdeki kitap sayısı, evde yardımcı eğitim kaynaklarına erişim, şiir kitabına sahip olma durumu ve dünya klasiklerine sahip olma durumu arasında anlamlı bir ilişki olduğunu göstermiştir.

Hacıoğlu (2024), çalışmasında PISA 2015 ve 2018 verilerini kullanarak, sınavlara katılan Türk öğrencilerin fen bilimleri, matematik ve okuma becerileri başarı skorlarını etkileyen öğrenci, aile ve okul faktörlerini belirlemeyi ve öğrencilerin akademik

performanslarını makine öğrenmesi algoritmaları ile tahmin etmeyi amaçlamıştır. Bu doğrultuda, öğrenci başarı skorlarını tahmin etmeyi hedefleyen makine öğrenmesi algoritmalarını anlamlı biçimde etkileyen değişkenler tespit edilmiştir. Bu bağlamda, Singapur ve Çin gibi ülkelerin PISA performansları ile söz konusu değişkenlerin öğrencilerin akademik başarıları üzerindeki etkileri karşılaştırılmıştır. Araştırmanın yöntemi kapsamında, makine öğrenmesi algoritmalarından “Random Forest (RF)”, “XGBoost”, “Destek Vektör Regresyonu (SVR)” ve “Bayesyen Düzenlemeli Yapay Sinir Ağları (BRNN)” kullanılarak tahmin modelleri oluşturulmuştur. Elde edilen bulgular, mevcut eğitim politikalarının yeniden değerlendirilerek modern anlamda revize edilmesine ihtiyaç duyulduğunu göstermektedir.

2.2 Eğitim Çalışmalarında Makine Öğrenmesi

2.2.1 Dünyadaki Çalışmalar

Golino ve Gomes (2014), çalışmalarında, psikometri ve makine öğrenimi alanlarını entegre ederek lise öğrencilerinin akademik başarısını tahmin etmek için nispeten yeni bir doğrusal olmayan yöntem kullanmışlardır. 14 ila 19 yaşları arasındaki 135 lise öğrencisinden oluşan bir örneklem ile öğrenciler “Tümevarımsal Muhakeme Gelişim Testi (TDRI)”, “Meta Bilişsel Kontrol Testi (TCM)” ve “Brezilya Öğrenme Yaklaşımları Ölçeği (BLAS-Derin Yaklaşım)” adlı üç psikolojik araca yanıt vermişlerdir. Öğrencilerin her teste/ölçeğe verdiği yanıtlar “Rasch Modeli” kullanılarak analiz edilmiştir. Sınıf üyeliğini tahmin etmek için “Rastgele Orman (RF)” adlı doğrusal olmayan bir makine öğrenimi algoritması kullanılmıştır. Çalışma bitimindeki Rastgele Orman sonucunda; eğitim setinde %75 genel doğruluk, %73,69 özgüllük ve %68 duyarlılık ortaya çıkmıştır.

Gabriel, Signolet ve Westwell (2017), araştırmalarında, PISA 2012 veri setini kullanarak psikolojik eğilimler ve matematik okuryazarlığı arasındaki ilişkileri keşfetmek için makine öğrenmesi algoritmalarını kullanmışlardır. Çalışma 15 yaşındaki Avustralyalı öğrenciler üzerinde uygulanmıştır. Elde edilen sonuçlar, matematik öz yeterliliğinin önceliğini yeniden teyit etmiş ve matematik öz yeterliliği, matematik kaygısı ve sosyoekonomik statü arasındaki yeni karmaşık etkileşimleri vurgulamıştır. Tüm analizler “R Programlama Dili” kullanılarak gerçekleştirilmiştir. Çalışmada, eğitim araştırmacılarının çok boyutlu bir veri setinde karmaşık doğrusal olmayan ilişkileri ve yeni etkileşimleri bulmak için veri odaklı modelleme tekniklerini nasıl yararlı bir şekilde kullanabilecekleri gösterilmeye çalışılmıştır.

Soares (2024), çalışmasında PISA 2018 matematik, fen bilimleri ve okuma sınavlarındaki yeterlilikleri etkileyen faktörleri yorumlamak amacıyla makine öğrenmesini geleneksel istatistik teknikleriyle birleştirmeyi denemiştir. Bu çalışmada, tahmin edici unsurlar geniş bir değişken yelpazesini içermekte olup, sosyodemografik özellikleri, öğretim ve öğrenme süreçlerini ve bilişsel olmayan becerileri kapsamaktadır. Çalışmanın yönteminde, makine öğrenmesi algoritmalarından biri olan “Rastgele Orman (Random Forest, RF)” analizi kullanılarak matematik, fen bilimleri ve okuma alanlarındaki yeterlilik için tahmin edicilerin önemi karşılaştırılmıştır. Ardından, en önemli tahmin ediciler ile sonuçlar arasındaki ilişkiyi incelemek amacıyla çok düzeyli lojistik regresyon analizleri uygulanmıştır. Çalışmanın sonucunda, belirlenen üç yeterlilik alanı için en önemli öngörücüler arasında "yıllık hane geliri", "ebeveynlerin en yüksek mesleki statüsü" ve "erken çocukluk eğitimi ve bakımı" yer almış ve bunların matematik, fen bilimleri ve okuma yeterlilikleriyle olumlu ilişkili olduğu tespit edilmiştir. Buna karşılık, "sınıf tekrarı" ve "ek eğitim" gibi faktörlerin bu sonuçlarla olumsuz ilişkili olduğu belirlenmiştir.

Chen ve diğerleri (2021), çalışmalarında PISA anketine dayalı olarak öğrenci, aile ve okul düzeyindeki faktörleri eşzamanlı olarak ele alarak, fen bilimleri başarı skorlarında en yüksek performansı gösteren öğrencilerin en önemli bağlamsal özelliklerini belirlemeyi amaçlamışlardır. Bu kapsamda, 141 bağlamsal değişken, bir makine öğrenmesi algoritması kullanılarak en iyi performans gösteren öğrencilerle olan ilişkilerine göre sıralanmıştır. Ardından, “Destek Vektör Makinesi (SVM)” sınıflandırıcısının sağladığı yüksek doğruluk nedeniyle, en iyi ve en düşük performans gösteren öğrencileri sınıflandırma ve tahmin etmede etkili olan bu özellikler arasından 20 faktörden (değişken) oluşan optimum bir set seçilmiştir. Araştırmanın bulguları, öğretmenlerin öğretim uygulamalarının kalitesi, ebeveynlerin eğitim ve mesleki statüsü, sınıfın disiplin iklimi, öğrenmeye ayrılan zaman ve katılım düzeyi, okulların kitle iletişim araçları tesisleri ve ekipmanları, okuldaki öğretmen sayısı ve öğrencilerin öz yeterliliklerinin, fen bilimleri başarı skorlarında en belirleyici rolleri oynadığını göstermektedir.

Yoo ve Rho (2020), çalışmalarında Koreli öğrencilerin uluslararası karşılaştırmalı araştırmalarda yüksek akademik başarılarına rağmen, öğretmenlerinin iş tatmininin OECD ortalamasının altında kaldığını vurgulamışlardır. Bu nedenle, araştırmalarında OECD'nin “2013 Uluslararası Öğretme ve Öğrenme Anketi (TALIS)” verilerini makine öğrenimi yöntemleriyle analiz etmişlerdir. Makalenin sonuçlarına göre, önceki araştırmalarla uyumlu

olarak "işbirlikçi okul iklimleri" ve "öğretmen öz yeterliği" gibi değişkenleri de içeren 558 tahmin edici arasından 18 tanesi belirlenmiştir. Öğretmen iş memnuniyetine ilişkin yeni tespit edilen değişkenler arasında "öğretmen geri bildirimi", "katılımcı okul iklimleri" ve "mesleki gelişime yönelik algılanan engeller" yer almaktadır.

Hu, Peng ve Ma (2021), çalışmalarında eğitim modellerinin dinamik bir tasarımını kullanarak, yüksek ve düşük performans gösteren öğrencilerin fen bilimleri başarı skorlarını sinerjik olarak etkileyen temel bağlamsal faktörleri belirlemeyi amaçlamışlardır. Araştırmada, 53 ülke ve ekonomiden 79.963 ortaöğretim öğrencisine ait fen bilimleri başarı skorları ile öğrencilerin ve okul müdürlerinin PISA anketlerine verdikleri yanıtları içeren PISA 2015 veri seti kullanılmıştır. “Destek Vektör Makinesi (SVM)” ve “SVM-Tekrarlı Özellik Eleme (SVM-RFE)” yöntemleri sırasıyla uygulanarak; okul, sınıf ve öğrenci düzeylerinde yüksek ve düşük başarı gösterenleri sinerjik olarak ayırt eden toplam 127 bağlamsal faktör arasından 30 temel faktör başarıyla belirlenmiştir. Ayrıca, çalışmanın sonuçları, bağlamsal faktörlerin çok boyutluluğunu dikkate alarak eğitim etkinliğinin dinamik modelinin geçerliliğini destekleyen kanıtlar sunmuştur.

Lezhninan ve Kismihók (2021), çalışmalarında PISA 2015 ve 2018 verilerini kullanarak, Alman öğrencilerin bilgi ve iletişim teknolojilerine (BİT) yönelik tutumlarını matematik ve fen bilimleri başarı skorları ile ilişkili olarak incelemişlerdir. Bu amaçla, makine öğrenimi algoritmalarından yararlanmışlardır. Çalışmada, eksik verilerin tamamlanması ve öğrencilerin matematik ve fen bilimleri alanlarındaki yeterlilik seviyelerinin tahmin edilmesi için “Rastgele Orman (Random Forest, RF)” algoritması uygulanmıştır. Ayrıca, verilerin hiyerarşik yapısı dikkate alınarak, BİT'e yönelik tutumlar ile matematik ve fen bilimleri başarı skorları arasındaki ilişkileri keşfetmek amacıyla “Hiyerarşik Doğrusal Modeller (HLM)” oluşturulmuştur. Araştırmanın sonuçlarına göre, BİT özerkliği değişkeni, Rastgele Orman modellerinde önemli bir değişken olarak belirlenmiş ve bu tutum ile HLM'deki okuryazarlık puanları arasındaki ilişkilerin anlamlı ve olumlu olduğu tespit edilmiştir. Diğer BİT tutumları olan sosyal etkileşimde BİT kullanımı, BİT yeterliliği ve BİT ilgisi için ise ilişkilerin olumsuz olduğu gözlemlenmiştir.

Immekus, Jeong ve Yoo (2022), çalışmalarında öğrencilerin PISA 2015 verilerine dayalı okul çalışmalarıyla ilgili kaygılarını değerlendirmek amacıyla, geleneksel regresyon yaklaşımları da dahil olmak üzere tahmin edici değişken seçimi için “Azaltma Operatörü

(LASSO)”, “Elastik Ağ (ENET)” ve “Rastgele Orman (RF)” yöntemlerini karşılaştırmışlardır. Çalışmanın sonuçları, LASSO ve Enet'in 100 modelleme yinelemesinde RF'den daha iyi performans gösterdiğini, ancak tahmin performansı açısından birbirlerinden anlamlı bir fark olmadığını ortaya koymuştur. Genel olarak LASSO ve ENET, RF'ye kıyasla daha üstün performans sergilemiş ve okul çalışmalarıyla ilgili kaygının belirleyicilerinin başarı hedefleri, aidiyet duygusu ve bilimsel olguyu açıklama güveni gibi kişisel ve çevresel faktörleri içeren etkenler olduğu vurgulanmıştır.

Resino, Constante-Amores, Castro ve Navarro (2024), çalışmalarında PISA 2018'e katılan İspanyol öğrencilerden oluşan bir örneklem kullanarak, okuduğunu anlama alanında yüksek performansla ilişkili okul etkinliği faktörlerini incelemeyi amaçlamışlardır. Veriler, “Rastgele Orman (Random Forest, RF)” algoritması ve ikili çok düzeyli “Lojistik Regresyon” yöntemleri kullanılarak analiz edilmiştir. Temel bulgular arasında, en önemli değişkenlerin "okuma keyfi" ve "metabiliş" adı verilen süreç değişkenleri olduğu vurgulanmıştır. Çalışmanın sonuçları, okuma sevgisini ve metabilişsel kapasiteyi teşvik eden eğitim politikalarının geliştirilmesi yoluyla, İspanyol bağlamında yüksek performanslı öğrencilerin sayısının artırılmasının mümkün olduğunu göstermektedir (Resino, Constante-Amores, Castro ve Navarro, 2024).

ÜÇÜNCÜ BÖLÜM

MATERYAL VE YÖNTEMLER

3.1 PISA Uygulaması

PISA, 2000 yılında uygulanmaya başlayan, beraberinde OECD (Ekonomik Kalkınma ve İş birliği Örgütü) tarafından üç yıl aralıklar ile gerçekleştirilmeye devam edilen bir sınavlar bütünüdür (EARGED, 2010). PISA, dünyadaki ekonomik güç yelpazesinin çok büyük bir paydasını oluşturan OECD üyesi olan ülkeler ve programa katılımında bulunan OECD üyesi olmayan ülkelerde öğrenim gören öğrencilerin, yaşamsal adaptasyonlarına yardımcı olan becerilere sahip olma durumlarını tespit etmeyi, aynı zamanda öğrencilerin bireysel, ebeveyn özelinde ve çevreleri ile bağlantılı kazanımlarında değerlendirme yapmayı hedeflemektedir (MEB, 2016).

PISA'nın yöntemsel bakış açısı, öğrencilerin günlük yaşamda karşılaştıkları sorunlara karşı çözüm oluşturma durumlarını anlamak için tasarlanmıştır. Çünkü PISA'nın temel hedefleri arasında katılımcı ülkelerin eğitim-öğretim politikalarını irdelemek, eğitim stratejilerini geliştirmelerine yardımcı olmak ve öğrencilerde mevcut bulunan yaşamsal becerilere katkı sağlamak için ayrıntılı bir küme meydana getirmek gibi amaçlar da bulunmaktadır (MEB, 2010).

PISA'nın Türkiye'deki uygulaması çerçevesinde yapılan birtakım hazırlıklar bulunmaktadır. Öncelikle PISA 2015 için geliştirilen yeni test maddeleri ve anket soruları, çevirilerinden sonra branş uzmanlarınca incelenerek son haline getirilmektedir. Ardından örneklem için seçilen okulların uygulamaya hazırlanması amacıyla birtakım eğitim toplantıları organize edilir. Hazırlanan eğitim materyalleri aracılığı ile tüm paydaşların PISA'yı anlamaları ve ihtiyaç olan hazırlıkları yapmaları için ortamlar hazır hale getirilmektedir. Öğrenciler, bilgisayar üzerinden soruları yanıtladıkları larken bilgisayarda yaşanacak olası sorunlarda öğrencilere yardımcı olmaları için ilgili görevlendirmeler sağlanmaktadır. Bu sayede farklı nedenlerden oluşabilecek kayıp veri oranı minimuma indirilmeye çalışılmaktadır. Uygulama bitiminde öğrencilerce yanıtlanan açık uçlu soruların puanlanması ilgili branş uzmanları tarafından yapılarak elde edilen veriler Uluslararası Merkez'e iletilmektedir (MEB, 2016).

PISA uygulamalarında öğrenci başarılarını ölçmek için çeşitli değişkenler kullanılmaktadır. Bu değişkenler PISA'nın genel hedeflerini karşılamak amacıyla eğitim-öğretim stratejilerini değerlendirmeye almak ve karşılaştırmalı olarak analiz etmek için sisteme dâhil edilmişlerdir. Değişkenler; “fırsat eşitliği”, “öğrenmeye ayrılan zaman”, “öğrencilerin akademik beklentileri”, “okul öncesi eğitim durumları”, “öğrenci devamsızlığı”, “öğrenmeyi etkileyen öğretmen davranışları”, “öğretmenlerin mesleki gelişimleri”, “öğretmen başına düşen öğrenci sayısı”, “okul kaynakları”, ve “sosyoekonomik göstergeler” olmak üzere sıralanmıştır.

PISA 2015, 35'i OECD üyesi ülke olmak üzere 72 ülkenin katılımıyla gerçekleştirilmiştir. Yaklaşık 30 milyon öğrenciyi temsil eden sınavda 550 bin öğrenci PISA 2015 uygulamasına dâhil edilmiştir. Türkiye’de de 61 ildeki 187 okulda öğrenim gören 5895 öğrencinin katılımı ile gerçekleştirilen uygulamada, %97’lik bir pay ile 9. sınıf ve üzeri sınıflarda okuyan öğrenciler katılım sağlamıştır (OECD, 2014). Türkiye’de gerçekleştirilen PISA uygulamasında da diğer ülkelerde olduğu gibi belli başlı branşlar yerlerini almışlardır.

Tablo 1: PISA uygulamalarındaki temel ve ağırlıklı branşlar (kalın olarak gösterilmiştir)

PISA 2006	PISA 2009	PISA 2012	PISA 2015	PISA 2018	PISA 2022
Okuma	Okuma	Okuma	Okuma	Okuma	Okuma
Becerileri	Becerileri	Becerileri	Becerileri	Becerileri	Becerileri
Matematik	Matematik	Matematik	Matematik	Matematik	Matematik
Okuryazarlık	Okuryazarlık	Okuryazarlık	Okuryazarlık	Okuryazarlık	Okuryazarlık
1	1	1	1	1	1
Fen	Fen	Fen	Fen	Fen	Fen
Okuryazarlık	Okuryazarlık	Okuryazarlık	Okuryazarlık	Okuryazarlık	Okuryazarlık
1	1	1	1	1	1

Tablo 1’te görüldüğü üzere temel branşlar ve ağırlıklı branşlar; “okuma becerileri”, “matematik okuryazarlığı” ve “fen okuryazarlığı” şeklinde üç alandan oluşmaktadır. Burada görülen, PISA’nın “okuryazarlık (literacy)” olarak adlandırdığı literatür kavramına, “Öğrencilerin, edindiği bilgileri günlük yaşama uyarlama, mantıksal çıkarımlarda bulunma, farklı durumlar ile ilgili problemleri yorumlama ve çözme kapasitesi” şeklinde bir tanım getirilmiştir (EARGED, 2007). Bu bağlamda ise PISA 2015 uygulamasında ağırlıklı branş olarak fen okuryazarlığı incelenmiştir. Fen okuryazarlığı, “Etkin bir vatandaş olarak fen

bilimleri ile ilgili düşünceler ve meseleler ile uğraşabilme yetisi” şeklinde tanımlanmıştır (MEB, 2016). PISA’ya göre ise bu tanımlama, “Öğrencilerin fen bilgisini nerede, nasıl ve ne ölçüde etkili kullandığı” şeklindedir (EARGED, 2010).

3.2 Araştırmanın Örneklemi ve Evreni

3.2.1 PISA 2015 Uygulamasında Türkiye Örneklem Yapısı

Türkiye’de uygulanan PISA 2015 sınavı; 15 yaş grubu öğrenci evreni 1324089 kişi, sınava katılım gösterebilecek ulaşılabilir Türkiye evreninde ise 925366 kişi ile gerçekleştirilmiştir. “Tabakalı Seçkisiz Örnekleme Yöntemi” ile okul örnekleme belirlenen PISA 2015 uygulaması için ilk etapta “İstatistikî Bölge Birimleri Sınıflaması (İBBS)” Düzey 1; “eğitim türü”, “okul türü”, “okulların bulundukları yer” ve “okulların idari biçimleri” şeklindeki tabakalar kullanılarak bu yöntem çerçevesinde kararlaştırılmış, son etapta ise belirlenen okullarda uygulamaya dahil olacak öğrenciler seçkisiz yöntem ile belirlenmiştir (MEB, 2016).



Şekil 1: İstatistikî Bölge Birimleri Sınıflaması, Düzey 1 haritası (MEB, 2016)

PISA 2015 uygulamasına Türkiye’de İBBS Şekil 1’e göre 12 bölgeyi temsil eden 61 ilden 187 okul ve 5895 öğrenci katılmıştır. İBBS Düzey 1 haritası Harita 2’de, PISA 2015 Türkiye örnekleminde yer alan okul ve öğrencilerin, bu bölgelere dağılımları ise Tablo 2’te gösterilmektedir (MEB, 2016) :

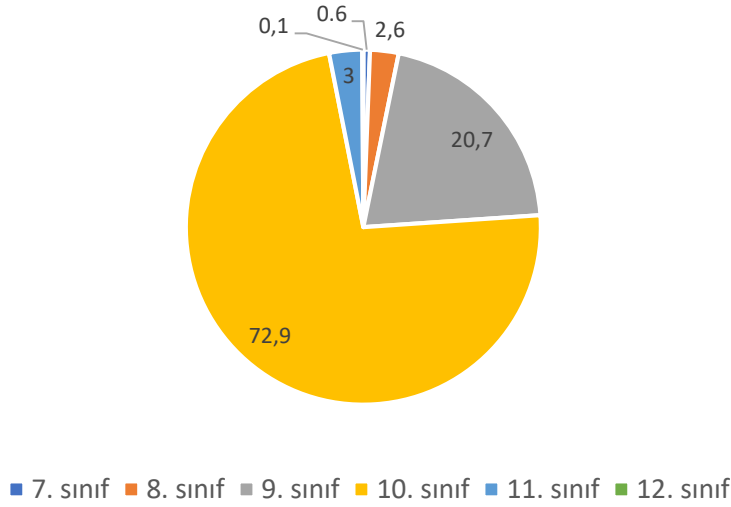
Tablo 2: İBBS’ye göre PISA 2015 Türkiye örneklemindeki öğrenci sayıları ve yüzdeleri

Bölge Kodu	Bölge İsmi	Katılan Öğrenci Sayısı	Katılan Öğrenci Yüzdesi
TR1	İstanbul	1070	18.15
TR2	Batı Marmara	245	4.16
TR3	Ege	707	11.99
TR4	Doğu Marmara	510	8.65
TR5	Batı Anadolu	553	9.38
TR6	Akdeniz	817	13,86
TR7	Orta Anadolu	334	5.67
TR8	Batı Karadeniz	303	5.14
TR9	Doğu Karadeniz	194	3.29
TRA	Kuzeydoğu Anadolu	199	3.38
TRB	Ortadoğu Anadolu	276	4.68
TRC	Güneydoğu Anadolu	687	11.65
	TOPLAM	5895	100

Tablo 2’ye bakıldığında, PISA 2015 Türkiye uygulamasında öğrenci sayısının en fazla olduğu bölgenin İstanbul (TR1) bölgesi, en az olduğu bölgenin ise Doğu Karadeniz (TR9) bölgesi olduğu görülmektedir. PISA 2015 sonuçları incelendiğinde, Türkiye’nin düşük başarı skorları ortaya çıkaran ülkeler arasında bulunduğu dikkat çekmektedir. İstatistiklere göre 72 ülkenin fen bilimleri başarı skoru ortalaması “465” puan iken Türkiye’nin ortalaması “425” şeklinde kayıtlara geçmiştir ve bu nedenler ile Türkiye 72 katılımcı ülke arasından 54. sırada yer almıştır (MEB, 2016). Fen bilimleri başarı skorları açısından ortalama olarak en iyi başarıyı gösteren ülkelerin başında Singapur ve Japonya varken en düşük performans gösteren ülkelerin başında Tunus ve Makedonya bulunmaktadır (MEB, 2016).

Türkiye’de PISA’ya katılan öğrencilerin büyük çoğunluğunu 9 ve 10. sınıf öğrencileri oluşturmaktadır. PISA 2015 uygulamasına katılan öğrencilerin sınıf seviyelerine göre dağılımları Şekil 2’de gösterilmiştir:

Sınıf Düzeyine Göre Dağılım



Şekil 2: 2015 PISA’ya katılan öğrencilerin sınıf düzeylerine göre dağılımı (%)

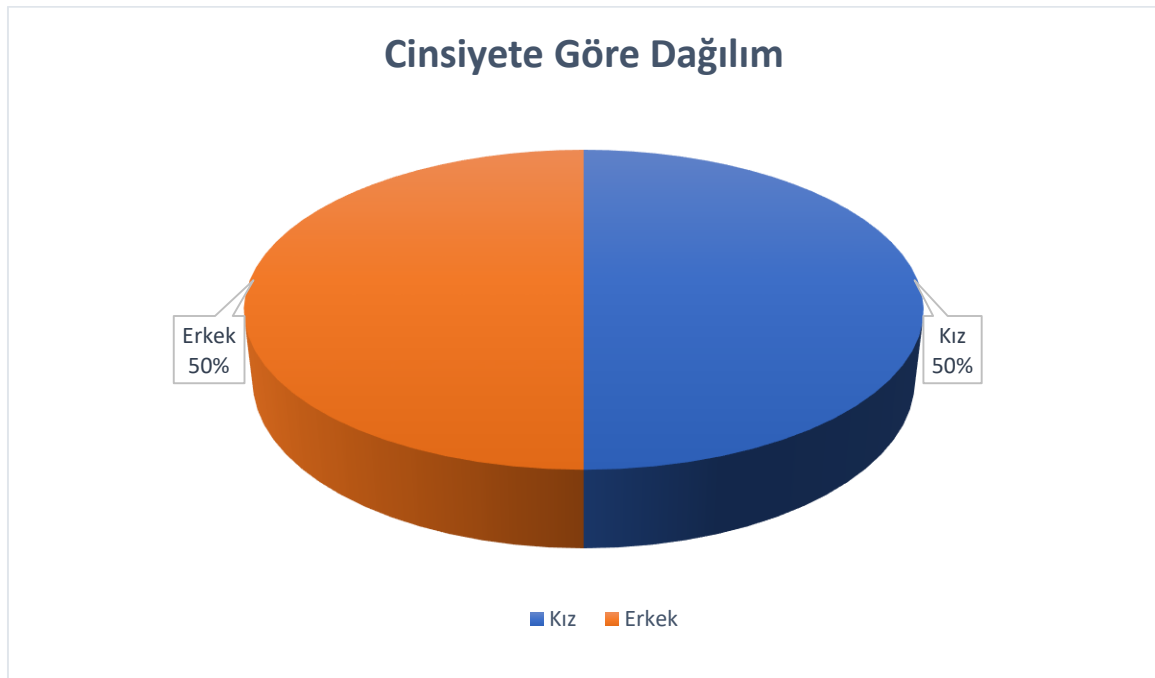
PISA 2012 sınavına katılım sağlayan öğrencilerin %27,6’sı 9. sınıf öğrencisi ve %65,4’ü 10. sınıf öğrencisi iken, PISA 2015 sınavına katılım sağlayan öğrencilerin %20,7’si, 9. sınıf öğrencisi ve %72,9’u 10. sınıf öğrencisidir. Bu durum, 2012 sınavı ile karşılaştırıldığında 2015 sınavındaki 15 yaş öğrencilerinin daha büyük bir bölümünün bulunması gereken sınıf seviyesinde olduğunu göstermektedir (MEB, 2016).

Tablo 3: PISA 2015 Türkiye örnekleminin okul türlerine göre dağılımı

Okul Türü	Öğrenci Yüzdesi
Ortaokul	2,0
Anadolu Lisesi	38,1
Fen Lisesi	2,1
Sosyal Bilimler Lisesi	1,4
Güzel Sanatlar Lisesi	0,7
Mesleki ve Teknik Anadolu Lisesi	36,4
Çok Programlı Anadolu Lisesi	4,1
Anadolu İmam Hatip Lisesi	14,4
Cevap Vermeyen	0,3
Ulaşılamayan	0,5
TOPLAM	100

Okul türlerine göre dağılım açısından ise PISA 2015 Türkiye örnekleminde öğrenci oranının en yüksek olduğu okul türleri, Anadolu Lisesi ile Mesleki ve Teknik Anadolu Lisesi'dir (%75). Öğrenci oranının en düşük olduğu okul türü ise Güzel Sanatlar Lisesi'dir (%0,7) (MEB, 2016). Bu veriler Tablo 3'te gösterilmiştir:

PISA 2015 Türkiye'den sınavına katılım sağlayan öğrencilerin cinsiyet dağılımları Şekil 3'te gösterilmektedir. Grafik incelendiğinde bu değişken bazındaki örneklemin eşit bir biçimde dağılım sergilediği görülmektedir:



Şekil 3: 2015 PISA'ya katılan öğrencilerin cinsiyet dağılımı

Türkiye'nin son dönem performanslarında gözlenen artış ile düşük başarı gösteren diğer ülkeler içinden önemli bir farkın açıkça ayırt edildiği görülmektedir (MEB, 2022). PISA Türkiye sonuçlarında dikkat çeken farklı bir özellik de kırsal bölgelerde yaşamlarını sürdüren öğrenciler ile şehirlerde yaşayan öğrenciler arasında bulunan başarı skorları farklarının diğer ülkelere kıyasla oldukça yüksek olmasıdır (MEB, 2022).

PISA yeterlik seviyelerine göre sonuçların raporlanmasında “asgari performans düzeyi” ve “üst performans düzeyi” olarak iki ayrı seviye bulunmaktadır. PISA uygulaması araştırmalarına göre asgari performans düzeyi, “15 yaş grubundaki öğrencilerin günlük

yaşamda kullanabileceği bilgi ve beceriler”i tanımlamaktadır. Üst performans düzeyi ise “PISA’da değerlendirilen alanlarda karmaşık bilişsel süreçlerin ve yeterlikler”in tamamıyla tanımlandığı seviyedir. Ülkelerin ekonomik kalkınmaları için gerekli beşeri sermayenin çoğunlukla bu düzeydeki öğrencilerden olabileceği kabul edilmektedir (OECD, 2022). PISA 2015’e göre PISA 2022’de OECD ülkeleri ve tüm ülkelerdeki üst performans düzeyindeki öğrenci oranının azaldığı görülürken Türkiye’de daha fazla oranda öğrencinin üst düzey yeterliklere sahip olduğu anlaşılmaktadır. PISA 2018 ile kıyaslandığında ise PISA 2022’de OECD ülkeleri ve tüm ülkelerde olduğu gibi Türkiye’de de üst performans düzeyindeki öğrenci oranı azalmıştır (OECD, 2022).

Tablo 4: Yıllara Göre Fen Okuryazarlığı Ortalama Puanları

	2022	2018	2015	2012	2009	2006
OECD Ortalaması	485	487	493	501	495	498
Tüm Ülkeler Ortalaması	447	453	465	477	471	478
Türkiye Ortalaması	476	466	425	463	454	424
Türkiye Sıralaması	39	40	54	43	42	47
Katılan Ülke Sayısı	81	79	72	65	65	57

3.3 Çalışmanın Metodolojik İşlem Adımları

Bu çalışmanın temelini oluşturan metodolojik yaklaşımın 5 temel adımda özetlemek mümkündür. Bunlar:

Adım 1: Veri Kaynaklarının Seçimi ve Tanımlanması:

Araştırmada kullanılan veriler, OECD tarafından düzenlenen PISA 2015 sınavına ait veri setlerinden elde edilmiştir. Bu veri seti, öğrencilerin fen bilimleri başarı skorları ile sosyo-ekonomik, demografik ve eğitimsel faktörleri kapsamaktadır (OECD, 2015).

Veri Setinin Yapısı: Veri seti, her bir öğrenciye ait bağımsız değişkenleri (önceki bir çalışmada (Hacıoğlu, 2024) tespit edilmiş olan önem seviyesi yüksek 99 tane öğrenci odaklı

değişken) ve bağımlı değişkenleri (her öğrenci için ayrı ayrı hesaplanan 10 adet fen başarı skorunun ortalaması olan puan seviyesi) içermektedir.

Örnekleme: Araştırmada, Türkiye'yi temsil eden 12 istatistiki bölge birimleri (TR1-TRC) baz alınmış ve her bölgeden ayrıntıları Kısım 3.2 de verilmiş olan belirli sayıda öğrenci verisi kullanılmıştır.

Veri Temizleme: Eksik, hatalı ve aykırı değerler, sistematik olarak tespit edilmiş ve veri işleme öncesinde imputasyon yöntemleri kullanılarak düzeltilmiştir. R programlama ortamında kullanılan "**mice**" (van Buuren ve Groothuis-Oudshoorn, 2011) kütüphanesi yardımıyla verilerin eksiksiz hale getirilmesi sağlanmıştır.

Adım 2: Veri Setinin Hazırlanması ve Bölgesel Ayrımalar

Veriler, her bir istatistiki bölgeye göre ayrıştırılmış ve modelleme öncesinde öğrenme algoritmalarına göre işlem yapılabilir biçime uygun hale getirilmiştir:

Eğitim ve Test Veri Setlerinin Oluşturulması: Veriler %80 eğitim ve %20 test veri seti olacak şekilde ayrılmıştır (Gholamy, Kreinovich ve Kosheleva, 2018).

Normalize Edilmiş Veriler: Tüm değişkenler ortalama ve standart sapma kullanılarak normalize edilmiş ve yanlı (bias) öğrenmenin önüne geçilmiştir.

İstatistiki Özetler: Her bir bölge için fen başarıları skorlarının ortalama, medyan ve standart sapma gibi özet istatistikleri hesaplanmıştır. Böylece tahmin sonuçları ile istatistiksel olarak karşılaştırılabilir standartlar elde edilmiştir.

Adım 3: Makine Öğrenmesi Modellerinin Kurulması:

Öğrencilerin PISA 2015 fen başarı skorlarının tahmini için, ayrıntıları Kısım 3.4 de verilmiş olan dört farklı makine öğrenmesi algoritması kullanılmıştır. Her model, çapraz doğrulama yöntemleriyle test edilmiş ve hiperparametre optimizasyonu R ortamında çalışan "caret" (Kuhn ve diğerleri, 2020) paketi yardımıyla gerçekleştirilmiştir.

Adım 4: Model Performansının Değerlendirilmesi

Modellerin tahmin performansları ayrıntıları Kısım 3.5 de verilmiş olan istatistiksel metrikler kullanılarak değerlendirilmiştir.

Adım 5: Bulguların Yorumlanması

Performans değerlendirilmesi sonucunda test veri setini en doğru şekilde temsil edecek biçimde tahmin performansı gösteren model tespit edilmiş ve bulgular SHAP değerleri, istatistiksel benzerlik yaklaşımları ve jaccard eş benzerlik indeksine göre karşılaştırılmıştır.

Önem Derecesi Belirleme: Değişkenlerin modeldeki katkılarını belirlemek için SHAP değerleri karşılaştırması yapılmıştır. SHAP değerlerinin tespitine ilişkin ayrıntılı bilgi Kısım 3.6 da verilmiştir.

İstatistiki Analizler: Tahmin sonuçları ve gerçek skorlar arasındaki merdağılım ve merkezi ölçü birimi benzerlikleri istatistiki testlerle sınanırken, önem seviyesi yüksek değişkenlerin farklı bölgelerdeki benzerlik düzeyleri, Jaccard indeksi yardımıyla analiz edilmiştir. Bu indekse ait ayrıntılı bilgi Kısım 3.7 de verilmiştir.

3.4 Makine Öğrenme Algoritmaları

3.4.1 Aşırı Gradyan Artırma (Extreme Gradient Boosting, XGBoost)

XGBoost (eXtreme Gradient Boosting), modern makine öğrenmesi uygulamalarında sıklıkla kullanılan, yüksek performanslı ve ölçeklenebilir bir algoritmadır (T. Chen ve Guestrin, 2016b; Guang, 2021). Bu algoritmanın tahmin performansında yüksek doğruluk göstermesinin temelinde, gradyan artırma (gradient boosting) prensibi (Friedman, 2001) ve performansı artıran çeşitli optimizasyon teknikleri yatmaktadır (Sheng ve Yu, 2022; Su, Chen ve Qian, 2024). Boosting, zayıf öğrencilerin bir araya getirilerek güçlü bir öğrenci oluşturulmasını sağlayan bir topluluk öğrenme yöntemidir. XGBoost, bu yöntemi karar ağaçlarını zayıf öğrenci olarak kullanarak uygular. Algoritma, iteratif bir iyileştirme süreciyle çalışır. Her adımda, bir önceki modelin hatalarını (artıklarını) minimize etmeye odaklanan yeni bir karar ağacı eklenir (Zhang, Jia ve Shang, 2022). Hatalı tahmin edilen örneklere daha fazla ağırlık verilerek modelin her adımda daha iyi hale getirilmesi amaçlanır (M. Chen ve diğerleri, 2024; Li, Yin, Quan ve Zhang, 2019; Su ve diğerleri, 2024).

XGBoost, geleneksel gradyan artırma algoritmalarına kıyasla bir dizi yenilikçi özellik sunar:

- **Düzenlileştirme (Regularization):** XGBoost, L1 ve L2 düzenlileştirme tekniklerini kullanarak model karmaşıklığını kontrol altına alır ve aşırı öğrenmeyi engeller (T. Chen ve Guestrin, 2016). L1 düzenlileştirme, daha fazla sıfır ağırlık oluşturarak modelin

sadeleştirilmesine yardımcı olurken, L2 düzenleme ağırlıkları küçültmek aşırı öğrenmeyi (overfitting) önlemeyi amaçlar (Açıışlı Çelik ve Yeşilkanat, 2023; Zhang ve diğerleri, 2022).

- Ağaç Budama (Tree Pruning): XGBoost, oluşturulan karar ağaçlarını budayarak (prune) modelin genelleme yeteneğini artırır. Geleneksel yöntemlerden farklı olarak XGBoost, ağacı önce tamamen büyütür ve maksimum kazancı (maximum gain) sağlamayan dalları keser (Sakata, Ohama ve Taniguchi, 2018).

- Parallelleştirme: XGBoost, karar ağaçlarının oluşturulması ve eğitimi sırasında paralel işlem yapabilme yeteneğiyle diğer boosting yöntemlerinden (örneğin, GBM) daha hızlıdır (Mitchell ve Frank, 2017).

- Seyrek Verilere Destek: XGBoost, eksik verileri (missing values) otomatik olarak işleyebilir ve seyrek veri kümelerinde etkili bir şekilde çalışabilir (T. Chen ve Guestrin, 2016).

- Ağırlıklandırılmış Nicelik Çizimi (Weighted Quantile Sketch): XGBoost, büyük veri setlerinde ağaçların daha hızlı bir şekilde oluşturulmasını sağlayan ağırlıklandırılmış nicelik çizimi tekniğini kullanır. Bu teknik, veriyi parçalara (quantile) bölerek ve parçaları temsil eden ağırlıkları kullanarak hesaplama karmaşıklığını azaltır (Wen, He, Kotagiri, Lu ve Shi, 2018).

XGBoost algoritması, kayıp fonksiyonunu minimize etmek için gradyan iniş yöntemini kullanır. Bu işlemde temel amaç, tahmin edilen değerler ile gerçek değerler arasındaki farkı temsil eden kayıp fonksiyonunu mümkün olduğunca küçültmektir. XGBoost'un amaç fonksiyonu, aşağıdaki gibi ifade edilir (Açıışlı Çelik ve Yeşilkanat, 2023) (Açıışlı Çelik ve Yeşilkanat, 2023; Sheng ve Yu, 2022):

$$\text{AMAÇ} = \sum_{i=1}^n \ell(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

Burada:

- $\ell(y_i, \hat{y}_i)$: Tahmin edilen $\hat{y}_i$ değeri ile gerçek y_i değeri arasındaki kayıp fonksiyonunu temsil eder. Örnek olarak, ortalama karesel hata (MSE) veya log-kayıp fonksiyonları kullanılabilir.

- f_k : Karar ağacı modelini temsil eder.

- $\Omega(f_k)$: Modelin karmaşıklığını cezalandıran düzenleme terimini ifade eder. Bu terim aşağıdaki gibi temsil edilmektedir:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_j \omega_j^2$$

Burada:

- T : Karar ağacındaki yaprak sayısı,
- w_j : Yaprak değerleri,
- λ ve γ : hiper parametre değerleridir.

Gradyan Artırma Adımları ise 3 adımda özetlenebilir: Bunlar;

1. Başlangıç: İlk olarak, basit bir sabit model tahmini $\hat{y}_i^{(0)}$ ile başlanır:

$$\hat{y}_i^{(0)} = \arg \min_{\theta} \sum_{i=1}^n \ell(y_i, \theta)$$

2. Yeni Ağaç Eğitimi: Her bir adımda, yeni bir karar ağacı modeli (f_k) eğitilir. Amaç, bir önceki modelin tahmin hatalarını düzeltmektir:

$$r_i^{(m)} = - \frac{\partial \ell(y_i, \hat{y}_i^{(m-1)})}{\partial \hat{y}_i^{(m-1)}}$$

Burada $r_i^{(m)}$, m 'ninci iterasyon sonunda modelin gradyanı (rezidüel hatası) olarak adlandırılır.

3. Model Güncellemesi: Yeni model, aşağıdaki şekilde güncellenir:

$$\hat{y}_i^{(m)} = \hat{y}_i^{(m-1)} + \eta f_m(x_i)$$

Burada η , modelin aşırı öğrenmesini engellemek için kullanılan öğrenme oranıdır.

XGBoost algoritması, kayıp fonksiyonunu optimize etmek için Taylor serisinin ikinci dereceden yaklaşımını kullanır. Bu yaklaşım, optimizasyonu daha hassas hale getirir. İkinci dereceden yaklaşım şu şekilde ifade edilir:

$$\ell(y_i, \hat{y}_i + f(x_i)) \approx \ell(y_i, \hat{y}_i) + g_i f(x_i) + \frac{1}{2} h_i f(x_i)^2$$

Burada: g_i : Birinci türev (gradyan), h_i : İkinci türev (Hessian matrisi) olarak tanımlanır. Bu işlerin sonunda Yeni ağacın optimizasyon amacı, aşağıdaki şekilde sadeleştirilmiş olur:

$$\text{AMAÇ} = \sum_{i=1}^n \left(g_i f(x_i) + \frac{1}{2} h_i f(x_i)^2 \right) + \Omega(f)$$

3.4.2 Destek Vektör Regresyonu (Support Vector Regression, SVR)

Destek Vektör Regresyonu (Support Vector Regression, SVR), Destek Vektör Makineleri'nin (Support Vector Machines, SVM) (Cortes ve Vapnik, 1995) bir uzantısı olup, ilk olarak (Drucker, Surges, Kaufman, Smola ve Vapnik, 1997) Drucker ve arkadaşları (1997) tarafından geliştirilmiştir. SVM'ler, sınıflandırma problemlerinde veri noktalarını en iyi şekilde ayıran en uygun hiperdüzlemi bulmayı amaçlar. Bu yaklaşım, farklı sınıflara ait verilerin maksimum marj ile ayrılmasını sağlar (Y. Lin, Lee ve Wahba, 2002). SVR ise bu prensibi regresyon problemlerine uyarlayarak, tahmin edilen değerlerin hedef değerlerin belirli bir epsilon (ϵ) tolerans aralığı içinde kalmasını hedefler (Yu, Chen ve Chang, 2006). Bu epsilon aralığı, modelin kabul edilebilir hata payını belirler ve tahminlerin ne kadar hassas olması gerektiğini ifade eder. Model, parametrelerini üst ve alt sınırları tanımlayan sapmaları en aza indirecek şekilde optimize eder (Suganyadevi ve Babulal, 2014). Optimizasyon sürecinde, modelin karmaşıklığını kontrol etmek ve belirlenen hata eşliğini aşan sapmalara cezalar uygulamak için düzenleme (regularization) terimleri kullanılır (Balabin ve Lomakina, 2011; Yeşilkanat ve Akkoyun, 2023). Bu sayede, model hem yüksek doğruluklu tahminler yapar hem de aşırı uyum (overfitting) riskini azaltmış olur.

SVR'nin temelini oluşturan kavram, Yapısal Risk Minimizasyonu (Structural Risk Minimization, SRM) teorisidir (Wu, Chau ve Li, 2008). SRM, istatistiksel öğrenme teorisinin bir parçası olarak, modelin genel hata oranını en aza indirirken aynı zamanda

model karmaşıklığını da kontrol altında tutmayı hedefler. Bu yaklaşım, modelin sadece eğitim verisinde değil, aynı zamanda daha önce görülmemiş yeni veriler üzerinde de iyi performans göstermesini sağlar (De, 2003; Kim ve Cho, 2008). Ayrıca SRM, modelin genelleme yeteneğini artırarak, aşırı uyum riskini azaltır ve böylece modelin gerçek dünya uygulamalarında daha güvenilir ve tutarlı sonuçlar üretmesine katkıda bulunur. Bu teoriye dayanan SVR, model performansını optimize etmek için eğitim hatası ile model karmaşıklığı arasında bir denge kurar.

SVR, genellikle giriş verilerini yüksek boyutlu bir özellik uzayına dönüştürmek için Radyal Taban Fonksiyonu (Radial Basis Function, RBF) gibi çeşitli çekirdek fonksiyonları kullanır (Wu ve diğerleri, 2019). Çekirdek fonksiyonları, orijinal veriyi daha yüksek boyutlu bir uzaya haritalayarak doğrusal olmayan veri ilişkilerinin doğrusal bir biçimde modellenmesini mümkün kılar. Özellikle RBF çekirdek fonksiyonu, veriler arasındaki mesafeye dayalı olarak benzerlik ölçümü yapar ve bu sayede modelin karmaşık desenleri ve ilişkileri öğrenmesine olanak tanır. Bu yöntem, karmaşık ve doğrusal olmayan ilişkiler içeren veri kümelerinde modelin performansını belirgin ölçüde artırabilir. Ayrıca, çekirdek fonksiyonlarının kullanımı, SVR'nin esnekliğini artırarak farklı türdeki veri setlerine uyum sağlamasını kolaylaştırır.

Genel olarak SVR algoritması, modelin hata toleransı ile sınır parametreleri arasında bir denge kurmayı amaçlayan dışbükey optimizasyon tekniklerini kullanır. Optimizasyon sürecinde, modelin karmaşıklığını azaltarak tahmin hatalarını düşüren düzenlileştirme terimleri yer alır. Bu düzenlileştirme, model parametrelerinin büyüklüğünü kontrol ederek aşırı uyumu önler ve genelleme kapasitesini güçlendirir (Hastie, Friedman ve Tibshirani, 2009). Sonuç olarak SVR modelleri, eğitim veri kümesinde aşırı öğrenmeyi en aza indirerek yeni veriler üzerinde daha iyi performans sergiler. Bu özellikleri sayesinde SVR, regresyon problemlerinde güçlü ve esnek bir yaklaşım sunmaktadır.

3.4.3 Bayes Düzenlemeli Yapay Sinir Ağları (Bayesian Regulation Artificial Neural Network, BRNN)

Yapay Sinir Ağları (YSA), insan beyninin işleyişini taklit etmek amacıyla tasarlanmış gelişmiş hesaplama araçlarıdır (Rumelhart, Hinton ve Williams, 1986). Bu ağlar, biyolojik sinir hücrelerine benzer yapay nöronlar temelinde inşa edilmiştir. Nöronlar, sinaptik ağırlıklar aracılığıyla birbirleriyle etkileşime girerek verileri işler ve bu sayede ileri beslemeli ağlarda veri akışını kolaylaştırır (Akkoyun, 2020). Veriler, sinaptik ağırlıklar

üzerinden ardışık nöron katmanlarına iletilirken dönüşüme uğrar. Her bir nöronda, gelen veriler önceden tanımlanmış fonksiyonlar kullanılarak birleştirilir ve net bir girdi oluşturulur; ardından bu girdi, başka bir aktivasyon fonksiyonu ile işlenir (Yesilkanat, Kobya, Taskin ve Çevik, 2017). YSA metodolojilerinin temel amacı, nöronları birbirine bağlayan sinaptik ağırlıkları optimize etmektir. Başlangıçta rastgele değerler olarak atanan bu ağırlıklar, beklenen ve gerçek ağ çıktıları arasındaki farkı minimize etmek için yinelemeli olarak ayarlanır. İyi yapılandırılmış bir YSA, karmaşık desenleri algılayabilir ve bunlara uyum sağlayarak hedef değerlere yakın sonuçlar üretebilir (Akkoyun ve Bayram, 2014).

Geleneksel YSA'lara “Bayes” temelli düzenleme metodolojilerinin eklenmesiyle Bayes Düzenlemeli Yapay Sinir Ağları (BRNN) geliştirilmiştir (Kwok ve Yeung, 1996). BRNN, Bayes yönteminin model belirsizliğini ele alma ve YSA'lardaki aşırı öğrenmeyi hafifletme kapasitesini entegre eder. Bayes düzenleme bağlamında, model parametrelerine önceden tahmin edilen olasılık dağılımları atanır; bu tahminler, öğrenme sürecinde elde edilen verilerle geliştirilir ve böylece olası parametre değerleri hakkında hem mevcut bilgi hem de yeni veriler kullanılmış olur (Fryer, Strümke ve Nguyen, 2020). BRNN çerçevesi, ağırlıklarını sadece eğitim verilerine değil, aynı zamanda ön bilgiye bağlı olarak gerçekleştirir. Böylece parametre tahminlerinin gerçekçi ve güvenilir kalması sağlanarak ve aşırı öğrenme durumu önlenmeye çalışılır (Burden ve Winkler, 1999). BRNN, modellemedeki belirsizlikleri yönetmek için net bir yaklaşım sunar; belirli bir girdi için yalnızca tek bir tahmin üretmekle kalmaz, aynı zamanda o tahminin kesinliğini de belirli dağılımlar çerçevesinde değerlendirir (Dong, An, Lu ve Geng, 2022). Bu durum, özellikle pratik uygulamalarda tahminlerin doğruluğu ve risk değerlendirmeleriyle ilgili önemli içgörüler sağlar. Ayrıca, Yapay sinir ağları için önemli bir tehlike olan, model karmaşıklığının otomatik olarak yönetilmesini mümkün kılar. Bu nedenlerden dolayı, BRNN, geleneksel YSA'lara kıyasla daha güvenilir, sağlam ve uyarlanabilir modelleme sonuçları üretir. Ayrıca, güvenilirliği ve çok yönlülüğü nedeniyle, modelin uygulama kapsamını önemli ölçüde genişletir. Aşırı öğrenmeye direnme yeteneği ve belirsizliği yönetmek için gelişmiş özellikleri birleştirerek, modelin karmaşık veri setlerinden temel sınıflandırma ve tahmin faaliyetleri için bilgiyi daha rasyonel bir şekilde yorumlamasını sağlar.

BRNN modelleri, özellikle değişkenlik ve belirsizlikle karakterize edilen veri setlerini işlerken, geleneksel YSA'lara kıyasla daha sağlam sonuçlar üretme kapasitesine sahiptir (F.

Burden ve Winkler, 2009). Bayes düzenlileştirmenin uygulanmasıyla model, öğrenme aşamasında parametreler hakkında geniş bir perspektif elde eder. Böylece modelin genelleştirme kabiliyetini artırırken YSA'nın zayıf noktası olan aşırı öğrenme durumu önlenmiş olur. Ayrıca, model belirsizliğinin kapsamlı bir analizi, karar verme süreçlerinde daha bilinçli seçimler yapılmasını kolaylaştırır. Bayes yöntemi, modelin öğrenme sürecinde yalnızca gerekli parametreleri kullanmasına yardımcı olur. Bu yaklaşım, hesaplama maliyetlerini azaltır ve modelin genelleştirme kabiliyetini artırır. Böyle bir strateji, eğitim süresini azaltarak hesaplama kaynaklarının kullanımını optimize ettiği için, büyük veri setlerini işlerken çok önemli bir özellik olarak öne çıkmaktadır.

3.4.4 Rastgele Orman (Random Forest, RF)

Rastgele Orman (Random Forests, RF) algoritması, makine öğrenmesi alanında sınıflandırma, regresyon ve değişken önem sıralaması gibi çeşitli görevler için yaygın olarak kullanılan güçlü ve esnek bir algoritmadır (Verikas, Gelzinis ve Bacauskiene, 2011). Breiman'ın 2001'de kavramsallaştırdığı RF algoritması (Breiman, 2001), Ho'nun 1998'de geliştirdiği Rastgele Alt Uzaylar yöntemi ile yine Breiman'ın 1996'da tanıttığı Bagging (Bootstrap Aggregating) tekniğinin birleşimi ile ortaya çıkmıştır (Breiman, 1996; (Ho, 1998). Bu birleşim, birden fazla karar ağacının birlikte kullanıldığı temel bir makine öğrenmesi yöntemi sunar. RF'nin temel yapısı, her biri verilerin rastgele bir alt kümesi üzerinde eğitilen ve özelliklerin rastgele bir alt kümesini kullanan çok sayıda karar ağacının bir araya getirilmesine dayanır (Bernard, Adam ve Heutte, 2012; Panov ve Džeroski, 2007). Bu topluluk (ensemble) öğrenme yöntemi ile, bireysel karar ağaçlarının aşırı uyum ve yüksek varyans sorunları hafifletilerek, daha sağlam ve genelleştirilebilir bir model yapısı oluşturulur. Bu modelde kullanılan karar ağaçları, veri kümesini bir dizi ikili bölme kuralına göre homojen alt kümelerle ayırmayı amaçlayan hiyerarşik yapılardır. Ancak, tek başına kullanılan karar ağaçları, eğitim verilerindeki küçük değişikliklere karşı aşırı duyarlılık göstererek genelleme performansını düşürebilir (Scornet, Biau ve Vert, 2015). RF algoritması, modelin tahmin doğruluğunu artırmak ve aşırı uyumu azaltmak için iki ana strateji uygular:

- **Bagging (Bootstrap Aggregating) işlemi:** Orijinal veri kümesinden rastgele örneklerle (bootstrap örnekleri) birden fazla eğitim kümesi oluşturulur. Her bootstrap örneği, orijinal veri kümesiyle aynı boyutta olup bazı gözlemleri tekrar içerirken bazılarını dışarıda bırakır. Bu çeşitli eğitim kümeleri üzerinde birden fazla karar ağacı eğitilir ve nihai tahmin,

tüm ağaçların tahminlerinin birleştirilmesiyle elde edilir (burada sınıflandırma için çoğunluk oylaması, regresyon için ortalama alma yöntemi kullanılır) (Biau ve Scornet, 2016; Breiman, 2001).

- **Rastgele Alt Uzay yöntemi:** Her karar ağacını oluştururken, özelliklerin (tahmin edici olarak kullanılan değişkenlerin) rastgele bir alt kümesi kullanılır. Bu yaklaşım, her ağacın verilerin farklı bir yönüne odaklanmasını sağlayarak topluluğun çeşitliliğini artırır ve genelleme yeteneğini geliştirir (Ho, 1998); Panov ve Džeroski, 2007).

Her bir karar ağacının inşası sırasında, RF algoritması her düğümde en iyi bölme noktasını belirlemek için Gini indeksi veya entropi gibi bir bölme kriteri kullanır. RF'deki ağaçlar genellikle budama yapılmaksızın maksimum derinliğe kadar büyütülür, bu da her ağacın mümkün olan en düşük saflık seviyesine ulaşmasını sağlar (Bernard ve diğerleri, 2012).

RF'nin önemli avantajlarından biri, tahmin için hangi özelliklerin en önemli olduğunu belirleyebilmesidir. Değişken önem ölçütleri, her özelliğin bölünmelerde ne sıklıkla kullanıldığına ve bu bölünmelerin modelin doğruluğunu ne kadar artırdığına dayanarak hesaplanır (Verikas ve diğerleri, 2011). Bootstrap örnekleme ve rastgele özellik seçimi, modelin çeşitli veri alt kümelerinden ve özellik kombinasyonlarından etkili bir şekilde öğrenmesini sağlar. Ayrıca, her bir ağacın performansı genel modelde eşit ağırlığa sahip olmayabilir; daha doğru tahminler yapan ağaçlar daha güçlü bir etkiye sahipken, daha fazla hata yapanların etkisi sınırlıdır. Bu durum, RF algoritmasında zayıf performans gösteren ağaçların katkısını azaltan bir "**ceza sistemi**"nin varlığını gösterir (Prasad, Iverson ve Liaw, 2006). RF algoritmasının matematiksel ve istatistiksel temelleri hakkında daha ayrıntılı bilgi için Biau ve Scornet (2016) ve Breiman (2001) gibi kaynaklara başvurulabilir.

3.5 Model Değerlendirme Metrikleri

Bu çalışmada, kullanarak Fen okuryazarlığı, matematik okuryazarlığı ve okuma becerileri puanlarını tahmin etmek için kullanılan makine öğrenimi modellerinin performansları çeşitli değerlendirme ölçütlerine göre yapılmıştır. Araştırmada modellerin tahmin hassasiyetini ölçmek için Ortalama Karekök Hata (RMSE), Ortalama Mutlak Hata (MAE), Belirlilik Katsayısı (R^2) ve Lin'in Uyum Korelasyon Katsayısı (ρ_c) gibi performans göstergeleri uygulanmıştır (Lin, 1989; Yesilkanat, 2020).

RMSE, tahmin edilen ve gerçek değerler arasındaki farkların büyüklüğüne odaklanarak modelin önemli tutarsızlıklara olan eğilimini göstererek daha büyük hataları vurgular. Buna karşılık MAE, tahmin edilen ve gerçek değerler arasındaki mutlak sapmaların ortalamasını hesaplayarak hem büyük hem de küçük hatalara eşit ağırlık verir ve böylece basit bir hata değerlendirmesi sağlar. R^2 , modelin bağımlı değişkenin davranışını ne kadar iyi tahmin ettiğini veya açıkladığını gösterir. Yüksek bir R^2 güçlü bir açıklama kabiliyetine işaret ederken, düşük bir R^2 modelin optimizasyon gerektirebileceği anlamına gelir. ρ_c , hem doğrusal korelasyonları hem de iki ölçüm seti arasındaki genel uyumu dikkate alarak, tahmin edilen ve gözlemlenen değerler arasındaki uyumu değerlendirir. Bu metrik, modelin hassasiyetinin ve güvenilirliğinin kapsamlı bir şekilde değerlendirilmesini kolaylaştırır.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (A_i - P_i)^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - P_i|$$

$$R^2 = 1 - \frac{\sum (A_i - P_i)^2}{\sum (A_i - \mu_A)^2}$$

$$\rho_c = \frac{2\rho\sigma_A\sigma_P}{\sigma_A^2 + \sigma_P^2 + (\mu_A - \mu_P)^2}$$

Burada, A_i gerçek değerleri, P_i tahmin edilen değerleri, n tam örnek sayısını, ρ Pearson korelasyon katsayısını, σ_A ve σ_P sırasıyla gerçek ve tahmin edilen değerlerin standart sapmalarını, ve μ_A ile μ_P sırasıyla gerçek ve tahmin edilen değerlerin ortalamasını göstermektedir.

3.6 Değişkenlerin Önem Parametrelerinin Tespiti: SHAP Yöntemi

SHAP (SHapley Additive exPlanations) yöntemi, makine öğrenimi modellerinin açıklanabilirliğini artırmak amacıyla oluşturulmuş, temeli oyun teorisine dayanan bir

yaklaşımıdır (Štrumbelj ve Kononenko, 2014). SHAP, model tahminlerinin hangi özelliklerden ne ölçüde etkilendiğini adil bir dağılım prensibine göre belirlemek için Shapley Değerleri kavramını kullanır (Fryer ve diğerleri, 2020; Shapley, 1953). Bu değerler, işbirlikçi oyun teorisinde bir oyuncunun toplam kazançtaki bireysel katkısını adil bir şekilde hesaplamak için geliştirilmiş bir ölçüttür. Makine öğrenimi bağlamında her bir özellik, oyundaki bir "oyuncu" olarak kabul edilir ve modelin toplam tahmini, bu oyuncuların birlikte elde ettiği kazancı temsil eder. SHAP, her özelliğin modele katkısını adil bir şekilde tahsis ederek, modelin kararlarının anlaşılmasını kolaylaştırır (Aromsuk, Kongnakorn ve Noonpakdee, 2023; Gebreyesus, Dalton, Nixon, De Chiara ve Chinnici, 2023; Merrick ve Taly, 2020). SHAP, model tahminini tüm özelliklerin katkılarının toplamı olarak ifade eder. Bir özelliğin tahmine yaptığı katkıyı belirlemek için, o özellik modele dahil edildiğinde ve edilmediğinde tahminin nasıl değiştiği hesaplanır (Van Den Broeck, Lykov, Suciu ve Schleich, 2022). Bu yaklaşım, her özelliğin farklı kombinasyonlardaki etkisini ortaya koyarak, özelliklerin modele yaptığı toplam katkıyı adil bir şekilde dağıtır. Model tahmini aşağıdaki formülizasyona göre doğrusal olarak ayrıştırılır:

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i$$

Burada, M , toplam özellik sayısını gösterirken, ϕ_0 ve ϕ_i değerleri sırasıyla temel değer (genellikle ortalama tahmini) ve her bir özelliğin modele yaptığı bireysel katkıyı ifade eden Shapley değerlerini temsil etmektedir. Bu ϕ_i değerleri aşağıdaki eşlilikle belirlenebilir (Molnar, 2022):

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)]$$

Burada, N tüm özelliklerin kümesini, S , i özelliği hariç olmak üzere özelliklerin alt kümesini, $v(S)$, S kümesindeki özelliklerle modelin tahmin sonucunu ve $v(S \cup \{i\}) - v(S)$, i özelliğinin modele eklenmesiyle tahminde oluşan değişikliği göstermektedir. Bu formülasyon, her özelliğin tüm olası alt kümelerdeki katkısının ağırlıklı ortalamasını alarak adil bir dağılım sağlar. Böylece, her özelliğin model üzerindeki etkisi kapsamlı bir şekilde değerlendirilir.

Pratik uygulamalarda SHAP, modelin hangi özelliklere dayandığını görselleştirerek model açıklanabilirliğini artırır. Bu da model tahminlerindeki hataların nedenlerini anlamaya ve düzeltmeye olanak tanır. Ayrıca, algoritmaların tarafsızlığını ve adaletini değerlendirmek için kullanılabilir, böylece adil karar verme süreçlerine katkı sağlar. Özellik seçimi aşamasında ise daha anlamlı ve etkili özelliklerin belirlenmesine yardımcı olarak model performansını artırır (M. Chen ve diğerleri, 2024; Rodríguez-Pérez ve Bajorath, 2020).

3.7 Değişkenlerin Benzerliklerinin Tespiti: Jaccard İndeksi

Jaccard benzerlik katsayısı olarak da bilinen Jaccard endeksi, iki kümenin ne kadar benzer olduğunu değerlendirmek için kullanılan istatistiksel bir tekniktir (Chung, Miasojedow, Startek ve Gambin, 2019; Jaccard, 1901). Kümelerin kesişiminin büyüklüğünün, birleşimlerinin büyüklüğüne oranı olarak tanımlanır. Matematiksel olarak, farklı sayıda elemana sahip iki A ve B kümesi için Jaccard endeksi $J(A,B)$ aşağıdaki gibi hesaplanır:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Burada, $|A \cap B|$ A ve B kümelerinin her ikisi tarafından paylaşılan ortak eleman sayısını ifade ederken, $|A \cup B|$ A ve B kümelerindeki elemanların toplam sayısını ifade eder. Jaccard endeksi 0 ile 1 arasında değişen bir aralık ölçeğidir. 0 puan, kümelerin hiç örtüşmediğini ve tamamen farklı olduklarını gösterirken, 1 puan kümelerin aynı olduğunu gösterir (Verma ve Aggarwal, 2020). Jaccard ölçütünün yorumlanması genellikle aşağıdaki şekilde yapılır:

- **Çok Düşük** [0.00 - 0.20]: Jaccard ölçütü 0.00 ile 0.20 arasında olduğunda, kümeler arasında çok düşük bir benzerlik olduğu söylenebilir. Bu durum genellikle rastgele veya çok az örtüşen veri kümelerini ifade eder.
- **Düşük** [0.20 - 0.40]: Endeksin bu aralığı, kümeler arasında az miktarda örtüşme olduğunu ifade eder. Ancak benzerlik düşük düzeydedir.
- **Orta**: [0.40 - 0.60]: Orta seviyedeki benzerlik, iki küme arasında daha anlamlı bir örtüşme olduğunu gösterir. Ancak tam bir uyumun bu aralıkta sağlanabildiği söylenemez.

- **Yüksek** [0.60 - 0.80]: Endeksin bu aralığı, kümelerin önemli ölçüde örtüştüğünü ve yüksek düzeyde benzerlik olduğunu ifade eder. Bu, iki veri kümesinin büyük oranda ortak elemanlara sahip olduğunu gösterir.

- **Çok Yüksek** [0.80 - 1.00]: Bu aralık, iki kümenin neredeyse tamamen örtüştüğünü ya da birebir aynı olduğunu ifade eder. 1.00 değeri, tam benzerliği (kümelerin tamamen aynı olduğunu) gösterir.

Jaccard endeksinin analizi, özellikle veri analizi, kümelenme ve desen tanıma gibi disiplinlerde büyük öneme sahiptir. Bu ölçüt, benzerlik arayışında olan veya kümeler arasındaki ilişkileri değerlendirmeye yönelik analitik kararlar ve stratejiler geliştirmek isteyenler için faydalı olabilir (Verma ve Aggarwal, 2020). Jaccard Endeksi, eski bir karşılaştırma ölçütü kriteri olmasına rağmen, günümüzde birçok farklı disiplinde ve çalışmada güvenilir bir benzerlik kıyaslaması sunduğu için kullanılmaya devam etmektedir. Örnek vermek gerekirse, biyoloji ve ekolojide tür çeşitliliği ve habitat benzerliği analizlerinde (Chung ve diğerleri, 2019), bilgi teknolojileri ve veri madenciliğinde öneri sistemleri veya metin benzerliği hesaplamalarında (Lee, 2017), tıp ve epidemiolojide hastalık kümelerinin analizinde (Salvatore ve diğerleri, 2020), sosyal bilimlerde ise anket verilerindeki benzer yanıtları gruplamada uygulanmaktadır (Woodhams, Grant ve Price, 2007). Jaccard endeksinin benzerlik karşılaştırma sorunlarında geniş çapta kullanımındaki iki temel gerekçe; yöntemin sunduğu basitlik ve anlaşılabilirlik ile iki kümenin ne ölçüde benzer veya farklı olduğunu açık bir şekilde ifade edebilmesidir.

3.8 Yazılım/Donanım Kaynakları ve Programlama

Veri analizi ve modelleme süreçlerini kapsayan bu araştırmanın her aşaması R programlama dili ve ilgili kütüphaneleri kullanılarak gerçekleştirilmiştir (Team, 2024). Veri analizi ve modelleme süreçlerini kapsayan bu çalışmanın tüm aşamaları R programlama dili ve ona bağlı kütüphaneler kullanılarak gerçekleştirilmiştir. RF algoritmalarının uygulanması için 'ranger' (Wright ve Ziegler, 2017) paketi, XGBoost modelleri için 'xgboost' (T. Chen ve Guestrin, 2016), SVR modelleri için 'kernlab' (Karatzoglou, Smola, Hornik ve Zeileis, 2004), ve BRNN için 'brnn' (Rodriguez ve Gianola, 2022) paketleri kullanılmıştır. Genel makine öğrenimi modelleme işlevlerinin entegrasyonu ve model seçimi 'caret' (Kuhn, 2019; Kuhn ve diğerleri, 2020) paketi aracılığıyla gerçekleştirilmiştir.

Veri kümesinin etkili bir şekilde yönetilmesini sağlayan veri manipölasyonu ve bölümlleme görevleri “dplyr” (Wickham, François, Henry, Müller ve Vaughan, 2023) paketi ile yapılmıştır. Elde edilen verilerin görselleştirilmesi ve analiz sonuçlarının grafiksel sunumu için “openair” (Carslaw ve Ropkins, 2012) ve “ggplot2” (Wickham, 2016) paketleri kullanılmıştır. Modelin yorumlanabilirliğini artırmak için SHAP (SHapley Additive exPlanations) değeri “fastshap” (Greenwell, 2024) paketi kullanılarak hesaplanmıştır. Ayrıca, kayıp verilerin işlenmesi için “mice” (van Buuren ve Groothuis-Oudshoorn, 2011) paketi uygulanmıştır. Her bir paket, bu araştırmanın bilimsel bütünlüğünü ve tekrarlanabilirliğini korumak için dikkatlice ayarlanmıştır. Hesaplamalar aşağıdaki özelliklere sahip bir bilgisayarda paralel işleme kullanılarak gerçekleştirilmiştir: İşlemci: Intel® Core™ i9-14900HX (5,80 GHz'e kadar, 24 fiziksel çekirdek), İşletim Sistemi: Windows11 Pro High-End Standart Sürüm, Grafik Kartı: RTX4090, GDDR6 16GB, Grafik Kartı Watt Gücü: 150W + 25W (Dynamic Boost ile), Bellek: DDR V 64GB (32GB*2, 5200MHz).

DÖRDÜNCÜ BÖLÜM

BULGULAR VE TARTIŞMA

4.1 TR1 Bölgesi PISA-2015 Bulguları

4.1.1 TR1 Bölgesi Hakkında Temel Bilgiler

2015 PISA uygulamasında Türkiye örneklemini 12 bölgeye ayırmıştır. 2015 PISA uygulamasında Türkiye örnekleminin birinci bölgesi TR1 şeklinde gösterilmiştir. TR1 bölgesi Marmara bölgesinde bulunan İstanbul ilinden katılan öğrencileri kapsamaktadır. İstanbul ilinden katılan öğrenci sayısı 1070'tir. 2015 PISA uygulamasına katılan 13-15 yaş aralığındaki toplam öğrenci sayısının %18.15'inin İstanbul ilinden katıldığı görülmektedir. 2015 PISA uygulamasında Türkiye örneklemine bakıldığında 12 bölge içerisinde öğrenci sayısının en fazla olduğu bölgenin TR1 (İstanbul) bölgesi olduğu görülmektedir.

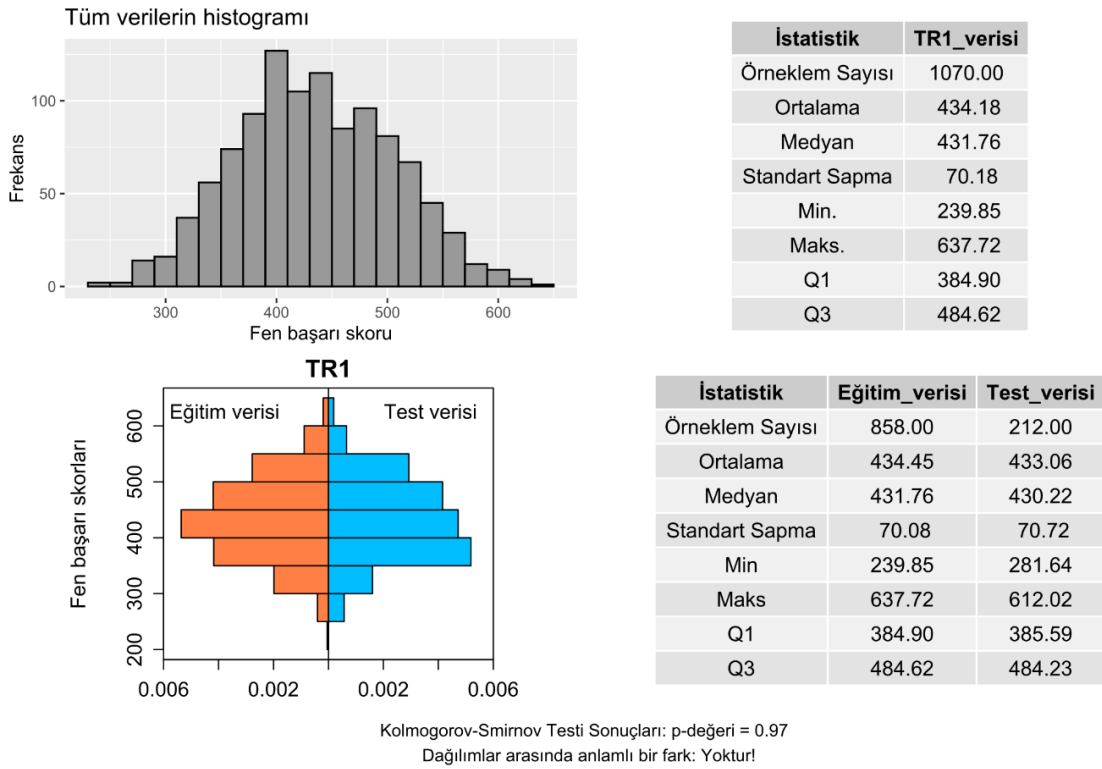
4.1.2 TR1 Bölgesinin Veri Yapısı

Şekil 4'te, Türkiye'nin TR1 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı skorlarına ait istatistiksel özeti sunulmaktadır. Şekil 4'ün üst panelinde, tüm veri setini temsil eden bir histogram ve istatistiksel özet bilgileri gösterilmektedir. Aynı şeklin alt panelinde ise, makine öğrenmesi modelinin oluşturulması ve modelin performansının değerlendirilmesi süreçlerinde kullanılacak olan eğitim ve test alt veri setlerine ait istatistiksel bilgiler bulunmaktadır. Bu alt veri setleri ana veri setinden rastgele bir şekilde ve literatüre uygun olarak %80 (Eğitim verisi) / %20 (Test verisi) oranında örneklenerek oluşturulmuştur (Gholamy ve diğerleri, 2018).

Şekil 4'ün üst panelindeki histograma göre, fen başarı skorlarının ortalama puan etrafında merkezi bir eğilimle dağıldığı görülmektedir. Bu TR1 bölgesindeki öğrencilerin (1070 kişi) Fen başarı puanlarının normal dağılıma uyduğunu belirtmektedir. Ayrıca bu bölge için tüm veri setine ait ortalama puan 434.18, medyan 431.76, standart sapma 70.18 olarak hesaplanmıştır. Bunlara ek olarak minimum ve maksimum puanlar sırasıyla 239.85 ve 637.72 olup, birinci çeyrek (Q1) 384.90 ve üçüncü çeyrek (Q3) 484.62 olarak belirlenmiştir.

Şekil 4'ün alt panelinde, eğitim veri setinin (858 kişi) ortalama puanı 434.45, medyanı 431.76 ve standart sapması 70.72 olarak tespit edilmiştir. Eğitim verisi için minimum ve maksimum puanlar sırasıyla 239.85 ve 637.72 olup, birinci çeyrek (Q1) 384.90 ve üçüncü çeyrek (Q3) 484.62 olarak belirlenmiştir. Bunların yanında Test veri setinin (212 kişi),

ortalama puanı 433.06, medyanı 430.22 ve standart sapması 70.08 olarak tespit edilmiştir. Test verisi için minimum ve maksimum puanlar sırasıyla 281.64 ve 612.02 olup, birinci çeyrek (Q1) 385.59 ve üçüncü çeyrek (Q3) 484.23 olarak belirlenmiştir. Bu veri seti için rasgele bir şekilde oluşturulan eğitim ve test alt veri setlerine ait dağılımlarının birbirlerinden anlamlı bir şekilde farklı olmadığı Kolmogorov-Smirnov testi sonucuna göre ($p > 0.05$) tespit edilmiştir.



Şekil 4: Türkiye'nin TR1 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)

4.1.3 TR1 Bölgesinde Makine Öğrenmesi

Tablo 5, TR1 (İstanbul) verileri için oluşturulan makine öğrenme algoritmalarının eğitim verilerinin 10 katlı ve 10 tekrarlı çapraz doğrulama performans metriklerini ve optimize edilmiş Hiperparametre değerlerini göstermektedir. Bu tabloda elde edilen bulgular; XGBoost, RF, SVR ve BRNN algoritmaları ile Kısım 3.3'te oluşturulan makine öğrenme modellerin 10 tekrarlı ve 10 katlı çapraz doğrulama tahmin performansları elde edilmiştir. Bu tablonun ayrıntılı sonuçları ve kısaltmalar için gerekli açıklamalar EK-1

içerisinde sunulmuştur. İncelenen ve oluşturulan makine öğrenmesi modellerin başarı performansları RMSE, R^2 ve MAE metrikleri üzerinden değerlendirilmiş ve XGBoost modelinin en düşük RMSE (46.03), en yüksek R^2 (0.57) ve düşük MAE (36.64) değerleri ile eğitim verileri için en iyi performansı gösterdiği tespit edilmiştir. Buna karşın SVR, RF ve BRNN modelleri sırasıyla 46.39, 47.57 ve 47.72 RMSE değerleri ile XGBoost algoritmasının ardından sıralanmışlardır. Bu veriler, TR1 bölgesinin eğitim verileri için XGBoost modelinin Fen okuryazarlık puanlarının tahmininde diğer modellere göre daha iyi bir öğrenme ortaya koyduğunu göstermektedir.

Tablo 5: TR1 (İstanbul) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri (Bu tablo için ayrıntılar EK-1 de sunulmuştur)

Model	RMSE	R^2	MAE	Hiperparametre değerleri
XGBoost	46.03	0.57	36.64	nrounds 282.00000000 max_depth 2.00000000 eta 0.05379505 gamma 8.64185906 colsample_bytree 0.48235216 min_child_weight 17.00000000 subsample 0.35974788
SVR	46.39	0.56	36.94	sigma 0.005336266 C 0.560407785
RF	47.57	0.56	37.84	mtry "75" splitrule "extratrees" min.node.size "1"
BRNN	47.72	0.54	38.11	neurons 1

Tablo 6' da TR1 bölgesine ait verileri için oluşturulan makine öğrenme algoritmalarının test verisi performans göstergeleri detaylı bir şekilde sunulmaktadır. Bu tabloya göre, XGBoost modelinin RMSE değeri 43.78, R^2 değeri 0.62 ve MAE değeri 34.75 olarak belirlenmiş ve diğer modellerden daha üstün bir performans sergilediği tespit edilmiştir. Ayrıca, XGBoost modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.748, güven aralığının alt sınırı 0.690 ve üst sınırı 0.796 ile yüksek bir uyumu gösterdiği belirlenmiştir. Buna karşın ikinci en iyi performans gösteren SVR modelinde, RMSE değeri 46.77, R^2 değeri 0.56 ve MAE değeri 36.65 ile XGBoost'a göre daha düşük olup, uyum korelasyon katsayısı merkezi tahmini 0.710 (lwr.ci: 0.645- upr.ci: 0.765) olarak hesaplanmıştır. BRNN ve RF modellerinde ise sırasıyla RMSE değerleri 47.51 ve 47.98, R^2

değerleri 0.55 ve 0.56, MAE değerleri 37.77 ve 38.25 olarak belirlenmiştir. BRNN modelinin uyum korelasyon katsayısı merkezi tahmini 0.715 (lwr.ci: 0.648- upr.ci: 0.771) iken, RF modelinin uyum korelasyon katsayısı merkezi tahmini 0.661 (lwr.ci: 0.597- upr.ci: 0.717) olarak hesaplanmıştır. Bu sonuçlar, XGBoost modelinin en düşük hata oranları ve en yüksek açıklama gücü ile diğer modellerden daha üstün performans sergilediğini ortaya koymaktadır.

Tablo 6: TR1 (İstanbul) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri

Model	RMSE	R ²	MAE	ρ_c
XGBoost	43.78	0.62	34.75	XGBoost.est 0.7477969 XGBoost.lwr.ci 0.6902296 XGBoost.upr.ci 0.7959574
SVR	46.77	0.56	36.65	SVR.est 0.7103163 SVR.lwr.ci 0.6453577 SVR.upr.ci 0.7650755
BRNN	47.51	0.55	37.77	BRNN.est 0.7150375 BRNN.lwr.ci 0.6478956 BRNN.upr.ci 0.7711542
RF	47.98	0.56	38.25	RF.est 0.6611722 RF.lwr.ci 0.5972945 RF.upr.ci 0.7167146

ρ_c = Uyum Korelasyon Katsayısı (Concordance Correlation Coefficient)

est = ρ_c için merkezi tahmin edilen değeri

lwr.ci = güven aralığının alt sınırını (lower confidence interval)

upr.ci = güven aralığının üst sınırını (upper confidence interval)

Tablo 5 ve Tablo 6 verilerine göre yapılan genel değerlendirmede, XGBoost modelinin hem eğitim hem de test verilerinde en yüksek performansı sergilediği görülmektedir. Bu durum, modelin aşırı uyum (overfitting) yapmadığını ve genelleştirme kabiliyetinin güçlü olduğunu göstermektedir. Ayrıca test verilerinin uyum korelasyon katsayısının (ρ_c) yüksek olması, modelin tahminlerinin gerçek verilerle yüksek uyum gösterdiğini belirtmektedir. Bu, XGBoost modelinin hem eğitim hem de test verilerinde tutarlı ve güvenilir bir performans sergilediğinin kanıtıdır.

Diğer yandan SVR, BRNN ve RF modelleri ise eğitim ve test verileri arasında benzer hata oranlarına sahip olsalar da performansları XGBoost modellerine kıyasla daha düşük seviyelerde kalmıştır. Bu durum SVR, BRNN ve RF modellerinin genelleştirme kabiliyetinin sınırlı olduğunu ve yeni veriler üzerinde daha düşük performans sergilediğini

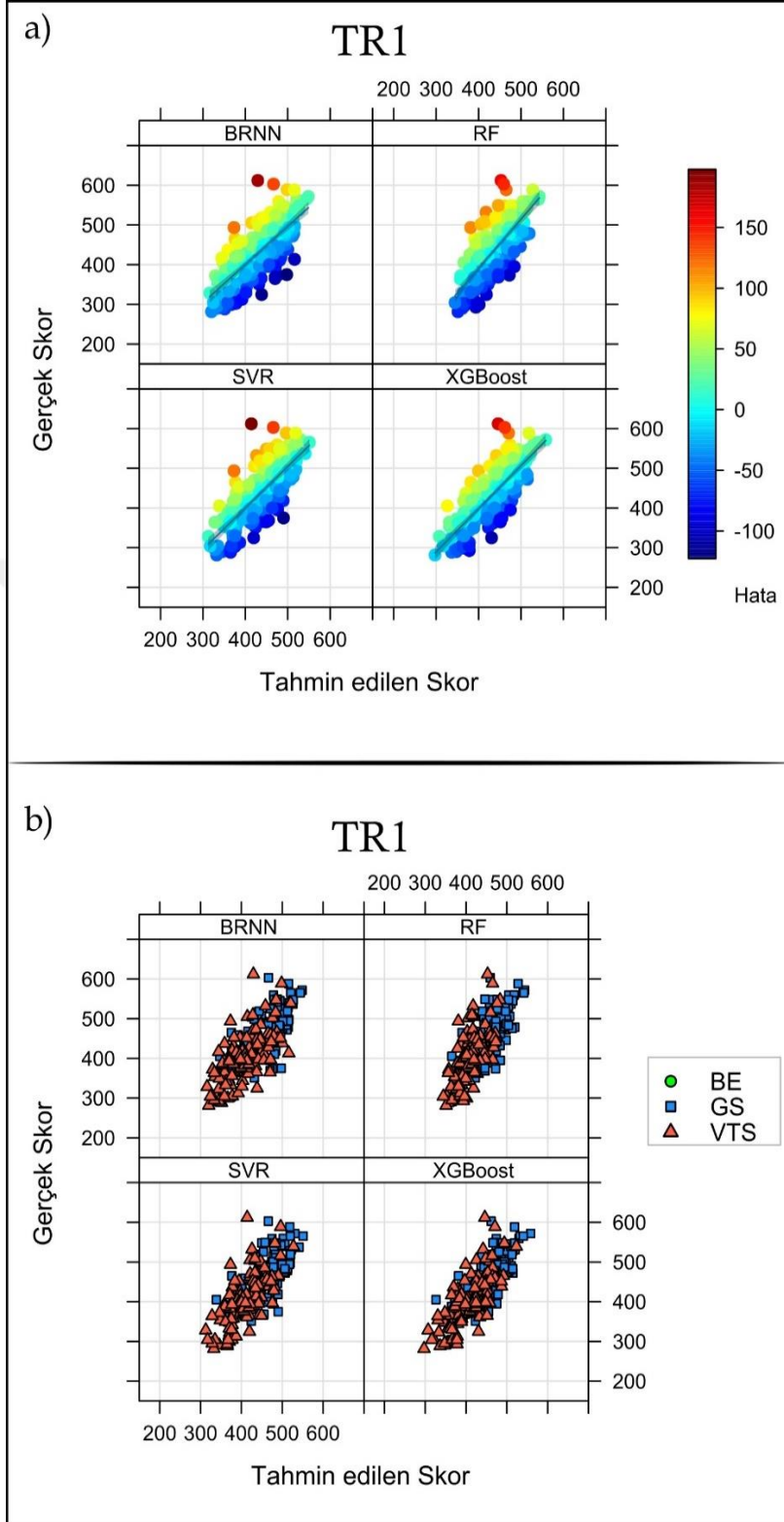
göstermektedir. Her üç modelin de ρ_c değerleri, tahminlerinin gerçek verilerle makul bir uyum gösterdiğini belirtmektedir.

Şekil 5, Tablo 6 da elde edilen performans metriklerine ait çapraz doğrulama diyagramlarını tanımlamaktadır. Şekil 5a'da tüm modellerin hata dağılımlarının regresyon çizgisi boyunca ve yakınlarında dağıldığı görülmektedir. Bu durum tahmin hatalarının dağılımının normal yapıda olduğunu ve hata ortalamasının da sıfıra yakın olduğunu göstermektedir. Bu sonuçlar Tablo 6 da elde edilen bulgular ile uyumludur. Bunlarla birlikte XGBoost modelinin yüksek tahmin performansı sergilediği, Şekil 5a' daki çapraz doğrulama diyagramının hata dağılımlarının regresyon çizgisi etrafında sıkı şekilde toplanmasından dolayı daha açık bir şekilde görülmektedir.

Şekil 5b, Aynı test verileri için modellerin farklı okul türlerine (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim) ait tahmin edilen ve gerçek skorlarının karşılaştırmasını göstermektedir. Burada özellikle tanımlayıcı regresyon çizgisinden sapan ve modellerin tahmin etme performansının düşmesine neden olan okul grubunun Mesleki ve Teknik Ortaöğretim (VTS) okul türü olduğu dikkat çekmektedir. Bununla birlikte XGBoost, her tüm okul türlerinde de tahmin edilen skorların gerçek skorlarla en uyumlu olduğu model olarak öne çıkmaktadır. Ayrıca TR1 bölgesindeki Fen başarı skorları için genel ortaokul (GS) başarı düzeyinin VTS okul türüne göre belirgin şekilde artış göstermiş olması dikkat çekicidir. Özellikle okul türünün modellerin oluşturulmasında bir değişken olarak kullanılmamış olmasına rağmen, makine öğrenme algoritmasının bu başarı farklılıklarını diğer değişkenlerin yardımı ile tespit edebilmesi ve tahminlerini buna göre gerçekleştirmesi önemli bir başarı olduğu düşünülmektedir. Bu bulgular, XGBoost modelinin eğitim verileri üzerindeki güçlü performansını ve genelleştirme kabiliyetini destekler niteliktedir. XGBoost'un, farklı okul türlerindeki öğrencilerin performansını doğru bir şekilde tahmin edebilmesi, eğitim politikaları ve müdahalelerinin planlanmasında doğru tahmin skorları önerebileceğini ve bu konuda değerli bir araç olarak kullanılabileceğini göstermektedir. Özellikle Mesleki ve Teknik Ortaöğretim (VTS) öğrencileri için tahminlerdeki sapmalara rağmen, XGBoost'un genel olarak diğer modellere göre daha yüksek doğrulukla sonuçlar vermesi, bu modelin TR1 bölgesi için geniş bir yelpazede güvenilir tahminler oluşturabileceğini göstermektedir.

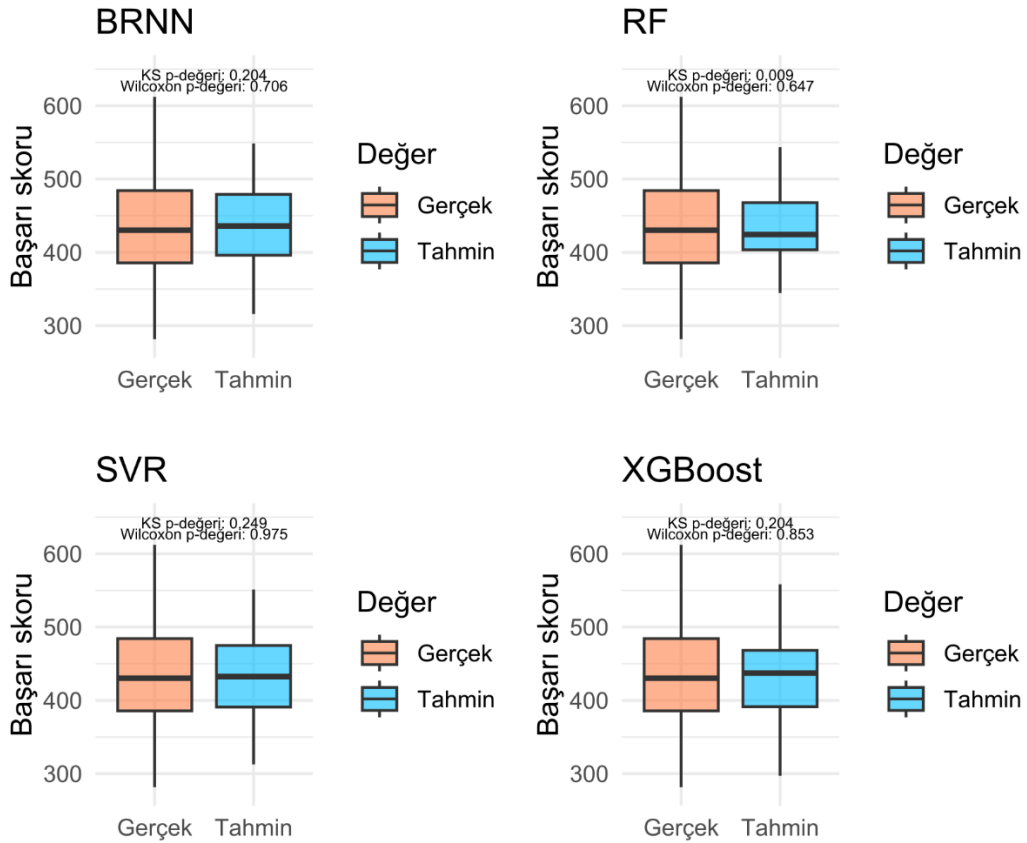
Şekil 6, farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost) tahmin ettiği başarı skorları ile gerçek başarı skorlarının karşılaştırmalı kutu grafiğini sunmaktadır. Şekil 5’te sunulan çapraz doğrulama bulgularına göre oluşturulmuş bu grafikler, modellerin tahmin performansını ve gerçek verilerle uyumunu değerlendirmek için önemli bir yaklaşım sağlamaktadır. Buna göre RF hariç tüm modellerin tahmin skorlarının hem Kolmogorov - Simirnov testi (KS) hem de Wilcoxon sıralama toplamı testine göre gerçek skorların genel dağılımlarından ve merkezi eğilimlerinden anlamlı ölçüde bir farklılığı ($p > 0.05$) olmadığı tespit edilmiştir. Bu durum makine öğrenmesi modellerinin tahmin edilen ve gerçek başarı skorları arasındaki uyumun yüksek olduğunu ortaya koymaktadır.





Şekil 5: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)

Şekil 6, en yüksek performans seviyesi gösteren XGBoost algoritması açısından değerlendirildiğinde, kutu grafiklerinin gerçek ve tahmin skorları için temsil ettiği ortalama ve medyan değerlerinin neredeyse örtüştüğü ve dağılımların (kutu dışındaki uç örnekler için yeterli seviyelerde olmasa da) Q1-Q3 aralığı için oldukça yakın olduğunu görülmektedir. Şekil 6'daki bulgulara göre tahmin edilen ve gerçek skorların genel dağılımlarının (KS p-değeri 0.204) ve merkezi eğilimlerinin (Wilcoxon p-değeri 0.853), istatistiksel olarak birbirlerinden farklı olmadığı belirlenmiştir. Bu durum, XGBoost algoritmasının TR1 verileri üzerinde en yüksek doğruluk ve güvenilirliği sunmakla birlikte hem bireysel gerçek skorları hem de gerçek skorlara ait dağılım yapısını doğru şekilde tahmin edebilme yeteneğini ortaya koymaktadır.



Şekil 6: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.

4.1.4 TR1 Bölgesi İçin Değişkenlerin Önem Seviyesi

PISA 2015 Türkiye örnekleme TR1 bölgesi Fen okuryazarlık alanı için en iyi tahmin performans gösteren XGBoost algoritmasına göre SHAP (SHapley Additive exPlanations) değerleri kullanılarak değişkenlerin önem parametreleri belirlenmiştir.

Sonuçlar her bir değişkenin ortalama SHAP değerlerinin mutlak değerine göre azalan sırasinda Tablo 7’de, sıralanmıştır. Bu sıralamaya göre diğer bölgelerdeki değişkenlerin karşılaştırması ise kısım 4.13’te Jaccard indeksine göre yapılacaktır. Burada model için önemli olan sıralı değişkenler üç önem grubuna ayrılmış ve bu işlem model için kullanılan toplam değişkenleri 3 eşit parçaya ayırarak gerçekleştirilmiştir. Burada yüksek derecede önemli değişkenler 1. Grup (ilk 33 tanesi), orta derecede önemli değişkenler 2. Grup (34 - 66) ve düşük derecede önemli değişkenler 3. Grup (67 - 99) olarak etiketlenmiştir.

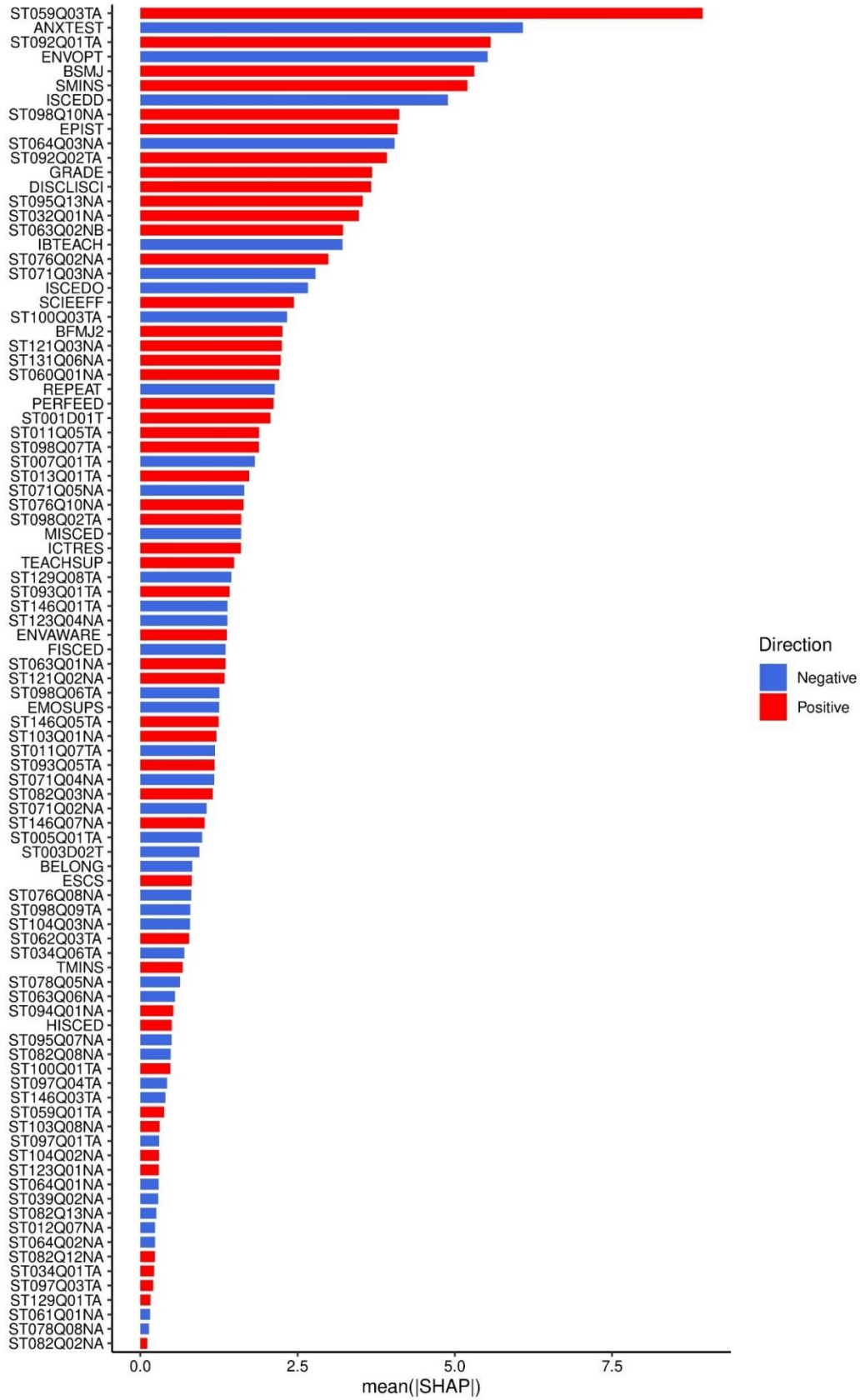
Ayrıca Şekil 7 ile Tablo 7’de özetlenen değişkenlerin ortalama SHAP değerleri görselleştirilmiştir. Bu şekil her bir tahmin edici değişkenin model tahminlerine olan ortalama etkisini ve yönünü (pozitif ve negatif katkıları) vurgulamaktadır. Şekil 8 her bir değişkenin test veri setindeki her bir gözlemdeki etkisini göstermektedir. Bu şekil içerisinde her bir nokta, bir veri noktasının bir özelliği için SHAP değerini temsil eder. Bu şeklin sağ tarafında pozitif SHAP değerleri, sol tarafında ise negatif SHAP değerleri bulunmaktadır. Tahmin edici değişkenlerin model çıktısına olan etkisi, bu SHAP değerlerinin büyüklüğüne göre belirlenir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Tablo7, Şekil 7 ve 8’deki bulgulara göre, TR1 bölgesi için model tahmininde önem seviyesi yüksek ilk 10 değişken, parantez içleri SHAP değerlerini ve model katkı yönlerini göstermek üzere, şu şekilde sıralanmıştır: ST059Q03TA (8.939, pozitif), ANXTEST (6.083, negatif), ST092Q01TA (5.569, pozitif), ENVOPT (5.523, negatif), BSMJ (5.310, pozitif), SMINS (5.200, pozitif), ISCEDD (4.890, negatif), ST098Q10NA (4.116, pozitif), EPIST (4.090, pozitif), ST064Q03NA (4.043, negatif).

Şekil 4’te kırmızı ve mavi renkler, sırasıyla pozitif ve negatif yönlü etkileri temsil etmektedir. Bu, modelin tahminlerinin hangi özellikler tarafından ne kadar ve hangi yönde etkilendiğini daha net bir şekilde anlaşılmasına yardımcı olmaktadır. Buna göre negatif özellikteki değişkenlerin (ANXTEST, ENVOPT, ISCEDD, ST064Q03NA) değerlerindeki artış öğrencilerin Fen başarı skorlarını düşürmekte buna karşı pozitif özellikteki

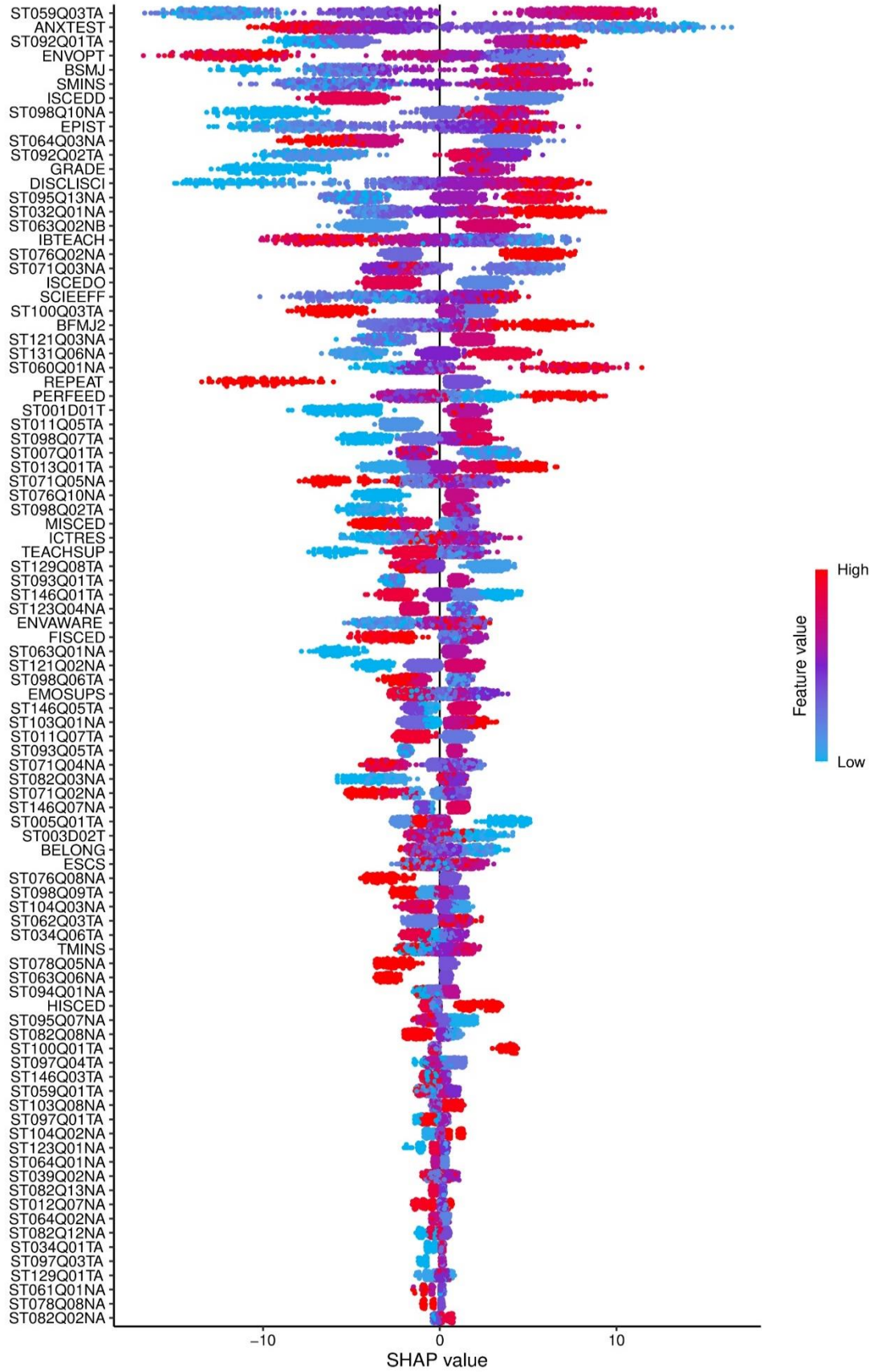
değişkenlerin değerlerindeki artış ise öğrencilerin Fen başarı skorlarını arttırıcı yönde etki etmektedir.

Tablo 7: TR1 (İstanbul) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin ortalama SHAP değerlerine göre önem seviyelerinin sıralanması.

ID	Değişken	SHAP değeri	Önem Grubu	ID	Değişken	SHAP değeri	Önem Grubu
1	ST059Q03TA	8.939	Grup 1	55	ST082Q03NA	1.152	Grup 2
2	ANXTEST	6.083	Grup 1	56	ST071Q02NA	1.053	Grup 2
3	ST092Q01TA	5.569	Grup 1	57	ST146Q07NA	1.024	Grup 2
4	ENVOPT	5.523	Grup 1	58	ST005Q01TA	0.984	Grup 2
5	BSMJ	5.310	Grup 1	59	ST003D02T	0.940	Grup 2
6	SMINS	5.200	Grup 1	60	BELONG	0.826	Grup 2
7	ISCEDD	4.890	Grup 1	61	ESCS	0.819	Grup 2
8	ST098Q10NA	4.116	Grup 1	62	ST076Q08NA	0.810	Grup 2
9	EPIST	4.090	Grup 1	63	ST098Q09TA	0.795	Grup 2
10	ST064Q03NA	4.043	Grup 1	64	ST104Q03NA	0.790	Grup 2
11	ST092Q02TA	3.921	Grup 1	65	ST062Q03TA	0.776	Grup 2
12	GRADE	3.684	Grup 1	66	ST034Q06TA	0.700	Grup 2
13	DISCLISCI	3.670	Grup 1	67	TMIN	0.675	Grup 3
14	ST095Q13NA	3.536	Grup 1	68	ST078Q05NA	0.633	Grup 3
15	ST032Q01NA	3.476	Grup 1	69	ST063Q06NA	0.553	Grup 3
16	ST063Q02NB	3.223	Grup 1	70	ST094Q01NA	0.525	Grup 3
17	IBTEACH	3.214	Grup 1	71	HISCED	0.499	Grup 3
18	ST076Q02NA	2.990	Grup 1	72	ST095Q07NA	0.496	Grup 3
19	ST071Q03NA	2.784	Grup 1	73	ST082Q08NA	0.484	Grup 3
20	ISCEDO	2.667	Grup 1	74	ST100Q01TA	0.479	Grup 3
21	SCIEEFF	2.443	Grup 1	75	ST097Q04TA	0.427	Grup 3
22	ST100Q03TA	2.334	Grup 1	76	ST146Q03TA	0.399	Grup 3
23	BFMJ2	2.261	Grup 1	77	ST059Q01TA	0.381	Grup 3
24	ST121Q03NA	2.250	Grup 1	78	ST103Q08NA	0.306	Grup 3
25	ST131Q06NA	2.233	Grup 1	79	ST097Q01TA	0.302	Grup 3
26	ST060Q01NA	2.210	Grup 1	80	ST104Q02NA	0.298	Grup 3
27	REPEAT	2.138	Grup 1	81	ST123Q01NA	0.294	Grup 3
28	PERFEED	2.120	Grup 1	82	ST064Q01NA	0.291	Grup 3
29	ST001D01T	2.070	Grup 1	83	ST039Q02NA	0.286	Grup 3
30	ST011Q05TA	1.890	Grup 1	84	ST082Q13NA	0.257	Grup 3
31	ST098Q07TA	1.888	Grup 1	85	ST012Q07NA	0.235	Grup 3
32	ST007Q01TA	1.821	Grup 1	86	ST064Q02NA	0.235	Grup 3
33	ST013Q01TA	1.733	Grup 1	87	ST082Q12NA	0.231	Grup 3
34	ST071Q05NA	1.655	Grup 2	88	ST034Q01TA	0.221	Grup 3
35	ST076Q10NA	1.643	Grup 2	89	ST097Q03TA	0.206	Grup 3
36	ST098Q02TA	1.604	Grup 2	90	ST129Q01TA	0.162	Grup 3
37	MISCED	1.603	Grup 2	91	ST061Q01NA	0.154	Grup 3
38	ICTRES	1.600	Grup 2	92	ST078Q08NA	0.138	Grup 3
39	TEACHSUP	1.491	Grup 2	93	ST082Q02NA	0.112	Grup 3
40	ST129Q08TA	1.448	Grup 2	94	ST093Q03TA	0.000	Grup 3
41	ST093Q01TA	1.422	Grup 2	95	ST063Q04NA	0.000	Grup 3
42	ST146Q01TA	1.386	Grup 2	96	ST098Q01TA	0.000	Grup 3
43	ST123Q04NA	1.382	Grup 2	97	ST104Q05NA	0.000	Grup 3
44	ENVAWARE	1.376	Grup 2	98	ST123Q02NA	0.000	Grup 3
45	FISCED	1.357	Grup 2	99	ST131Q03NA	0.000	Grup 3
46	ST063Q01NA	1.357	Grup 2				
47	ST121Q02NA	1.341	Grup 2				
48	ST098Q06TA	1.257	Grup 2				
49	EMOSUPS	1.254	Grup 2				
50	ST146Q05TA	1.247	Grup 2				
51	ST103Q01NA	1.208	Grup 2				
52	ST011Q07TA	1.189	Grup 2				
53	ST093Q05TA	1.181	Grup 2				
54	ST071Q04NA	1.176	Grup 2				



Şekil 7: TR1 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması



Şekil 8: TR1 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.

4.2 TR2 Bölgesi PISA-2015 Bulguları

4.2.1 TR2 Bölgesi Hakkında Temel Bilgiler

2015 PISA sınavında Türkiye örnekleminin ikinci bölgesi TR2 olarak adlandırılmıştır. TR2 bölgesi PISA bölgelerinden Batı Marmara bölgesinde bulunan illeri kapsamaktadır. Batı Marmara bölgesi Tekirdağ, Edirne, Kırklareli, Çanakkale ve Balıkesir’de bulunan 7. Sınıf ve üzeri sınıf düzeyinde öğrenim gören öğrencilerin başarı skorlarını göstermektedir. TR2 (Batı Marmara) bölgesinden 245 öğrencinin 2015 PISA sınavına katıldığı görülmektedir. PISA 2015 sınavına katılan 13 ve 15 yaş aralığında öğrencilerin %4.16’sının uygulamaya TR2 bölgesinden (Batı Marmara) katıldığını görmekteyiz.

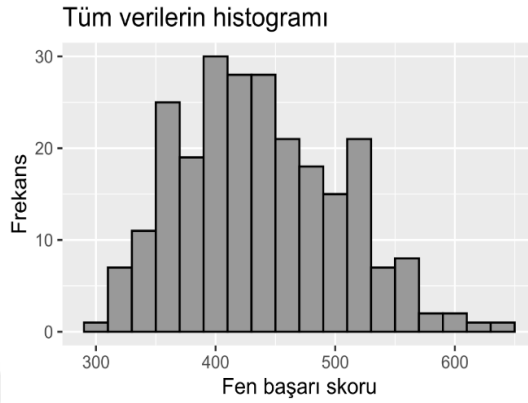
4.2.2 TR2 Bölgesinin Veri Yapısı

Şekil 9’da Türkiye’nin TR2 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özeti sunulmaktadır. Şekil 9’un üst panelinde, tüm veri setini temsil eden bir histogram ve istatistiksel özet bilgileri gösterilmektedir. Aynı şeklin alt panelinde ise, makine öğrenmesi modelinin oluşturulması ve modelin performansının değerlendirilmesi süreçlerinde kullanılacak olan eğitim ve test alt veri setlerine ait istatistiksel bilgiler bulunmaktadır. Bu alt veri setleri ana veri setinden rastgele bir şekilde ve literatüre uygun olarak %80 (Eğitim verisi) / %20 (Test verisi) oranında örneklenerek oluşturulmuştur (Gholamy ve diğerleri, 2018).

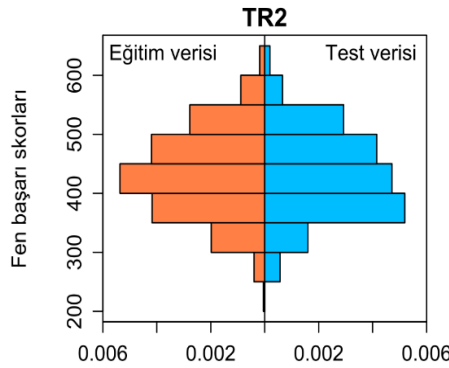
Şekil 9’un üst panelindeki histograma göre Fen başarı skorlarının ortalama puan etrafında merkezi bir eğilimle dağıldığı görülmektedir. Bu TR2 bölgesindeki öğrencilerin (245 öğrenci) Fen başarı skorlarının normal dağılıma uyduğunu belirtmektedir. Ayrıca bu bölge için tüm veri setine ait ortalama puan 437.42, medyan 433.61, standart sapma 65.77 olup, birinci çeyrek (Q1) 389.95 ve üçüncü çeyrek 484.74 olarak belirlenmiştir.

Şekil 9’un alt panelinde eğitim veri setinin (197 kişi) ortalama puanı 437.68, medyanı 433.61 ve standart sapması 64.90 olarak tespit edilmiştir. Eğitim verisi için minimum ve maksimum puanlar sırasıyla 319.30 ve 619.63 olup, birinci çeyrek (Q1) 389.95 ve üçüncü çeyrek 484,74 olarak belirlenmiştir. Bunların yanında Test veri setinin (48 kişi), ortalama puanı 436.39, medyanı 430.95 ve standart sapması 69.97 olarak tespit edilmiştir. Test verisi için minimum ve maksimum puanlar sırasıyla 306.35 ve 641.41 olup, birinci çeyrek (Q1) 388.11 ve üçüncü çeyrek 484.35 olarak belirlenmiştir. Bu veri seti için rasgele bir şekilde

oluşturulan eğitim ve test alt veri setine ait dağılımların birbirlerinden anlamlı bir şekilde farklı olmadığı Kolmogorov-Smirnov testi sonucuna göre ($p > 0.05$) tespit edilmiştir.



İstatistik	TR2_verisi
Örneklem Sayısı	245.00
Ortalama	437.42
Medyan	433.61
Standart Sapma	65.77
Min.	306.35
Maks.	641.41
Q1	389.95
Q3	484.74



İstatistik	Eğitim_verisi	Test_verisi
Örneklem Sayısı	197.00	48.00
Ortalama	437.68	436.39
Medyan	433.61	430.95
Standart Sapma	64.90	69.97
Min	319.30	306.35
Maks	619.63	641.41
Q1	389.95	388.11
Q3	484.74	484.35

Kolmogorov-Smirnov Testi Sonuçları: p-değeri = 0.78
Dağılımlar arasında anlamlı bir fark: Yoktur!

Şekil 9: Türkiye'nin TR2 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)

4.2.3 TR2 Bölgesinde Makine Öğrenme Modeli

Tablo 8, TR2 (Batı Marmara) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı ve 10 tekrarlı çapraz doğrulama performans metriklerini ve optimize edilmiş Hiperparametre değerlerini göstermektedir. Bu tabloda elde edilen bulgular, XGBoost, RF, SVR ve BRNN algoritmaları ile Kısım 3.3'te oluşturulan makine öğrenme modellerin 10 tekrarlı ve 10 katlı çapraz doğrulama tahmin performansları elde edilmiştir. Bu tablonun ayrıntılı sonuçları ve kısaltmalar için gerekli açıklamalar EK-1 içerisinde sunulmuştur. İncelenen ve oluşturulan makine öğrenmesi modellerin başarı performansları RMSE, R^2 ve MAE metrikleri üzerinden değerlendirilmiş ve XGBoost modelinin en düşük RMSE 46.77, en yüksek R^2 0.51 ve en düşük MAE 38.09 değeri ile

eğitim verileri için en iyi performansı gösterdiği tespit edilmiştir. Buna karşın SVR, RF ve BRNN modelleri sırasıyla 46.89, 47.59 ve 52.01 RMSE değerleri ile XGBoost algoritmasının ardından sıralanmışlardır. Bu veriler, TR2 bölgesinin eğitim verileri için XGBoost modelinin Fen okuryazarlık puanlarının tahmininde diğer modellere göre daha iyi bir öğrenme ortaya koyduğunu göstermektedir.

Tablo 8: TR2 (Batı Marmara) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri

Model	RMSE	R2	MAE	Hiperparametre değerleri
XGBoost	46.77	0.51	38.09	nrounds 777.00000000 max_depth 2.00000000 eta 0.03748243 gamma 3.64823660 colsample_bytree 0.46808052 min_child_weight 15.00000000 subsample 0.99495393
SVR	46.89	0.49	38.47	sigma 0.008288906 C 0.917046224
RF	47.59	0.48	39.69	mtry "84" splitrule "extratrees" min.node.size "10"
BRNN	52.01	0.39	42.19	neurons 2

Tablo 9'da TR2 bölgesine ait verileri için oluşturulan makine öğrenme algoritmalarının test verisi performans göstergeleri detaylı bir şekilde sunulmaktadır. Bu tabloya göre, XGBoost modelinin RMSE değeri 42.31, R^2 değeri 0.66 ve MAE değeri 34.05 olarak belirlenmiş ve diğer modellerden daha üstün bir performans sergilediği tespit edilmiştir. Ayrıca, XGBoost modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.733, güven aralığının alt sınırı 0.613 ve üst sınırı 0.821 ile yüksek bir uyumu gösterdiği belirlenmiştir. Buna karşın ikinci en iyi performans gösteren SVR modelinde, RMSE değeri 43.73, R^2 değeri 0.61 ve MAE değeri 34.05 ile XGBoost'a göre daha düşük olup, uyum korelasyon katsayısı merkezi tahmini 0.737 (lwr.ci: 0.598- upr.ci: 0.833) olarak hesaplanmıştır. RF ve BRNN modellerinde sırasıyla RMSE değerleri 45.08 ve 43.56, R^2 değerleri 0.62 ve 0.61, MAE değerleri 36.72 ve 32.69 olarak belirlenmiştir. RF modelinin uyum korelasyon katsayısı 0.683 (lwr.ci: 0.554- upr.ci: 0.780) iken BRNN modelinin uyum korelasyon katsayısı merkezi tahmini 0.750 (lwr.ci: 0.611 - upr.ci: 0.844) olarak

hesaplanmıştır. Bu sonuçlar XGBoost modelinin en düşük hata oranları ve en yüksek açıklama gücü ile diğer modellerden daha üstün performans sergilediğini ortaya koymaktadır.

Tablo 9: TR2 (Batı Marmara) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri

Model	RMSE	R^2	MAE	ρ_c
XGBoost	42.31	0.66	34.05	XGBoost.est 0.7335710 XGBoost.lwr.ci 0.6129560 XGBoost.upr.ci 0.8207752
SVR	43.73	0.61	34.99	SVR.est 0.7370038 SVR.lwr.ci 0.5982063 SVR.upr.ci 0.8328684
RF	45.08	0.62	36.72	RF.est 0.6833053 RF.lwr.ci 0.5538256 RF.upr.ci 0.7805176
BRNN	43.56	0.61	32.69	BRNN.est 0.7502889 BRNN.lwr.ci 0.6113308 BRNN.upr.ci 0.8443648

ρ_c = Uyum Korelasyon Katsayısı (Concordance Correlation Coefficient)

est = ρ_c için merkezi tahmin edilen değeri

lwr.ci = güven aralığının alt sınırını (lower confidence interval)

upr.ci = güven aralığının üst sınırını (upper confidence interval)

Tablo 8 ve Tablo 9 verilerine göre yapılan genel değerlendirmede, XGBoost modelinin hem eğitim hem de test verilerinde en yüksek performansı sergilediği görülmektedir. Bu durum, modelin aşırı uyum (overfitting) yapmadığını ve genelleştirme kabiliyetinin güçlü olduğunu göstermektedir. Ayrıca test verilerinin uyum korelasyon katsayısının (ρ_c) yüksek olması, modelin tahminlerinin gerçek verilerle yüksek uyum gösterdiğini belirtmektedir. Bu, XGBoost modelinin hem eğitim hem de test verilerinde tutarlı ve güvenilir bir performans sergilediğinin kanıtıdır.

Diğer yandan SVR, RF ve BRNN modelleri ise eğitim ve test verileri arasında benzer hata oranlarına sahip olsalar da performansları XGBoost modellerine kıyasla daha düşük seviyelerde kalmıştır. Bu durum SVR, BRNN ve RF modellerinin genelleştirme kabiliyetinin sınırlı olduğunu ve yeni veriler üzerinde daha düşük performans sergilediğini göstermektedir. Her üç modelin de ρ_c değerleri, tahminlerinin gerçek verilerle makul bir uyum gösterdiğini belirtmektedir.

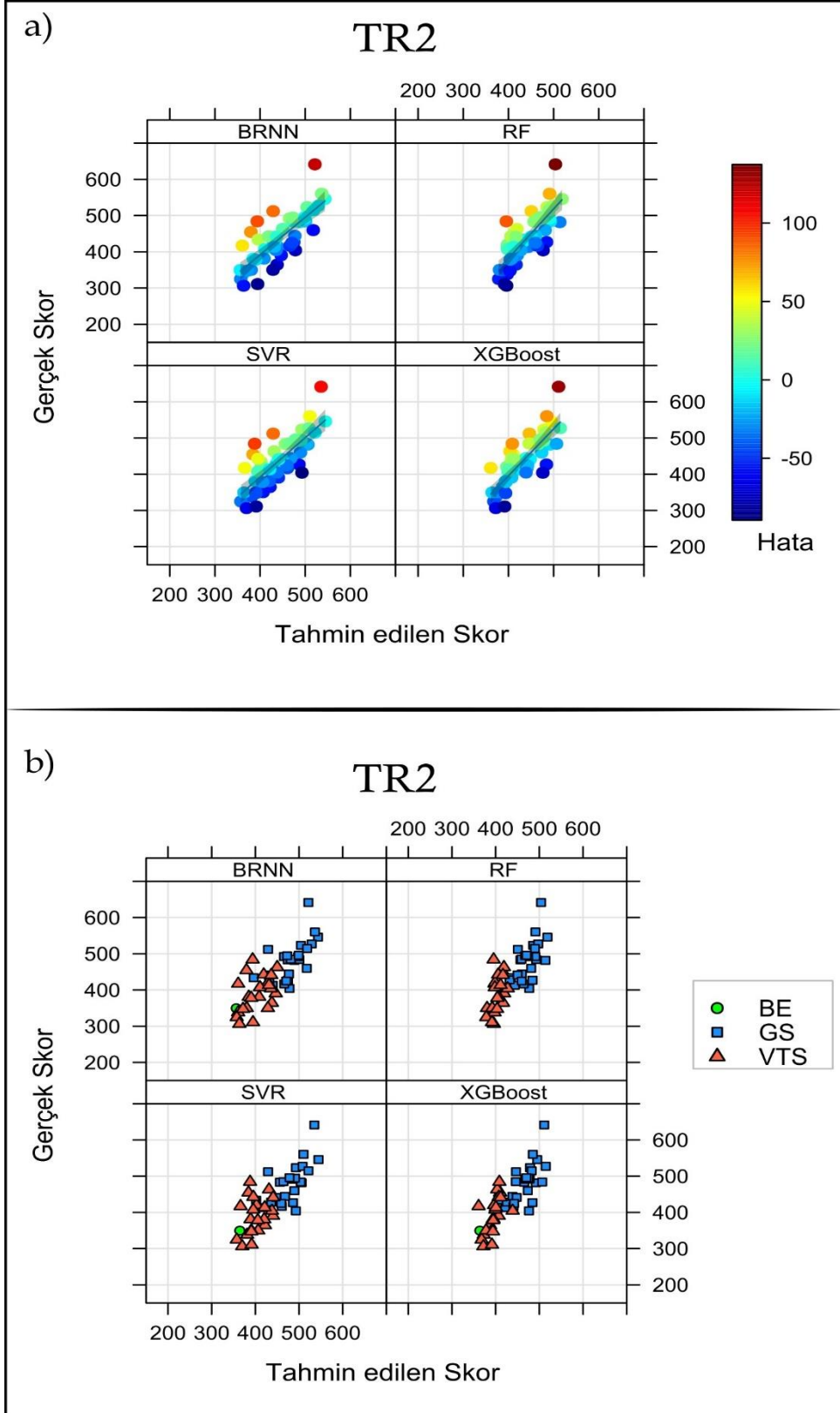
Şekil 10, Tablo 9 da elde edilen performans metriklerine ait çapraz doğrulama diyagramlarını tanımlamaktadır. Şekil 10a' da tüm modellerin hata dağılımlarının regresyon çizgisi boyunca ve yakınlarında dağıldığı görülmektedir. Bu durum tahmin hatalarının dağılımının normal yapıda olduğunu ve hata ortalamasının da sıfıra yakın olduğunu göstermektedir. Bu sonuçlar Tablo 9 ile elde edilen bulgular ile uyumludur. Bunlarla birlikte XGboost modelinin yüksek tahmin performansı sergilediği Şekil a'daki çapraz doğrulama diyagramının hata dağılımlarının regresyon çizgisi etrafında sıkı şekilde toplanmasından dolayı daha açık bir şekilde görülmektedir.

Şekil 10b, aynı test verileri için modellerin farklı okul türlerine (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim) ait tahmin edilen ve gerçek skorlarının karşılaştırmasını göstermektedir. Burada özellikle tanımlayıcı regresyon çizgisinden sapan ve modellerin tahmin etme performansının düşmesine neden olan okul grubunun Mesleki ve Teknik Ortaöğretim (VTS) okul türü olduğu dikkat çekmektedir. Bununla birlikte XGBoost, her tüm okul türlerinde de tahmin edilen skorların gerçek skorlarla en uyumlu olduğu model olarak öne çıkmaktadır. Ayrıca TR2 bölgesindeki Fen başarı skorları için genel ortaokul (GS) başarı düzeyinin VTS okul türüne göre belirgin şekilde artış göstermiş olması dikkat çekicidir. Özellikle okul türünün modellerin oluşturulmasında bir değişken olarak kullanılmamış olmasına rağmen, makine öğrenme algoritmasının bu başarı farklılıklarını diğer değişkenlerin yardımı ile tespit edebilmesi ve tahminlerini buna göre gerçekleştirmesi önemli bir başarı olduğu düşünülmektedir. Bu bulgular, XGBoost modelinin eğitim verileri üzerindeki güçlü performansını ve genelleştirme kabiliyetini destekler niteliktedir. XGBoost'un, farklı okul türlerindeki öğrencilerin performansını doğru bir şekilde tahmin edebilmesi, eğitim politikaları ve müdahalelerinin planlanmasında doğru tahmin skorları önerebileceğini ve bu konuda değerli bir araç olarak kullanılabileceğini göstermektedir. Özellikle Mesleki ve Teknik Ortaöğretim (VTS) öğrencileri için tahminlerdeki sapmalara rağmen, XGBoost'un genel olarak diğer modellere göre daha yüksek doğrulukla sonuçlar vermesi, bu modelin TR2 bölgesi için geniş bir yelpazede güvenilir tahminler oluşturabileceğini göstermektedir.

Şekil 11, farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost) tahmin ettiği başarı skorları ile gerçek başarı skorlarının karşılaştırmalı kutu grafiğini sunmaktadır. Şekil 10 da sunulan çapraz doğrulama bulgularına göre oluşturulmuş bu grafikler, modellerin tahmin performansını ve gerçek verilerle uyumunu değerlendirmek için önemli

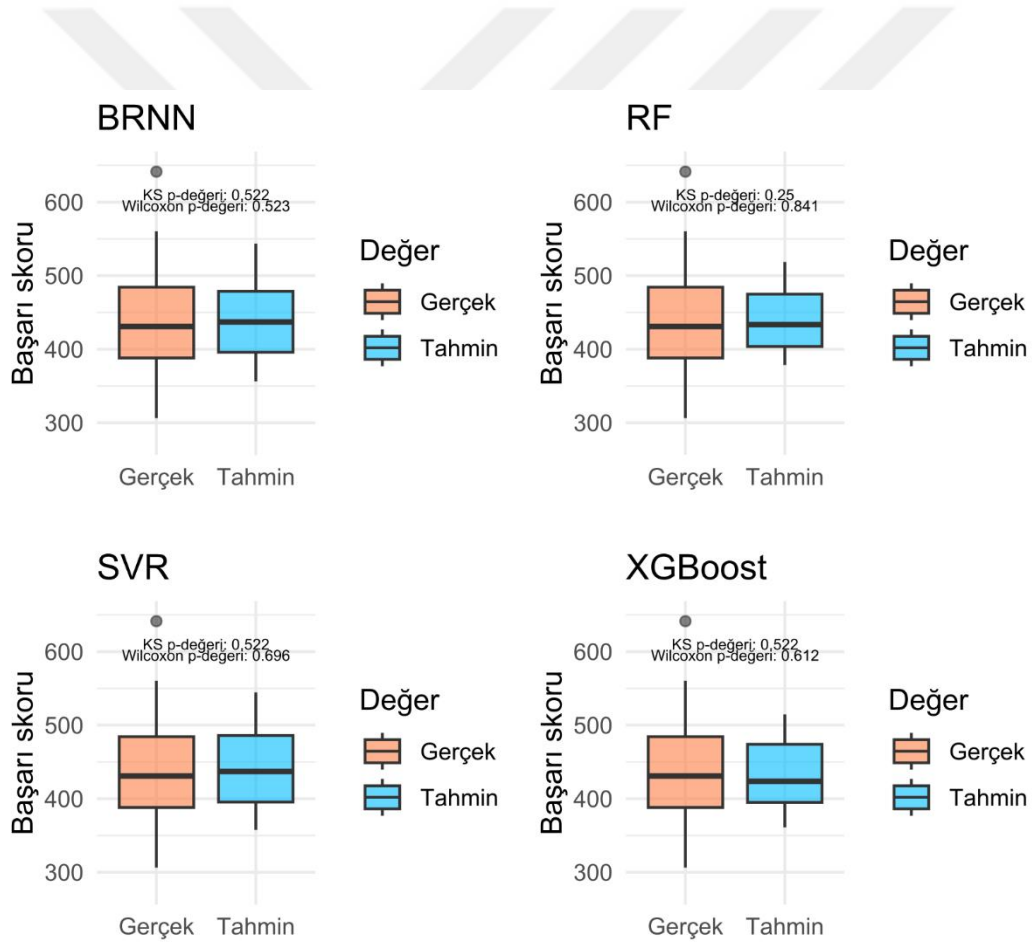
bir yaklaşımları sağlamaktadır. Buna göre tüm modellerin tahmin skorlarının hem Kolmogorov-Smirnov testi (KS) hem de Wilcoxon sıralama toplamı testine göre gerçek skorların genel dağılımlarından ve merkezi eğilimlerinden anlamlı ölçüde bir farklılığı ($p > 0.05$) olmadığı tespit edilmiştir. Bu durum makine öğrenmesi modellerinin tahmin edilen ve gerçek başarı skorları arasındaki uyumun yüksek olduğunu ortaya koymaktadır.





Şekil 10: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)

Şekil 11 en yüksek performans seviyesi gösteren XGBoost algoritması açısından değerlendirildiğinde, kutu grafiklerinin gerçek ve tahmin skorları için temsil ettiği ortalama ve medyan değerlerinin neredeyse örtüştüğü ve dağılımların (kutu dışındaki uç örnekler için yeterli seviyelerde olmasa da) Q1-Q3 aralığı için oldukça yakın olduğunu görülmektedir. Şekil 4'teki bulgulara göre tahmin edilen ve gerçek skorların genel dağılımlarının (KS p-değeri 0.522) ve merkezi eğilimlerinin (Wilcoxon p-değeri 0.612), istatistiksel olarak birbirlerinden farklı olmadığı belirlenmiştir. Bu durum, XGBoost algoritmasının TR2 verileri üzerinde en yüksek doğruluk ve güvenilirliği sunmakla birlikte hem bireysel gerçek skorları hem de gerçek skorlara ait dağılım yapısını doğru şekilde tahmin edebilme yeteneğini ortaya koymaktadır.



Şekil 11: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.

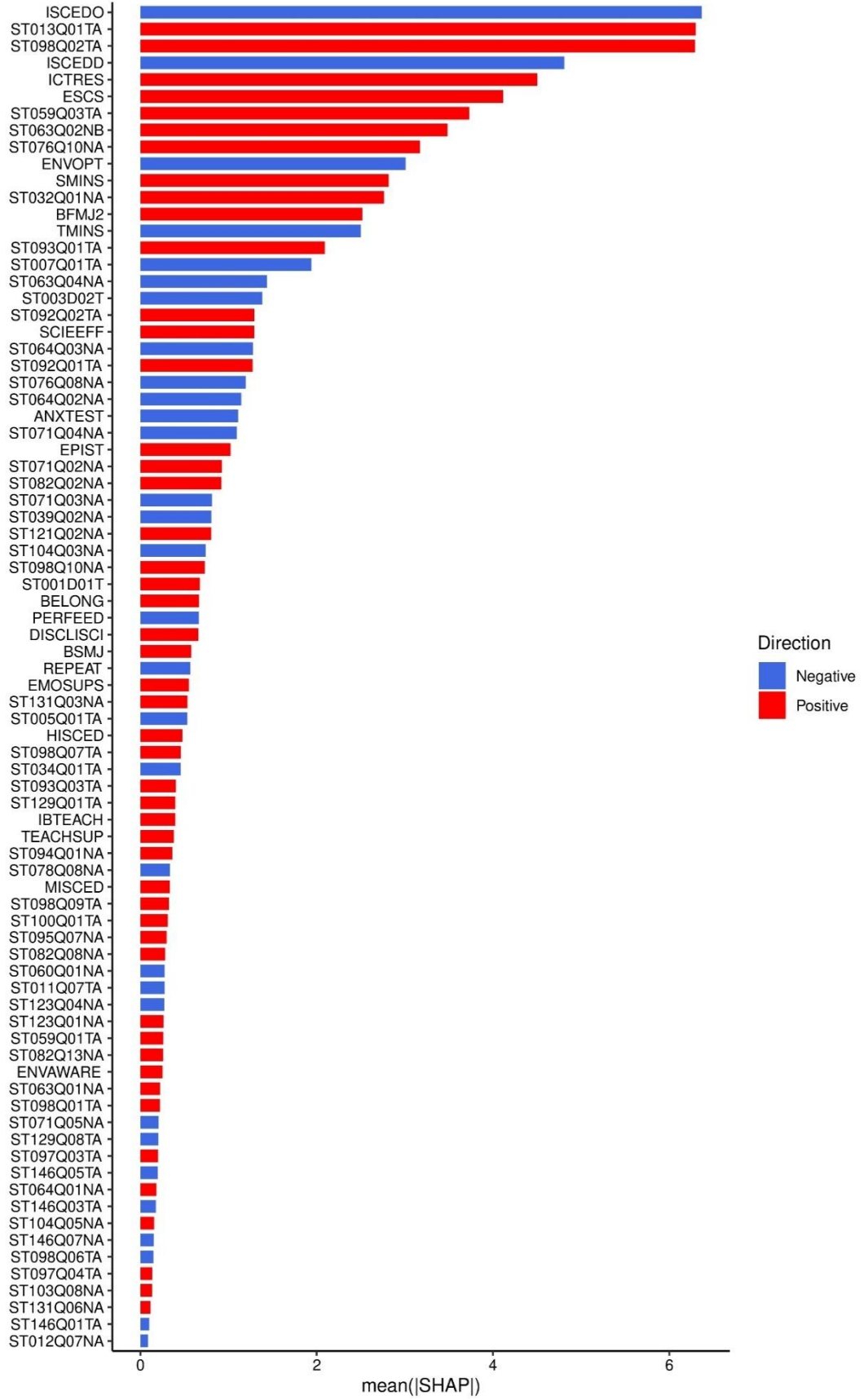
4.2.4 TR2 Bölgesi İçin Değişkenlerin Önem Seviyeleri

PISA 2015 Türkiye örnekleme TR2 bölgesi Fen okuryazarlık alanı için en iyi tahmin performans gösteren XGBoost algoritmasına göre SHAP (SHapley Additive exPlanations) değerleri kullanılarak değişkenlerin önem parametreleri belirlenmiştir. Sonuçlar her bir değişkenin ortalama SHAP değerlerinin mutlak değerine göre azalan sırada Tablo 10'da, sıralanmıştır. Bu sıralamaya göre diğer bölgelerdeki değişkenlerin karşılaştırması ise Kısım 4.13'te Jaccard indeksine göre yapılacaktır. Burada model için önemli olan sıralı değişkenler üç önem grubuna ayrılmış ve bu işlem model için kullanılan toplam değişkenleri 3 eşit parçaya ayırarak gerçekleştirilmiştir. Burada yüksek derecede önemli değişkenler 1. Grup (ilk 33 tanesi), orta derecede önemli değişkenler 2. Grup (34 - 66) ve düşük derecede önemli değişkenler 3. Grup (67 - 99) olarak etiketlenmiştir. Ayrıca Şekil 12 ile Tablo 10 da özetlenen değişkenlerin ortalama SHAP değerleri görselleştirilmiştir. Bu şekil her bir tahmin edici değişkenin model tahminlerine olan ortalama etkisini ve yönünü (pozitif ve negatif katkıları) vurgulamaktadır. Şekil 12 her bir değişkenin test veri setindeki her bir gözlemdeki etkisini göstermektedir. Bu şekil içerisinde her bir nokta, bir veri noktasının bir özelliği için SHAP değerini temsil eder. Bu şeklin sağ tarafında pozitif SHAP değerleri, sol tarafında ise negatif SHAP değerleri bulunmaktadır.

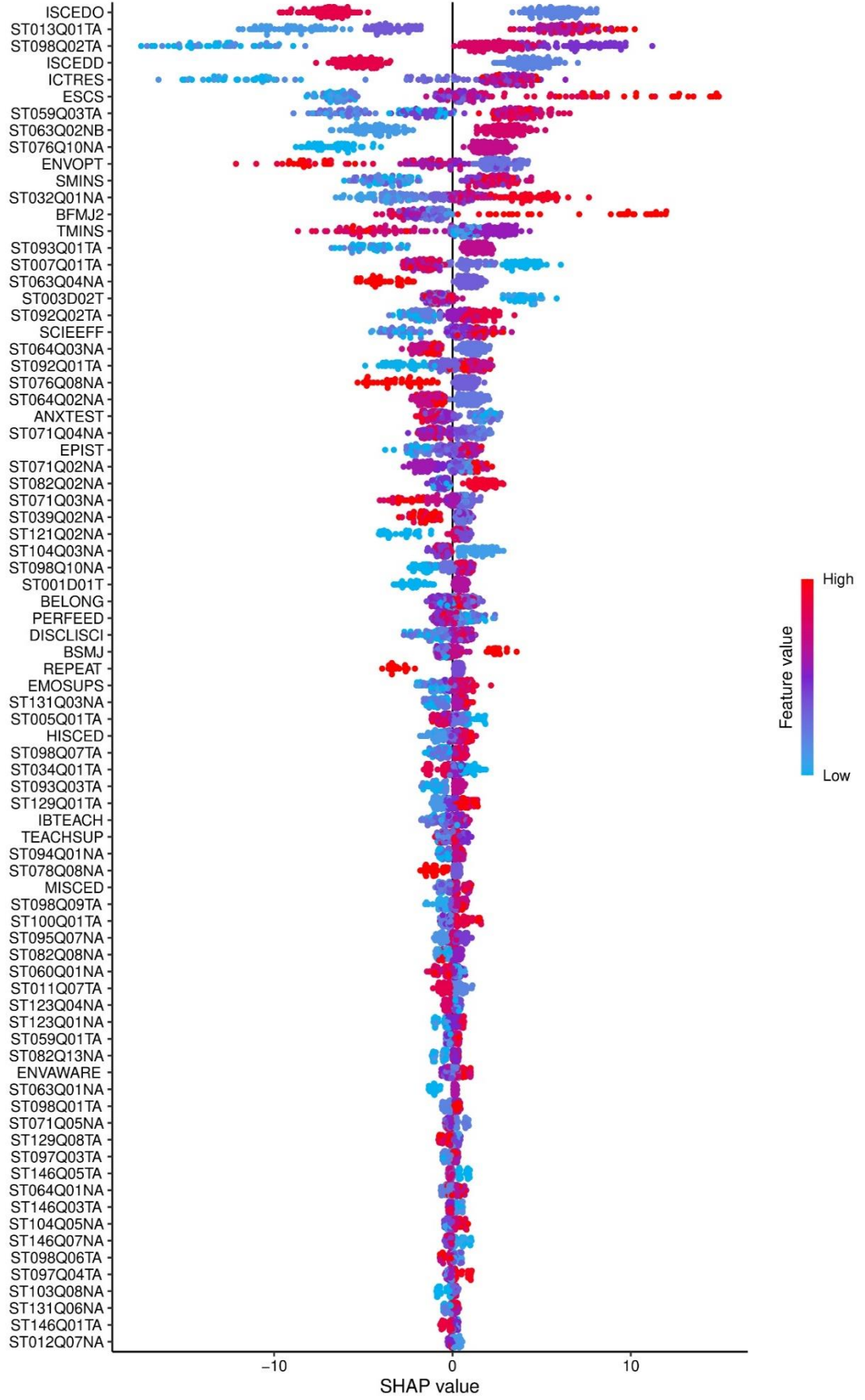
Tahmin edici değişkenlerin model çıktısına olan etkisi, bu SHAP değerlerinin büyüklüğüne göre belirlenir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Tablo 10, Şekil 12 ve 13'teki bulgulara göre, TR2 bölgesi için model tahmininde önem seviyesi yüksek ilk 10 değişken, parantez içleri SHAP değerlerini ve model katkı yönlerini göstermek üzere, şu şekilde sıralanmıştır: ISCEDO (6.367, pozitif), ST013Q01TA(6.300, negatif), ST098Q02TA (6.291, pozitif), ISCEDD (4.808, negatif), ICTRES (4.503, pozitif), ESCS (4.117, pozitif), ST059Q03TA (3.731, negatif), ST063Q02NB (3.484, pozitif), ST076Q10NA (3.172, pozitif), ENVOPT (3.008, negatif). Şekil 13'te kırmızı ve mavi renkler sırasıyla pozitif ve negatif yönlü etkileri temsil etmektedir. Bu, modelin tahminlerinin hangi özellikler tarafından ne kadar ve hangi yönde etkilendiğini daha net bir şekilde anlaşılmasına yardımcı olmaktadır. Buna göre negatif özellikteki değişkenlerin (ST013Q01TA, ISCEDD, ST059Q03TA, ENVOPT) değerlerindeki artış öğrencilerin Fen başarı skorlarını düşürmekte buna karşı pozitif özellikteki değişkenlerin değerlerindeki artış ise öğrencilerin Fen başarı skorlarını artırıcı yönde etki etmektedir.

Tablo 10: TR2 (Batı Marmara) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.

ID	Değişken	SHAP değeri	Önem Grubu	ID	Değişken	SHAP değeri	Önem Grubu
1	ISCEDO	6.367	Grup 1	55	ST100Q01TA	0.312	Grup 2
2	ST013Q01TA	6.300	Grup 1	56	ST095Q07NA	0.298	Grup 2
3	ST098Q02TA	6.291	Grup 1	57	ST082Q08NA	0.280	Grup 2
4	ISCEDD	4.808	Grup 1	58	ST060Q01NA	0.275	Grup 2
5	ICTRES	4.503	Grup 1	59	ST011Q07TA	0.274	Grup 2
6	ESCS	4.117	Grup 1	60	ST123Q04NA	0.271	Grup 2
7	ST059Q03TA	3.731	Grup 1	61	ST123Q01NA	0.262	Grup 2
8	ST063Q02NB	3.484	Grup 1	62	ST059Q01TA	0.258	Grup 2
9	ST076Q10NA	3.172	Grup 1	63	ST082Q13NA	0.258	Grup 2
10	ENVOPT	3.008	Grup 1	64	ENVAWARE	0.251	Grup 2
11	SMINS	2.817	Grup 1	65	ST063Q01NA	0.223	Grup 2
12	ST032Q01NA	2.763	Grup 1	66	ST098Q01TA	0.222	Grup 2
13	BFMJ2	2.517	Grup 1	67	ST071Q05NA	0.207	Grup 3
14	TMINS	2.500	Grup 1	68	ST129Q08TA	0.203	Grup 3
15	ST093Q01TA	2.091	Grup 1	69	ST097Q03TA	0.200	Grup 3
16	ST007Q01TA	1.940	Grup 1	70	ST146Q05TA	0.197	Grup 3
17	ST063Q04NA	1.436	Grup 1	71	ST064Q01NA	0.182	Grup 3
18	ST003D02T	1.381	Grup 1	72	ST146Q03TA	0.175	Grup 3
19	ST092Q02TA	1.294	Grup 1	73	ST104Q05NA	0.155	Grup 3
20	SCIEEFF	1.293	Grup 1	74	ST146Q07NA	0.152	Grup 3
21	ST064Q03NA	1.277	Grup 1	75	ST098Q06TA	0.148	Grup 3
22	ST092Q01TA	1.273	Grup 1	76	ST097Q04TA	0.136	Grup 3
23	ST076Q08NA	1.195	Grup 1	77	ST103Q08NA	0.133	Grup 3
24	ST064Q02NA	1.144	Grup 1	78	ST131Q06NA	0.114	Grup 3
25	ANXTEST	1.109	Grup 1	79	ST146Q01TA	0.101	Grup 3
26	ST071Q04NA	1.094	Grup 1	80	ST012Q07NA	0.087	Grup 3
27	EPIST	1.023	Grup 1	81	ST011Q05TA	0	Grup 3
28	ST071Q02NA	0.925	Grup 1	82	ST095Q13NA	0	Grup 3
29	ST082Q02NA	0.918	Grup 1	83	ST076Q02NA	0	Grup 3
30	ST071Q03NA	0.813	Grup 1	84	ST121Q03NA	0	Grup 3
31	ST039Q02NA	0.805	Grup 1	85	GRADE	0	Grup 3
32	ST121Q02NA	0.803	Grup 1	86	ST082Q03NA	0	Grup 3
33	ST104Q03NA	0.741	Grup 1	87	ST063Q06NA	0	Grup 3
34	ST098Q10NA	0.730	Grup 2	88	ST034Q06TA	0	Grup 3
35	ST001D01T	0.673	Grup 2	89	ST082Q12NA	0	Grup 3
36	BELONG	0.666	Grup 2	90	ST078Q05NA	0	Grup 3
37	PERFEED	0.662	Grup 2	91	ST062Q03TA	0	Grup 3
38	DISCLISCI	0.658	Grup 2	92	ST093Q05TA	0	Grup 3
39	BSMJ	0.577	Grup 2	93	ST123Q02NA	0	Grup 3
40	REPEAT	0.565	Grup 2	94	ST061Q01NA	0	Grup 3
41	EMOSUPS	0.550	Grup 2	95	FISCED	0	Grup 3
42	ST131Q03NA	0.533	Grup 2	96	ST104Q02NA	0	Grup 3
43	ST005Q01TA	0.532	Grup 2	97	ST100Q03TA	0	Grup 3
44	HISCED	0.477	Grup 2	98	ST103Q01NA	0	Grup 3
45	ST098Q07TA	0.458	Grup 2	99	ST097Q01TA	0	Grup 3
46	ST034Q01TA	0.457	Grup 2				
47	ST093Q03TA	0.401	Grup 2				
48	ST129Q01TA	0.396	Grup 2				
49	IBTEACH	0.396	Grup 2				
50	TEACHSUP	0.379	Grup 2				
51	ST094Q01NA	0.363	Grup 2				
52	ST078Q08NA	0.337	Grup 2				
53	MISCED	0.335	Grup 2				
54	ST098Q09TA	0.325	Grup 2				



Şekil 12: TR2 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.



Şekil 13: TR2 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.

4.3 TR3 Bölgesi PISA-2015 Bulguları

4.3.1 TR3 Bölgesi Hakkında Temel Bilgiler

2015 PISA uygulamasında Türkiye örnekleminin üçüncü bölgesi TR3 olarak adlandırılmıştır. TR3 bölgesi Türkiye'nin Ege bölgesinde bulunan illeri kapsamaktadır. TR3 bölgesi (Ege); İzmir, Manisa, Kütahya, Uşak, Afyonkarahisar, Aydın, Muğla illerinde bulunan 7. Sınıf ve üzeri sınıf düzeyinde öğrenim gören öğrencilerin Fen başarı skorlarını ifade etmektedir. TR3 (Ege) bölgesinden 707 öğrenci 2015 PISA uygulamasına katılmıştır. 2015 PISA uygulamasına katılan tüm öğrencilerin %11.99'unun Ege bölgesinden katıldığını görmekteyiz.

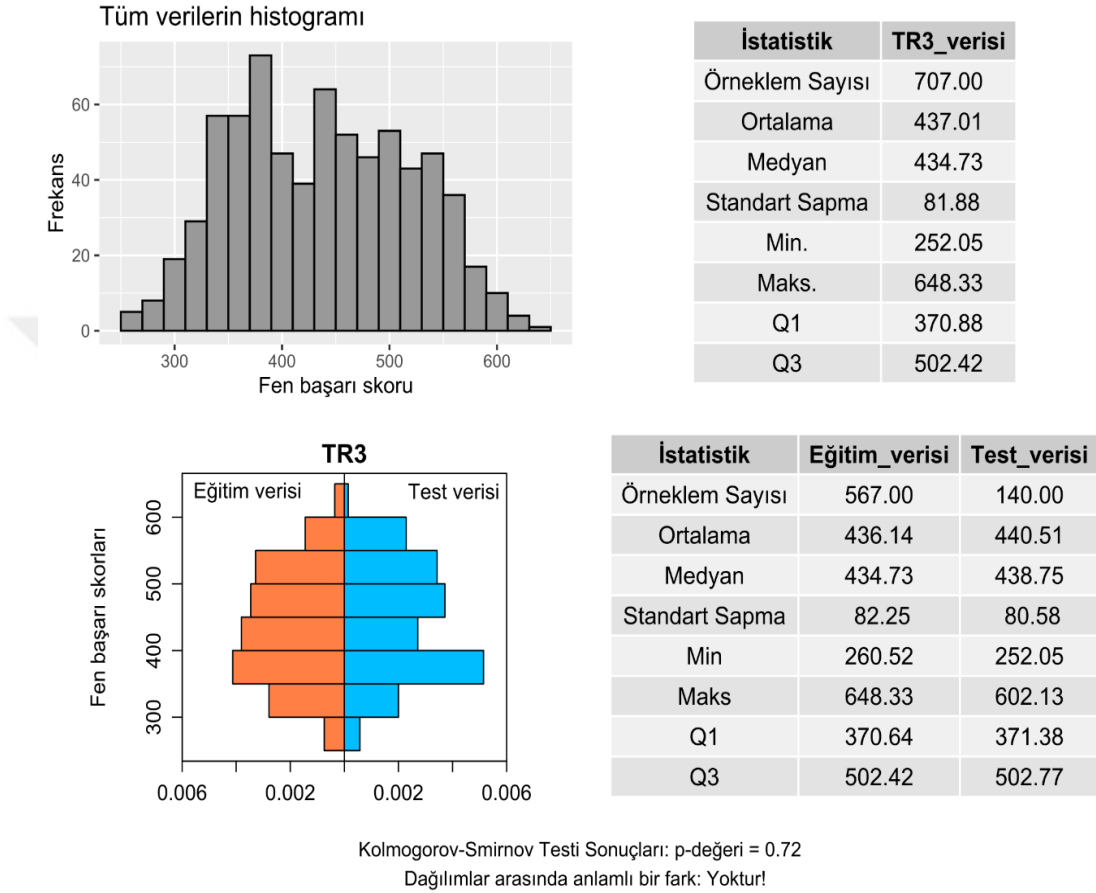
4.3.2 TR3 Bölgesinin Veri Yapısı

Şekil 14'te, Türkiye'nin TR3 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı skorlarına ait istatistiksel özeti sunulmaktadır. Şekil 14'ün üst panelinde, tüm veri setini temsil eden bir histogram ve istatistiksel özet bilgileri gösterilmektedir. Aynı şeklin alt panelinde ise, makine öğrenmesi modelinin oluşturulması ve modelin performansının değerlendirilmesi süreçlerinde kullanılacak olan eğitim ve test alt veri setlerine ait istatistiksel bilgiler bulunmaktadır. Bu alt veri setleri ana veri setinden rastgele bir şekilde ve literatüre uygun olarak %80 (Eğitim verisi) / %20 (Test verisi) oranında örneklenerek oluşturulmuştur (Gholamy ve diğerleri, 2018).

Şekil 14'ün üst panelindeki histograma göre, fen başarı skorlarının ortalama puan etrafında merkezi bir eğilimle dağıldığı görülmektedir. Bu TR3 bölgesindeki öğrencilerin (707 kişi) Fen başarı skorlarının normal dağılıma uyduğunu belirtmektedir. Ayrıca bu bölge için tüm veri setine ait ortalama puan 437.01, medyan 434.73, standart sapma 81.88 olarak hesaplanmıştır. Bunlara ek olarak minimum ve maksimum puanlar sırasıyla 252.05 ve 648.33 olup, birinci çeyrek (Q1) 370.88 ve üçüncü çeyrek (Q3) 502.42 olarak belirlenmiştir.

Şekil 14'ün üst panelinde, eğitim veri setinin (567 kişi) ortalama puanı 436.14, medyanı 434.73 ve standart sapması 82.25 olarak tespit edilmiştir. Eğitim verisi için minimum ve maksimum puanlar sırasıyla 260.52 ve 648.33 olup, birinci çeyrek (Q1) 370.64 ve üçüncü çeyrek (Q3) 502.42 olarak belirlenmiştir. Bunların yanında Test veri setinin (140 kişi), ortalama puanı 440.51, medyanı 438.75 ve standart sapması 80.58 olarak tespit edilmiştir. Test verisi için minimum ve maksimum puanlar sırasıyla 252.05 ve 602.13 olup, birinci çeyrek (Q1) 371.38 ve üçüncü çeyrek (Q3) 502.77 olarak belirlenmiştir. Bu veri seti

için rasgele bir şekilde oluşturulan eğitim ve test veri setlerine ait dağılımlarının birbirinden anlamlı bir şekilde farklı olmadığı Kolmogorov-Smirnov testi sonucuna göre ($p > 0.05$) tespit edilmiştir.



Şekil 14: Türkiye'nin TR3 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)

4.3.3 TR3 Bölgesinde Makine Öğrenme Modeli

Tablo 11 TR3 (Ege) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı ve 10 tekrarlı çapraz doğrulama performans metriklerini ve optimize edilmiş Hiperparametre değerlerini göstermektedir. Bu tabloda elde edilen bulgular, XGBoost, RF, SVR ve BRNN algoritmaları ile Kısım 3.3'te oluşturulan makine öğrenme modellerin 10 tekrarlı ve 10 katlı çapraz doğrulama tahmin performansları elde edilmiştir. Bu tablonun ayrıntılı sonuçları ve kısaltmalar için gerekli açıklamalar EK-1 içerisinde sunulmuştur. İncelenen ve oluşturulan makine öğrenmesi modellerin başarı performansları

RMSE, R^2 ve MAE metrikleri üzerinden değerlendirilmiş ve XGBoost modelinin en düşük RMSE (49.21), en yüksek R^2 (0.65) ve en düşük MAE (40.14) değerleri ile eğitim verileri için en iyi performansı gösterdiği tespit edilmiştir. Buna karşın SVR, BRNN ve RF modelleri sırasıyla 50.58, 51.59 ve 51.73 RMSE değerleri ile XGBoost algoritmasının ardından sıralanmışlardır. Bu veriler, TR3 bölgesinin eğitim verileri için XGBoost modelinin Fen okuryazarlık skorlarının tahmininde diğer modellere göre daha iyi bir öğrenme ortaya koyduğunu göstermektedir.

Tablo 11: TR3(Ege) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri

Model	RMSE	R^2	MAE	Hiperparametre değerleri
XGBoost	49.21	0.65	40.14	nrounds 777.00000000 max_depth 2.00000000 eta 0.03748243 gamma 3.64823660 colsample_bytree 0.46808052 min_child_weight 15.00000000 subsample 0.99495393
SVR	50.58	0.63	41.15	sigma 0.008288906 C 0.917046224
BRNN	51.59	0.62	42.09	neurons 20
RF	51.73	0.63	41.96	mtry "35" splitrule "extratrees" min.node.size "7"

Tablo 12’de TR3 bölgesine ait verileri için oluşturulan makine öğrenme algoritmalarının test verisi performans göstergeleri detaylı bir şekilde sunulmaktadır. Buna göre XGBoost modelinin RMSE değeri 48.65 R^2 değeri 0.63 ve MAE değeri 40.41 olarak belirlenmiş ve diğer modellerden daha üstün bir performans sergilediği tespit edilmiştir. Ayrıca XGBoost modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.778, güven aralığının alt sınırı 0.708 ve üst sınırı 0.833 ile yüksek bir uyumu gösterdiği belirlenmiştir. Buna karşın ikinci en iyi performans gösteren RF modelinde, RMSE değeri 50.27, R^2 değeri 0.62 ve MAE değeri 40.30 ile XGBoost’a göre daha düşük olup, uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.730 (lwr.ci: 0.659- upr.ci: 0.789) olarak hesaplanmıştır. SVR ve BRNN modellerinde sırasıyla RMSE değerleri 52.84 ve 56.58, R^2 değerleri 0.57 ve 0.52, MAE değerleri 44.13 ve 47.17 olarak belirlenmiştir. SVR modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.732 (lwr.ci: 0.651-

upr.ci: 0.797) iken, BRNN modelinin modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.706 (lwr.ci: 0.616- upr.ci: 0.778) olarak hesaplanmıştır. Bu sonuçlar XGBoost modelinin en düşük hata oranları ve en yüksek açıklama gücü ile diğer modellerden daha üstün performans sergilendiğini ortaya koymaktadır.

Tablo 12: TR3 (Ege) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri

Model	RMSE	R^2	MAE	ρ_c
XGBoost	48.65	0.63	40.41	XGBoost.est 0.7780267 XGBoost.lwr.ci 0.7079294 XGBoost.upr.ci 0.8329460
RF	50.27	0.62	40.30	RF.est 0.7305511 RF.lwr.ci 0.6588200 RF.upr.ci 0.7891184
SVR	52.84	0.57	44.13	SVR.est 0.7321346 SVR.lwr.ci 0.6510547 SVR.upr.ci 0.7966948
BRNN	56.58	0.52	47.17	BRNN.est 0.7066752 BRNN.lwr.ci 0.6164575 BRNN.upr.ci 0.7785786

ρ_c = Uyum Korelasyon Katsayısı (Concordance Correlation Coefficient)

est = ρ_c için merkezi tahmin edilen değeri

lwr.ci = güven aralığının alt sınırını (lower confidence interval)

upr.ci = güven aralığının üst sınırını (upper confidence interval)

Tablo 11 ve Tablo 12 verilerine göre yapılan genel değerlendirmede, XGBoost modelinin hem eğitim hem de test verilerinde en yüksek performansı sergilediği görülmektedir. Bu durum, modelin aşırı uyum (overfitting) yapmadığını ve genelleştirme kabiliyetinin güçlü olduğunu göstermektedir. Ayrıca test verilerinin uyum korelasyon katsayısının (ρ_c) yüksek olması, modelin tahminlerinin gerçek verilerle yüksek uyum gösterdiğini belirtmektedir. Bu, XGBoost modelinin hem eğitim hem de test verilerinde tutarlı ve güvenilir bir performans sergilediğinin kanıtıdır.

Diğer yandan RF, SVR ve BRNN modelleri ise eğitim ve test verileri arasında benzer hata oranlarına sahip olsalar da performansları XGBoost modellerine kıyasla daha düşük seviyelerde kalmıştır. Bu durum, RF, SVR ve BRNN modellerinin genelleştirme kabiliyetinin sınırlı olduğunu ve yeni veriler üzerinde daha düşük performans sergilediğini göstermektedir. Her üç modelin de ρ_c değerleri, tahminlerinin gerçek verilerle makul bir uyum gösterdiğini belirtmektedir.

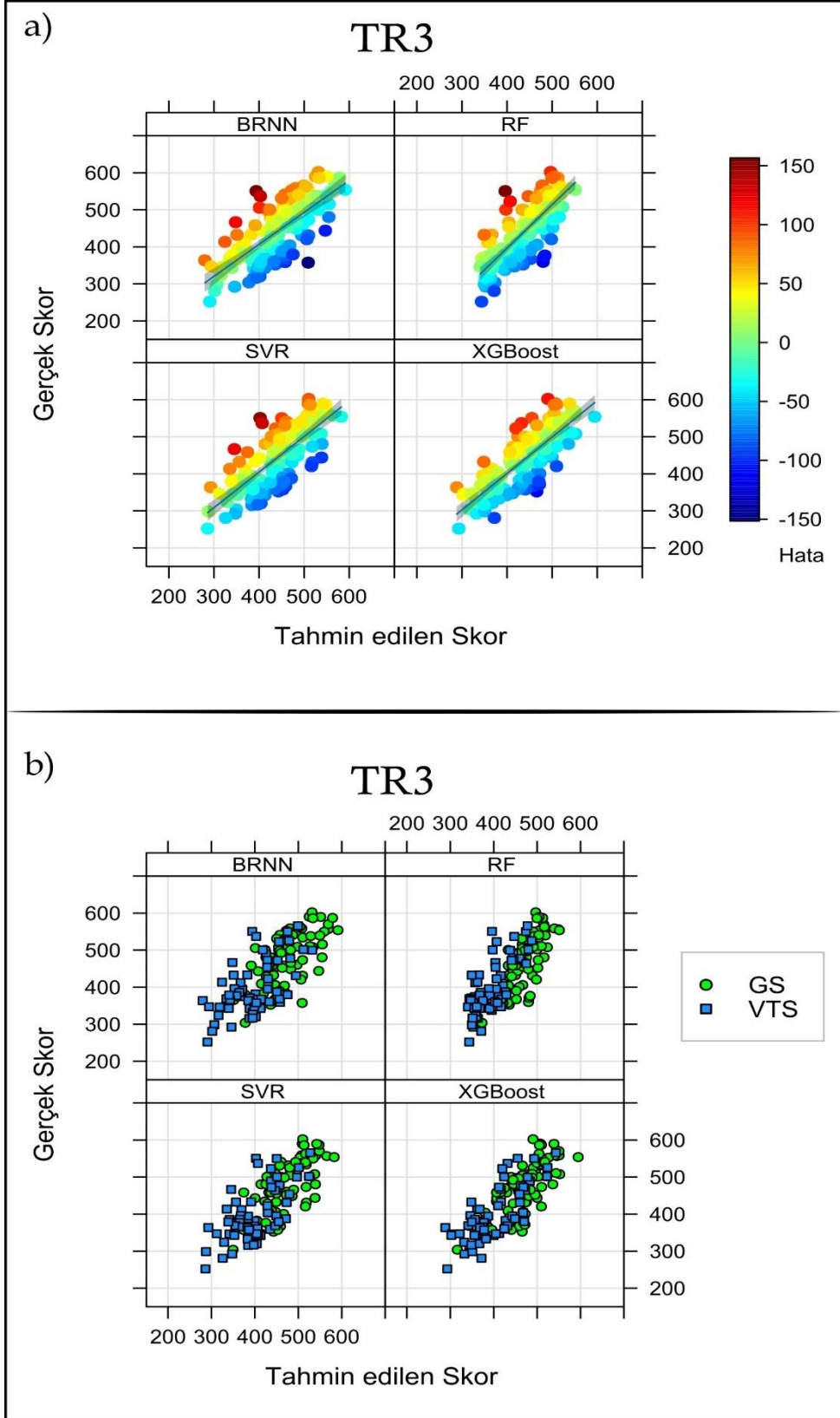
Şekil 15, Tablo 12 da elde edilen performans metriklerine ait çapraz doğrulama diyagramlarını tanımlamaktadır. Şekil B a’da tüm modellerin hata dağılımlarının regresyon çizgisi boyunca ve yakınlarında dağıldığı görülmektedir. Bu durum tahmin hatalarının dağılımının normal yapıda olduğunu ve hata ortalamasının da sıfıra yakın olduğunu göstermektedir. Bu sonuçlar Tablo 12 da elde edilen bulgular ile uyumludur. Bunlarla birlikte XGBoost modelinin yüksek tahmin performansı sergilediği, Şekil 15a’daki çapraz doğrulama diyagramının hata dağılımlarının regresyon çizgisi etrafında sıkı şekilde toplanmasından dolayı daha açık bir şekilde görülmektedir.

Şekil 15b, aynı test verileri için modellerin farklı okul türlerine (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim) ait tahmin edilen ve gerçek skorlarının karşılaştırmasını göstermektedir. Burada özellikle tanımlayıcı regresyon çizgisinden sapan ve modellerin tahmin etme performansının düşmesine neden olan okul grubunun Mesleki ve Teknik Ortaöğretim (VTS) okul türü olduğu dikkat çekmektedir. Bununla birlikte XGBoost, her tüm okul türlerinde de tahmin edilen skorların gerçek skorlarla en uyumlu olduğu model olarak öne çıkmaktadır. Ayrıca TR3 bölgesindeki Fen başarı skorları için genel ortaokul (GS) başarı düzeyinin VTS okul türüne göre belirgin şekilde artış göstermiş olması dikkat çekicidir. Özellikle okul türünün modellerin oluşturulmasında bir değişken olarak kullanılmamış olmasına rağmen, makine öğrenme algoritmasının bu başarı farklılıklarını diğer değişkenlerin yardımı ile tespit edebilmesi ve tahminlerini buna göre gerçekleştirmesi önemli bir başarı olduğu düşünülmektedir. Bu bulgular, XGBoost modelinin eğitim verileri üzerindeki güçlü performansını ve genelleştirme kabiliyetini destekler niteliktedir. XGBoost’un, farklı okul türlerindeki öğrencilerin performansını doğru bir şekilde tahmin edebilmesi, eğitim politikaları ve müdahalelerinin planlanmasında doğru tahmin skorları önerebileceğini ve bu konuda değerli bir araç olarak kullanılabileceğini göstermektedir. Özellikle Mesleki ve Teknik Ortaöğretim (VTS) öğrencileri için tahminlerdeki sapmalara rağmen, XGBoost’un genel olarak diğer modellere göre daha yüksek doğrulukla sonuçlar vermesi, bu modelin TR3 bölgesi için geniş bir yelpazede güvenilir tahminler oluşturabileceğini göstermektedir.

Şekil 16, farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost) tahmin ettiği başarı skorları ile gerçek başarı skorlarının karşılaştırmalı kutu grafiğini sunmaktadır. Şekil B de sunulan çapraz doğrulama bulgularına göre oluşturulmuş bu grafikler, modellerin tahmin performansını ve gerçek verilerle uyumunu değerlendirmek için önemli bir yaklaşım

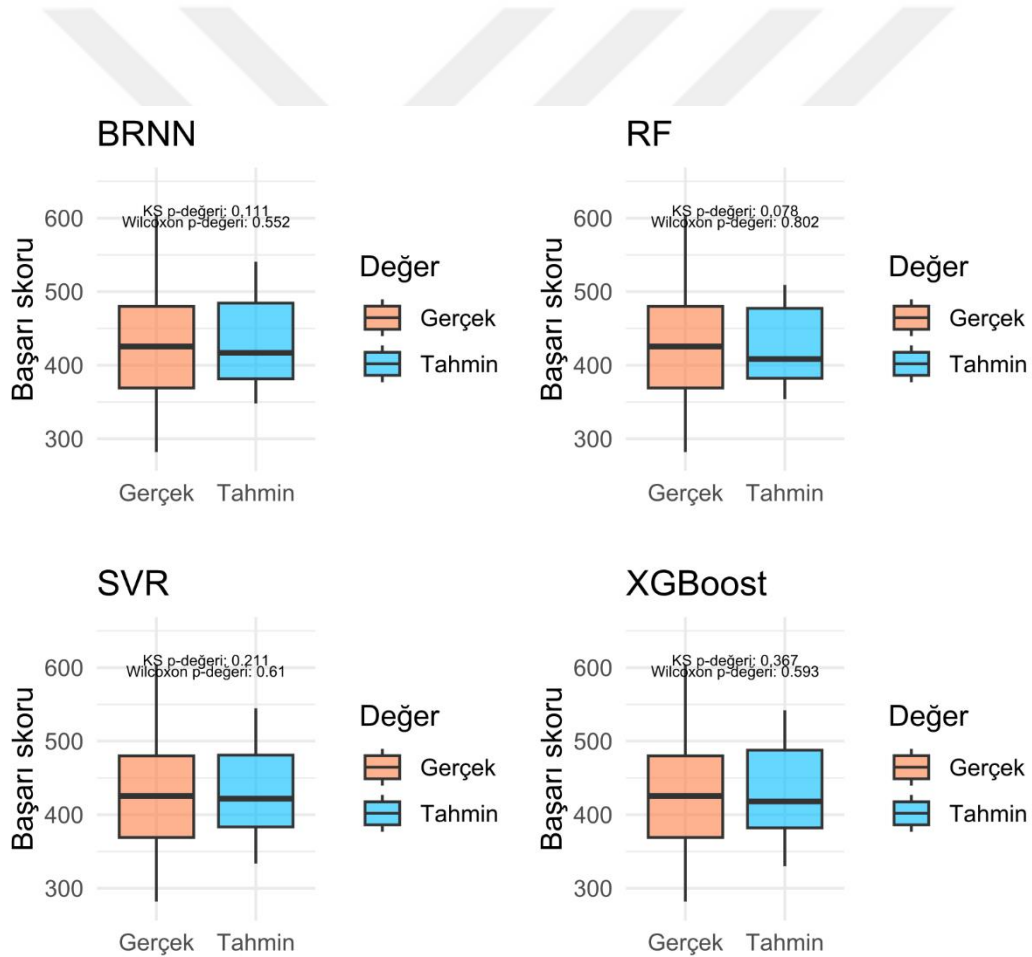
sağlamaktadır. Buna göre RF hariç tüm modellerin tahmin skorlarının hem Kolmogorov - Simirnov testi (KS) hem de Wilcoxon sıralama toplamı testine göre gerçek skorların genel dağılımlarından ve merkezi eğilimlerinden anlamlı ölçüde bir farklılığı ($p > 0.05$) olmadığı tespit edilmiştir. Bu durum makine öğrenmesi modellerinin tahmin edilen ve gerçek başarı skorları arasındaki uyumun yüksek olduğunu ortaya koymaktadır.





Şekil 15: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)

Şekil 16 en yüksek performans seviyesi gösteren XGBoost algoritması açısından değerlendirildiğinde, Kutu grafiklerinin gerçek ve tahmin skorları için temsil ettiği ortalama ve medyan değerlerinin neredeyse örtüştüğü ve dağılımların (kutu dışındaki uç örnekler için yeterli seviyelerde olmasa da) Q1-Q3 aralığı için oldukça yakın olduğunu görülmektedir. Şekil C'deki bulgulara göre tahmin edilen ve gerçek skorların genel dağılımlarının (KS p-değeri 0.367) ve merkezi eğilimlerinin (Wilcoxon p-değeri > 0.593), istatistiksel olarak birbirlerinden anlamlı bir şekilde farklı olmadığı belirlenmiştir. Bu durum, XGBoost algoritmasının TR3 verileri üzerinde en yüksek doğruluk ve güvenilirliği sunmakla birlikte hem bireysel gerçek skorları hem de gerçek skorlara ait dağılım yapısını doğru şekilde tahmin edebilme yeteneğini ortaya koymaktadır.



Şekil 16: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.

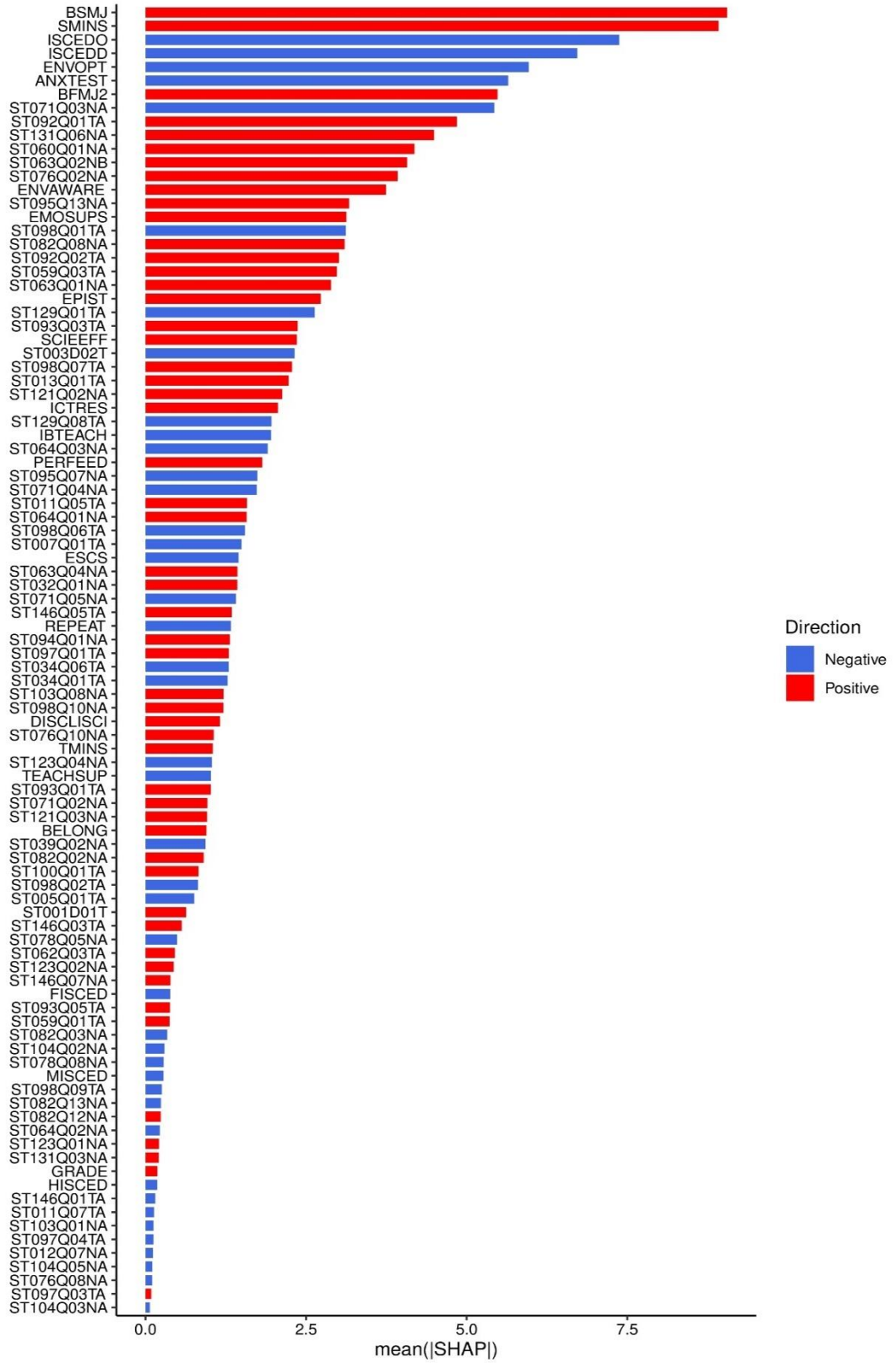
4.3.4 TR3 Bölgesi İçin Değişkenlerin Önem Seviyeleri

PISA 2015 Türkiye örnekleme TR3 bölgesi Fen okuryazarlık alanı için en iyi tahmin performans gösteren XGBoost algoritmasına göre SHAP (SHapley Additive exPlanations) değerleri kullanılarak değişkenlerin önem parametreleri belirlenmiştir. Sonuçlar her bir değişkenin ortalama SHAP değerlerinin mutlak değerine göre azalan sırasına Tablo 13'te, sıralanmıştır. Bu sıralamaya göre diğer bölgelerdeki değişkenlerin karşılaştırması ise Kısım 4.13'te Jaccard indeksine göre yapılacaktır. Burada model için önemli olan sıralı değişkenler üç önem grubuna ayrılmış ve bu işlem model için kullanılan toplam değişkenleri 3 eşit parçaya ayırarak gerçekleştirilmiştir. Burada yüksek derecede önemli değişkenler 1. Grup (ilk 33 tanesi), orta derecede önemli değişkenler 2. Grup (34- 66) ve düşük derecede önemli değişkenler 3. Grup (67- 99) olarak etiketlenmiştir.

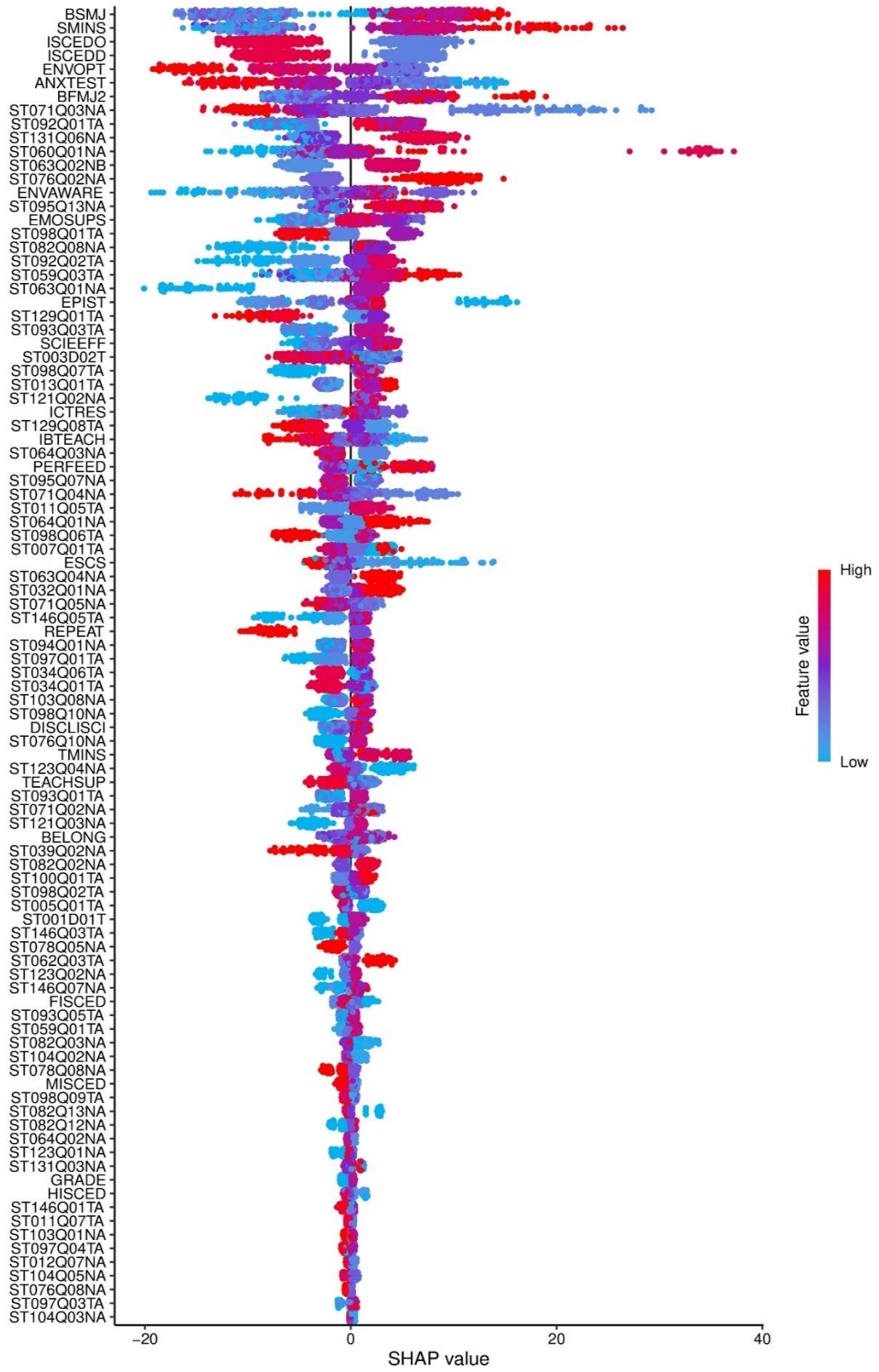
Ayrıca Şekil 17 ile Tablo 13'te özetlenen değişkenlerin ortalama SHAP değerleri görselleştirilmiştir. Bu şekil her bir tahmin edici değişkenin model tahminlerine olan ortalama etkisini ve yönünü (pozitif ve negatif katkıları) vurgulamaktadır. Şekil 18 her bir değişkenin test veri setindeki her bir gözlemdeki etkisini göstermektedir. Bu şekil içerisinde her bir nokta, bir veri noktasının bir özelliği için SHAP değerini temsil eder. Bu şeklin sağ tarafında pozitif SHAP değerleri, sol tarafında ise negatif SHAP değerleri bulunmaktadır. Tahmin edici değişkenlerin model çıktısına olan etkisi, bu SHAP değerlerinin büyüklüğüne göre belirlenir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Tablo 13, şekil 17 ve 18' deki bulgulara göre, TR3 bölgesi için model tahmininde önem seviyesi yüksek ilk 10 değişken, parantez içleri SHAP değerlerini ve model katkı yönlerini göstermek, şu şekilde sıralanmıştır: BSMJ (9.052, pozitif), SMINS (8.919, pozitif), ISCEDO (7.374, negatif), ISCEDD (6.722, negatif), ENVOPT (5.966, negatif), ANXTEST (5.645, negatif), BFMJ2(5.481, pozitif), ST071Q03NA(4.848, negatif), ST092Q01TA(5.429, pozitif), ST131Q06NA(4.492, pozitif). Şekil D'de kırmızı ve mavi renkler sırasıyla pozitif ve negatif yönlü etkileri temsil etmektedir. Bu, modelin tahminlerinin hangi özellikler tarafından ne kadar ve hangi yönde etkilendiğini daha net bir şekilde anlaşılmasına yardımcı olmaktadır. Buna göre negatif özellikteki değişkenlerin (ISCEDO, ISCEDD, ENVOPT, ANXTEST, ST071Q03NA) değerlerindeki artış öğrencilerin fen başarı skorlarını düşürmekte buna karşı pozitif özellikteki değişkenlerin değerlerindeki artış ise öğrencilerin Fen başarı skorlarını arttırıcı yönde etki etmektedir.

Tablo 13: TR3 (Ege) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.

ID	Değişken	SHAP değeri	Önem Grubu	ID	Değişken	SHAP değeri	Önem Grubu
1	BSMJ	9.052	Grup 1	55	TMINS	1.048	Grup 2
2	SMINS	8.919	Grup 1	56	ST123Q04NA	1.035	Grup 2
3	ISCEDO	7.374	Grup 1	57	TEACHSUP	1.018	Grup 2
4	ISCEDD	6.722	Grup 1	58	ST093Q01TA	1.017	Grup 2
5	ENVOPT	5.966	Grup 1	59	ST071Q02NA	0.965	Grup 2
6	ANXTEST	5.645	Grup 1	60	ST121Q03NA	0.958	Grup 2
7	BFMJ2	5.481	Grup 1	61	BELONG	0.950	Grup 2
8	ST071Q03NA	5.429	Grup 1	62	ST039Q02NA	0.936	Grup 2
9	ST092Q01TA	4.848	Grup 1	63	ST082Q02NA	0.906	Grup 2
10	ST131Q06NA	4.492	Grup 1	64	ST100Q01TA	0.825	Grup 2
11	ST060Q01NA	4.187	Grup 1	65	ST098Q02TA	0.818	Grup 2
12	ST063Q02NB	4.073	Grup 1	66	ST005Q01TA	0.760	Grup 2
13	ST076Q02NA	3.925	Grup 1	67	ST001D01T	0.633	Grup 3
14	ENVAWARE	3.743	Grup 1	68	ST146Q03TA	0.564	Grup 3
15	ST095Q13NA	3.169	Grup 1	69	ST078Q05NA	0.494	Grup 3
16	EMOSUPS	3.126	Grup 1	70	ST062Q03TA	0.458	Grup 3
17	ST098Q01TA	3.119	Grup 1	71	ST123Q02NA	0.438	Grup 3
18	ST082Q08NA	3.100	Grup 1	72	ST146Q07NA	0.388	Grup 3
19	ST092Q02TA	3.013	Grup 1	73	FISCED	0.384	Grup 3
20	ST059Q03TA	2.980	Grup 1	74	ST093Q05TA	0.380	Grup 3
21	ST063Q01NA	2.887	Grup 1	75	ST059Q01TA	0.377	Grup 3
22	EPIST	2.729	Grup 1	76	ST082Q03NA	0.338	Grup 3
23	ST129Q01TA	2.634	Grup 1	77	ST104Q02NA	0.296	Grup 3
24	ST093Q03TA	2.368	Grup 1	78	ST078Q08NA	0.282	Grup 3
25	SCIEEFF	2.356	Grup 1	79	MISCED	0.282	Grup 3
26	ST003D02T	2.321	Grup 1	80	ST098Q09TA	0.255	Grup 3
27	ST098Q07TA	2.281	Grup 1	81	ST082Q13NA	0.242	Grup 3
28	ST013Q01TA	2.228	Grup 1	82	ST082Q12NA	0.237	Grup 3
29	ST121Q02NA	2.128	Grup 1	83	ST064Q02NA	0.225	Grup 3
30	ICTRES	2.062	Grup 1	84	ST123Q01NA	0.211	Grup 3
31	ST129Q08TA	1.961	Grup 1	85	ST131Q03NA	0.208	Grup 3
32	IBTEACH	1.957	Grup 1	86	GRADE	0.184	Grup 3
33	ST064Q03NA	1.904	Grup 1	87	HISCED	0.183	Grup 3
34	PERFEED	1.819	Grup 2	88	ST146Q01TA	0.153	Grup 3
35	ST095Q07NA	1.742	Grup 2	89	ST011Q07TA	0.133	Grup 3
36	ST071Q04NA	1.730	Grup 2	90	ST103Q01NA	0.126	Grup 3
37	ST011Q05TA	1.577	Grup 2	91	ST097Q04TA	0.125	Grup 3
38	ST064Q01NA	1.575	Grup 2	92	ST012Q07NA	0.115	Grup 3
39	ST098Q06TA	1.549	Grup 2	93	ST104Q05NA	0.107	Grup 3
40	ST007Q01TA	1.494	Grup 2	94	ST076Q08NA	0.104	Grup 3
41	ESCS	1.450	Grup 2	95	ST097Q03TA	0.091	Grup 3
42	ST063Q04NA	1.431	Grup 2	96	ST104Q03NA	0.068	Grup 3
43	ST032Q01NA	1.429	Grup 2	97	ST063Q06NA	0	Grup 3
44	ST071Q05NA	1.407	Grup 2	98	ST061Q01NA	0	Grup 3
45	ST146Q05TA	1.341	Grup 2	99	ST100Q03TA	0	Grup 3
46	REPEAT	1.331	Grup 2				
47	ST094Q01NA	1.315	Grup 2				
48	ST097Q01TA	1.297	Grup 2				
49	ST034Q06TA	1.295	Grup 2				
50	ST034Q01TA	1.278	Grup 2				
51	ST103Q08NA	1.216	Grup 2				
52	ST098Q10NA	1.212	Grup 2				
53	DISCLSCI	1.158	Grup 2				
54	ST076Q10NA	1.063	Grup 2				



Şekil 17: TR3 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.



Şekil 18: TR3 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.

4.4 TR4 Bölgesi PISA-2015 Bulguları

4.4.1 TR4 Bölgesi Hakkında Temel Bilgiler

2015 PISA uygulamasında Türkiye örnekleminin dördüncü bölgesi TR4 olarak adlandırılmıştır. TR4 bölgesi 2015 PISA uygulamasında ayrılan bölgelerden biri olan Doğu Marmara bölgesidir. Doğu Marmara Bölgesi Kocaeli, Bursa, Sakarya, Düzce, Bilecik ve Bolu illerini kapsamaktadır. Doğu Marmara bölgesinden 7.sınıf ve üzeri düzeyde eğitim gören 510 öğrenci 2015 PISA sınavına katıldığı görülmektedir. 2015 PISA sınavına katılan tüm Türkiye örnekleminin %8.65'inin Doğu Marmara bölgesinden katıldığını görmekteyiz.

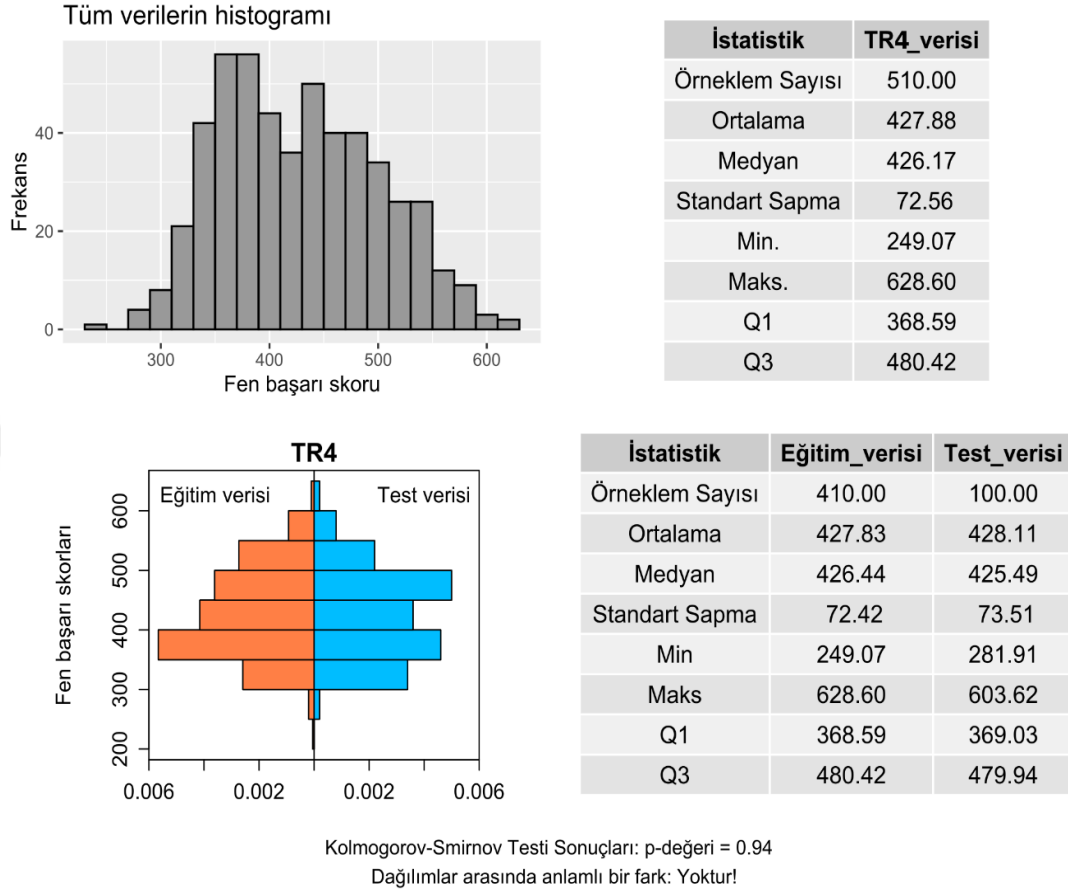
4.4.2 TR4 Bölgesinin Veri Yapısı

Şekil 19'da Türkiye'nin TR4 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özeti sunulmaktadır. Şekil 19'un üst panelinde, tüm veri setini temsil eden bir histogram ve istatistiksel özet bilgileri gösterilmektedir. Aynı şeklin alt panelinde ise, makine öğrenmesi modelinin oluşturulması ve modelin performansının değerlendirilmesi süreçlerinde kullanılacak olan eğitim ve test alt veri setlerine ait istatistiksel bilgiler bulunmaktadır. Bu alt veri setleri ana veri setinden rastgele bir şekilde ve literatüre uygun olarak %80 (Eğitim verisi) / %20 (Test verisi) oranında örneklemlenerek oluşturulmuştur (Gholamy ve diğerleri, 2018).

Şekil 19'un üst panelindeki histograma göre, fen puanlarının ortalama puan etrafında merkezi bir eğilimle dağıldığı görülmektedir. Bu TR4 bölgesindeki öğrencilerin (510 kişi) Fen başarı skorlarının normal dağılıma uyduğunu belirtmektedir. Ayrıca bu bölge için tüm veri setine ait ortalama puan 427.88, medyan 426.17, standart sapma 72.56 olarak hesaplanmıştır. Bunlara ek olarak minimum ve maksimum puanlar sırasıyla 249.07 ve 628.60 olup, birinci çeyrek (Q1) 368.59 ve üçüncü çeyrek (Q3) 480.42 olarak belirlenmiştir.

Şekil 19'un alt panelinde Eğitim veri setinin (410 kişi) ortalama puanı 427.83, medyanı 426.44 ve standart sapması 72.42 olarak tespit edilmiştir. Eğitim verisi için minimum ve maksimum puanlar sırasıyla 249.07 ve 628.42 olup, birinci çeyrek (Q1) 368.59 ve üçüncü çeyrek 480.42 olarak belirlenmiştir. Bunların yanında test veri setinin (100 kişi) ortalaması 428.11, medyanı 425.49 ve standart sapması 73.51 olarak tespit edilmiştir. Test veri seti için minimum ve maksimum puanlar sırasıyla 281.91 ve 603.62 olup, birinci çeyrek (Q1) 369.03 ve üçüncü çeyrek 479.94 olarak belirlenmiştir. Bu veri seti için rasgele bir şekilde

oluşturulan eğitim ve test alt veri setlerine ait dağılımlarının anlamlı bir şekilde farklı olmadığı Kolmogorov-Smirnov testi sonucuna göre ($p > 0.05$) tespit edilmiştir.



Şekil 19: Türkiye'nin TR4 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)

4.4.3 TR4 Bölgesinde Makine Öğrenme Modeli

Tablo 14 TR4 (Doğu Marmara) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı ve 10 tekrarlı çapraz doğrulama performans metriklerini ve optimize edilmiş Hiperparametre değerlerini göstermektedir. Bu tabloda elde edilen bulgular, XGBoost, RF, SVR ve BRNN algoritmaları ile Kısım 3.3'te oluşturulan makine öğrenme modellerin 10 tekrarlı ve 10 katlı çapraz doğrulama tahmin performansları elde edilmiştir. Bu tablonun ayrıntılı sonuçları ve kısaltmalar için gerekli açıklamalar EK-1 içerisinde sunulmuştur. İncelenen ve oluşturulan makine öğrenmesi modellerin başarı performansları RMSE, R^2 ve MAE metrikleri üzerinden değerlendirilmiş ve XGBoost modelinin en düşük RMSE (44.82), en yüksek R^2 (0.62), en düşük MAE (35.64) değerleri

ile eğitim verileri için en iyi performansı gösterdiği tespit edilmiştir. Buna karşın SVR, RF ve BRNN modelleri sırasıyla 45.26, 46.70 ve 47.27 RMSE değerleri ile XGBoost algoritmasının ardından sıralanmışlardır. Bu veriler TR4 bölgesinin eğitim verileri için XGBoost modelinin Fen okuryazarlık skorlarının tahmininde diğer modellere göre daha iyi bir öğrenme ortaya koyduğunu göstermektedir.

Tablo 14: TR4 (Doğu Marmara) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri (Bu tablo için ayrıntılar EK-1 de sunulmuştur)

Model	RMSE	R^2	MAE	Hiperparametre değerleri
XGBoost	44.82	0.62	35.64	nrounds 532.00000000 max_depth 7.00000000 eta 0.04449636 gamma 9.24671504 colsample_bytree 0.50804535 min_child_weight 16.00000000 subsample 0.75566734
SVR	45.26	0.61	36.22	sigma 0.00291816 C 1.80270740
RF	46.70	0.60	37.30	mtry "49" splitrule "maxstat" min.node.size "7"
BRNN	47.27	0.58	37.31	neurons 1

Tablo 15'te TR4 bölgesine ait verileri için oluşturulan makine öğrenme algoritmalarının test verisi performans göstergeleri detaylı bir şekilde sunulmaktadır. Bu tabloya göre, RF modelinin RMSE değeri 43.85, R^2 değeri 0.66 ve MAE değeri 34.13 olarak belirlenmiş ve diğer modellerden daha üstün bir performans sergilediği tespit edilmiştir. Ayrıca RF modelinin uyum performans katsayısı (ρ_c) merkezi tahmin edilen değeri 0.753, güven aralığının alt sınırı 0.673 ve üst sınırı 0.815 ile yüksek bir uyumu gösterdiği belirlenmiştir. Buna karşın ikinci en iyi performans gösteren SVR modelinde RMSE değeri 45.59, R^2 değeri 0.61 ve MAE değeri 36.52 ile RF 'ye göre daha düşük olup, uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.753 (lwr.ci: 0.664- upr.ci: 0.821) olarak hesaplanmıştır. XGBoost ve BRNN modellerinde ise sırasıyla RMSE değerleri 45.95 ve 48.19, R^2 değerleri 0.61 ve 0.57, MAE değerleri 36.34 ve 38.35 olarak belirlenmiştir. XGBoost modelinin uyum korelasyon katsayısı merkezi tahmini 0.757 (lwr.ci: 0.667- upr.ci: 0.826) iken, BRNN modelinin uyum korelasyon katsayısı merkezi tahmini 0.731 (lwr.ci: 0.632- upr.ci: 0.806) olarak hesaplanmıştır. Bu sonuçlar RF modelinin en düşük hata

oranları ve en yüksek açıklama gücü ile diğer modellerden daha üstün performans sergilediğini ortaya koymaktadır.

Tablo 15: TR4 (Doğu Marmara) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri

Model	RMSE	R^2	MAE	ρ_c
RF	43.85	0.66	34.13	RF.est 0.7528840 RF.lwr.ci 0.6733990 RF.upr.ci 0.8151614
SVR	45.59	0.61	36.52	SVR.est 0.7534201 SVR.lwr.ci 0.6642545 SVR.upr.ci 0.8214462
XGBoost	45.95	0.61	36.34	XGBoost.est 0.7575169 XGBoost.lwr.ci 0.6669364 XGBoost.upr.ci 0.8260339
BRNN	48.19	0.57	38.35	BRNN.est 0.7308396 BRNN.lwr.ci 0.6324731 BRNN.upr.ci 0.8060075

ρ_c = Uyum Korelasyon Katsayısı (Concordance Correlation Coefficient)

est = ρ_c için merkezi tahmin edilen değeri

lwr.ci = güven aralığının alt sınırını (lower confidence interval)

upr.ci = güven aralığının üst sınırını (upper confidence interval)

Tablo 14 ve Tablo 15 verilerine göre yapılan genel değerlendirmede eğitim verilerine göre XGBoost en iyi performansı göstermiş olsa da TR4 bölgesi için RF modeli test verilerine göre diğer modellerden daha üstün bir performans göstermiştir. XGBoost modelinin eğitim verilerinde çok iyi performans gösterip test verilerinde diğer modellere göre daha düşük performans göstermesi bu algoritmanın veri yapısı için ezberleme (overfitting) eğilimi göstermiş olduğu düşünülmektedir. Ayrıca RF test verilerinin uyum korelasyon katsayısının (ρ_c) yüksek olması, modelin tahminlerinin gerçek verilerle yüksek uyum gösterdiğini doğru tahminler elde ettiğini belirtmektedir.

Diğer yandan SVR, XGBoost ve BRNN modelleri ise eğitim ve test verileri arasında benzer hata oranlarına sahip olsalar da performansları RF modellerine kıyasla daha düşük seviyelerde kalmıştır. Bu durum SVR, XGBoost ve BRNN modellerinin genelleştirme kabiliyetinin sınırlı olduğunu ve yeni veriler üzerinde daha düşük performans sergilediğini göstermektedir. Her üç modelin de ρ_c değerleri, tahminlerinin gerçek verilerle makul bir uyum gösterdiğini belirtmektedir.

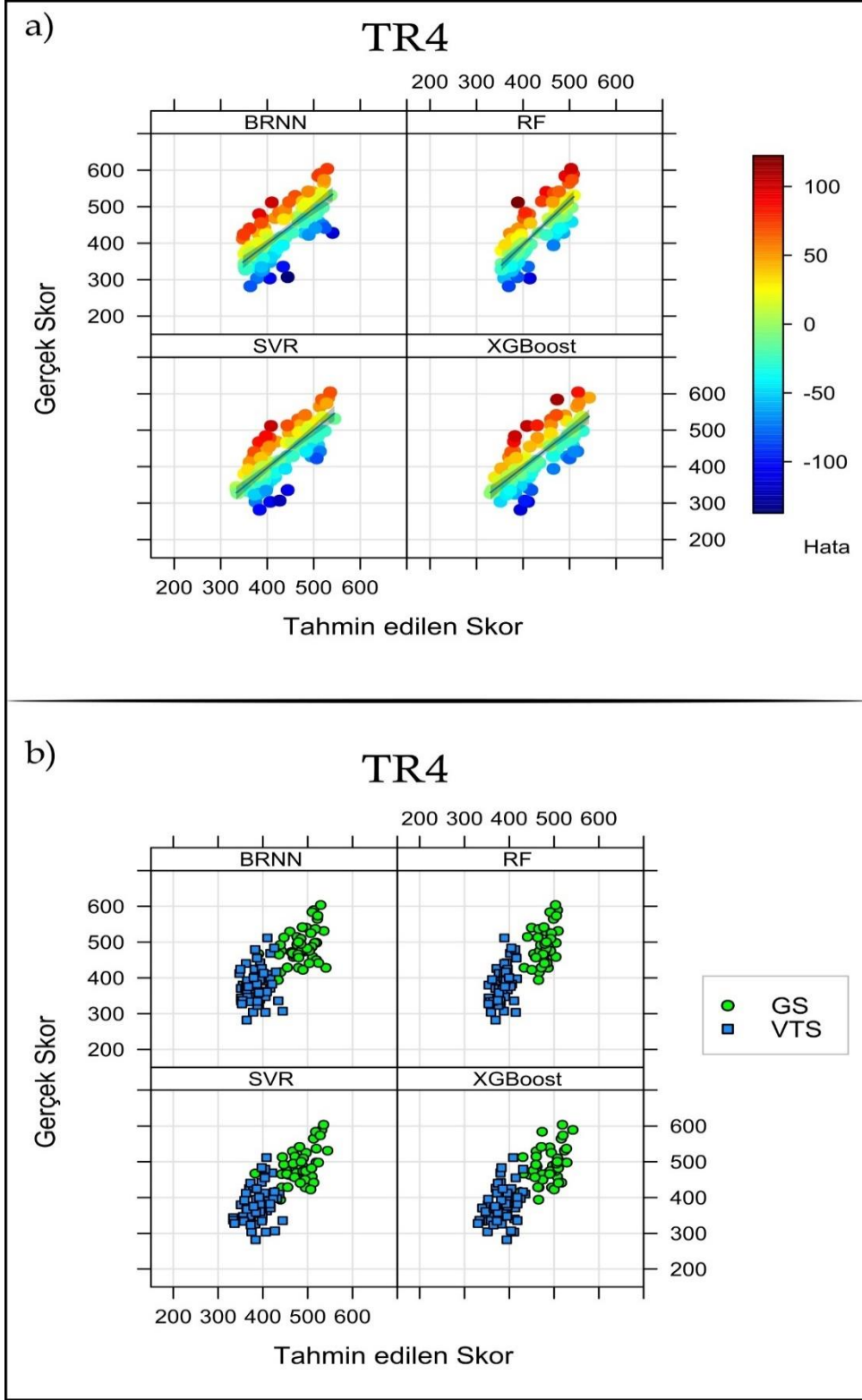
Şekil 20, Tablo 15’te elde edilen performans metriklerine ait çapraz doğrulama diyagramlarını tanımlamaktadır. Şekil 2a’da tüm modellerin hata dağılımlarının regresyon çizgisi boyunca ve yakınlarında dağıldığı görülmektedir. Bu durum tahmin hatalarının dağılımının normal yapıda olduğunu ve hata ortalamasının da sıfıra yakın olduğunu göstermektedir. Bu sonuçlar Tablo 15 de elde edilen bulgular ile uyumludur. Bunlarla birlikte XGBoost modelinin yüksek tahmin performansı sergilediği, Şekil 20a’ daki çapraz doğrulama diyagramının hata dağılımlarının regresyon çizgisi etrafında sıkı şekilde toplanmasından dolayı daha açık bir şekilde görülmektedir.

Şekil 20b, aynı test verileri için modellerin farklı okul türlerine (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim) ait tahmin edilen ve gerçek skorlarının karşılaştırmasını göstermektedir. Burada özellikle tanımlayıcı regresyon çizgisinden sapan ve modellerin tahmin etme performansının düşmesine neden olan okul grubunun Mesleki ve Teknik Ortaöğretim (VTS) okul türü olduğu dikkat çekmektedir. Bununla birlikte XGBoost, her tüm okul türlerinde de tahmin edilen skorların gerçek skorlarla en uyumlu olduğu model olarak öne çıkmaktadır. Ayrıca TR4 bölgesindeki Fen başarı skorları için genel ortaokul (GS) başarı düzeyinin VTS okul türüne göre belirgin şekilde artış göstermiş olması dikkat çekicidir. Özellikle okul türünün modellerin oluşturulmasında bir değişken olarak kullanılmamış olmasına rağmen, makine öğrenme algoritmasının bu başarı farklılıklarını diğer değişkenlerin yardımı ile tespit edebilmesi ve tahminlerini buna göre gerçekleştirmesi önemli bir başarı olduğu düşünülmektedir. Bu bulgular, XGBoost modelinin eğitim verileri üzerindeki güçlü performansını ve genelleştirme kabiliyetini destekler niteliktedir. XGBoost’un, farklı okul türlerindeki öğrencilerin performansını doğru bir şekilde tahmin edebilmesi, eğitim politikaları ve müdahalelerinin planlanmasında doğru tahmin skorları önerebileceğini ve bu konuda değerli bir araç olarak kullanılabileceğini göstermektedir. Özellikle Mesleki ve Teknik Ortaöğretim (VTS) öğrencileri için tahminlerdeki sapmalara rağmen, XGBoost’un genel olarak diğer modellere göre daha yüksek doğrulukla sonuçlar vermesi, bu modelin TR4 bölgesi için geniş bir yelpazede güvenilir tahminler oluşturabileceğini göstermektedir.

Şekil 21, farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost) tahmin ettiği başarı skorları ile gerçek başarı skorlarının karşılaştırmalı kutu grafiğini sunmaktadır. Şekil 20 de sunulan çapraz doğrulama bulgularına göre oluşturulmuş bu grafikler, modellerin tahmin performansını ve gerçek verilerle uyumunu değerlendirmek için önemli

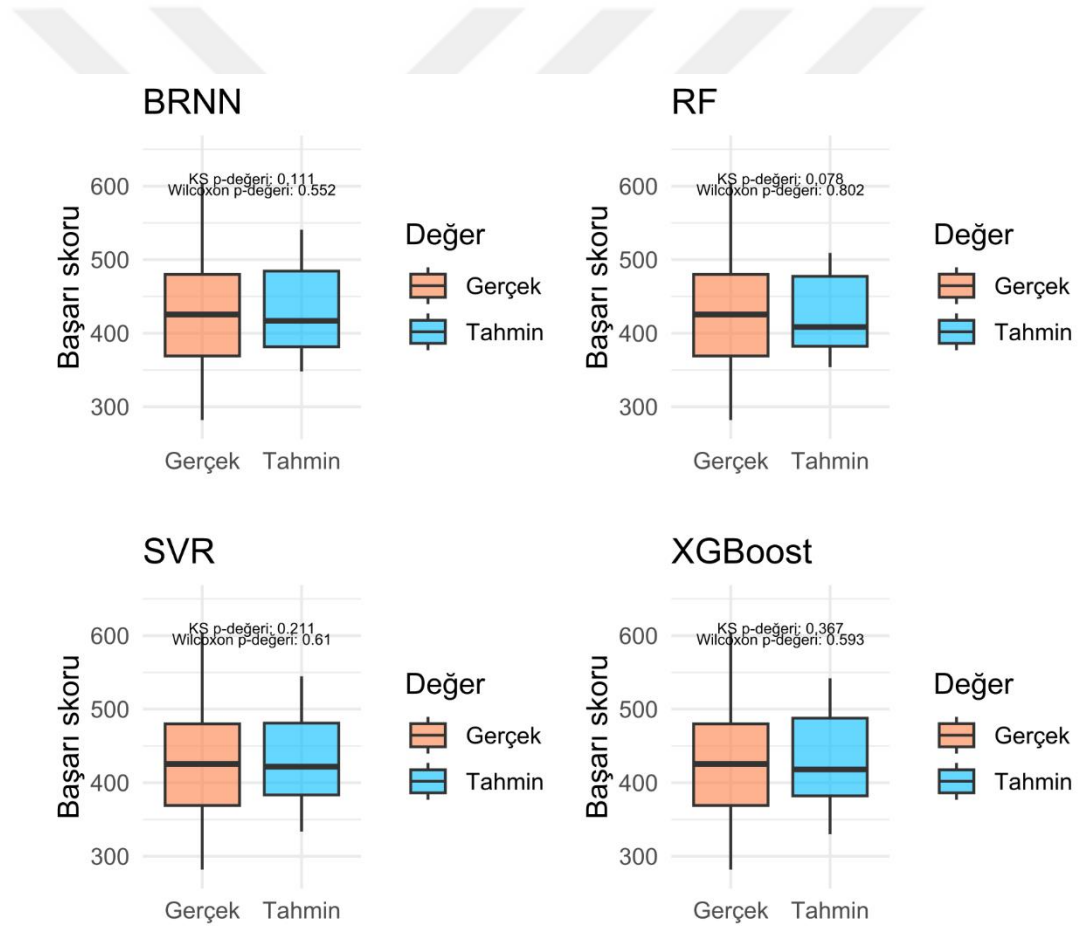
bir yaklaşım sağlamaktadır. Buna göre tüm modellerin tahmin skorlarının hem Kolmogorov-Smirnov testi (KS) hem de Wilcoxon sıralama toplamı testine göre gerçek skorların genel dağılımlarından ve merkezi eğilimlerinden anlamlı ölçüde bir farklılığı ($p > 0.05$) olmadığı tespit edilmiştir. Bu durum makine öğrenmesi modellerinin tahmin edilen ve gerçek başarı skorları arasındaki uyumun yüksek olduğunu ortaya koymaktadır.





Şekil 20: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)

Şekil 21 en yüksek performans seviyesi gösteren RF algoritması açısından değerlendirildiğinde, kutu grafiklerinin gerçek ve tahmin skorları için temsil ettiği ortalama ve medyan değerlerinin neredeyse örtüştüğü ve dağılımların (kutu dışındaki uç örnekler için yeterli seviyelerde olmasa da) Q1-Q3 aralığı için oldukça yakın olduğunu görülmektedir. Şekil 21’deki bulgulara göre tahmin edilen ve gerçek skorların genel dağılımlarının (KS p-değeri 0.078) ve merkezi eğilimlerinin (Wilcoxon p-değeri 0.802), istatistiksel olarak birbirlerinden farklı olmadığı belirlenmiştir. Bu durum, XGBoost algoritmasının TR4 verileri üzerinde en yüksek doğruluk ve güvenilirliği sunmakla birlikte hem bireysel gerçek skorları hem de gerçek skorlara ait dağılım yapısını doğru şekilde tahmin edebilme yeteneğini ortaya koymaktadır.



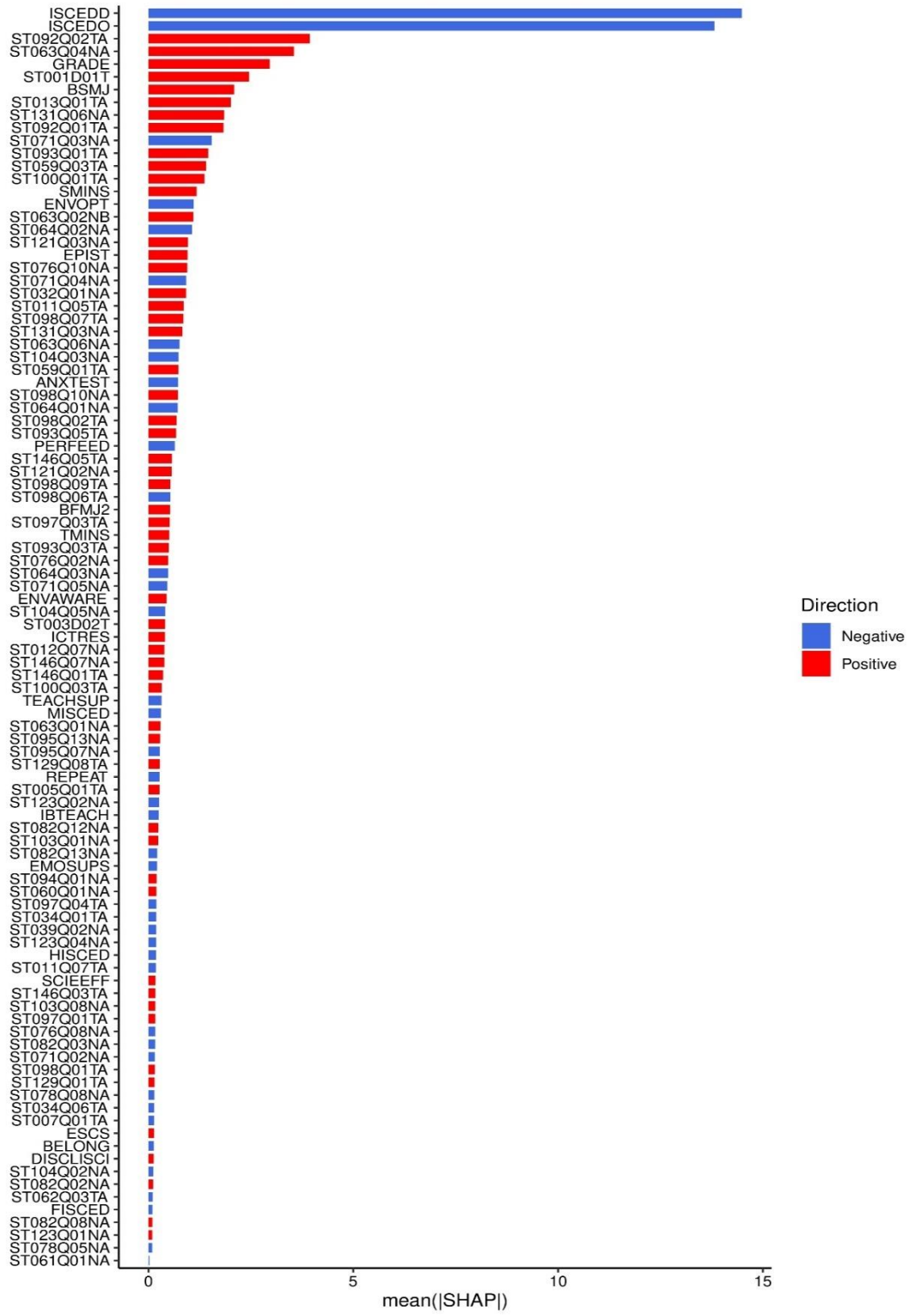
Şekil 21: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.

4.4.4 TR4 Bölgesi İçin Değişkenlerin Önem Seviyeleri

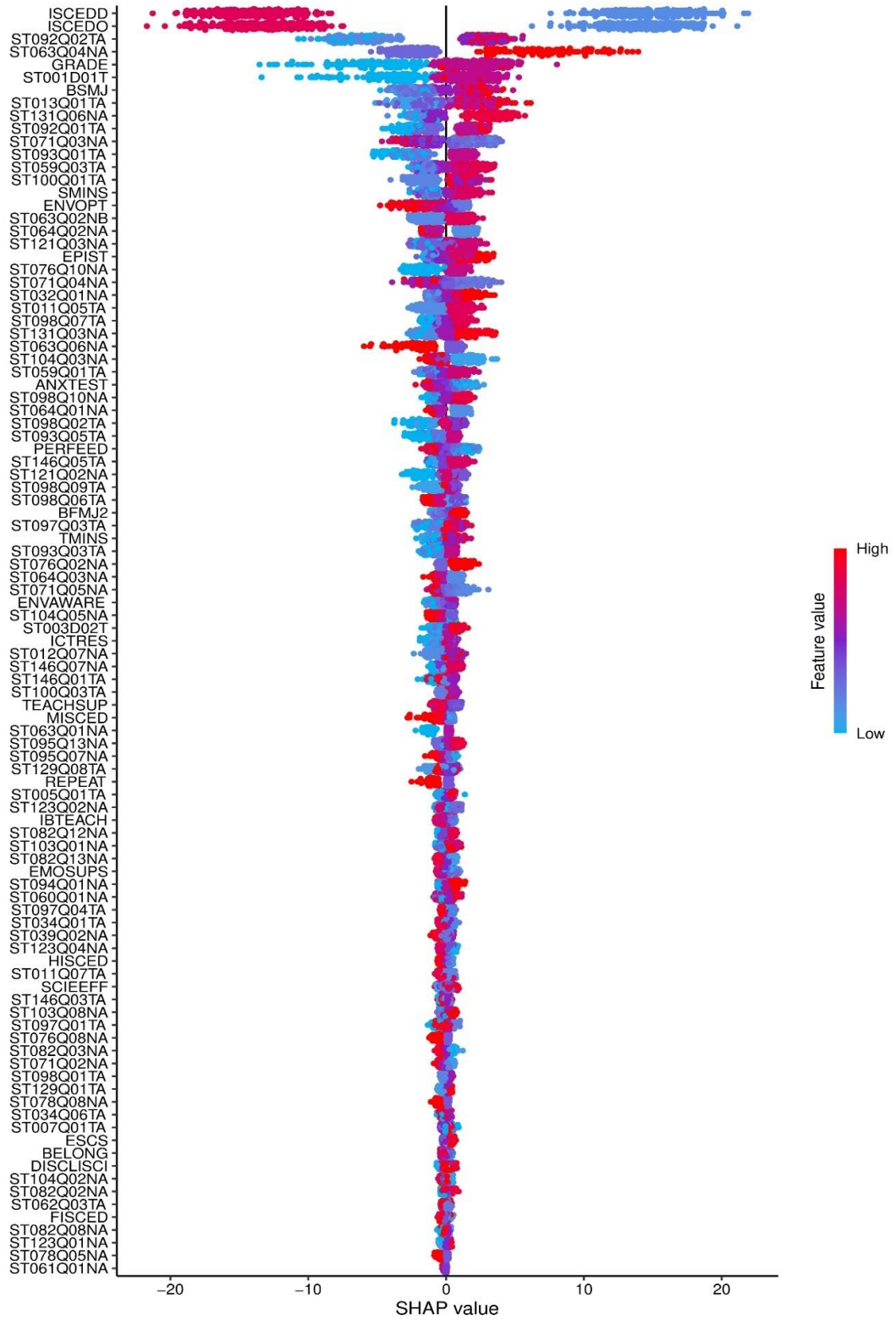
PISA 2015 Türkiye örnekleme TR4 bölgesi Fen okuryazarlık alanı için en iyi tahmin performans gösteren RF algoritmasına göre SHAP (SHapley Additive exPlanations) değerleri kullanılarak değişkenlerin önem parametreleri belirlenmiştir. Sonuçlar her bir değişkenin ortalama SHAP değerlerinin mutlak değerine göre azalan sırada Tablo 16’da sıralanmıştır. Bu sıralamaya göre diğer bölgelerdeki değişkenlerin karşılaştırması ise Kısım 4.13’te Jaccard indeksine göre yapılacaktır. Burada model için önemli olan sıralı değişkenler üç önem grubuna ayrılmış ve bu işlem model için kullanılan toplam değişkenleri 3 eşit parçaya ayırarak gerçekleştirilmiştir. Burada yüksek derecede önemli değişkenler 1. Grup (ilk 33 tanesi), orta derecede önemli değişkenler 2. Grup (34- 66) ve düşük derecede önemli değişkenler 3. Grup (66- 99) olarak etiketlenmiştir. Ayrıca Şekil 22 ‘ile Tablo 16 de özetlenen değişkenlerin ortalama SHAP değerleri görselleştirilmiştir. Bu şekil her bir tahmin edici değişkenin model tahminlerine olan ortalama etkisini ve yönünü (pozitif ve negatif katkıları) vurgulamaktadır. Şekil 23 her bir değişkenin test veri setindeki her bir gözlemdeki etkisini göstermektedir. Bu şekil içerisinde her bir nokta, bir veri noktasının bir özelliği için SHAP değerini temsil eder. Bu şeklin sağ tarafında pozitif SHAP değerleri, sol tarafında ise negatif SHAP değerleri bulunmaktadır. Tahmin edici değişkenlerin model çıktısına olan etkisi, bu SHAP değerlerinin büyüklüğüne göre belirlenir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Tablo3, Şekil 22 ve 23’teki bulgulara göre, TR4 bölgesi için model tahmininde önem seviyesi yüksek ilk 10 değişken, parantez içleri SHAP değerlerini ve model katkı yönlerini göstermek üzere, şu şekilde sıralanmıştır: ISCEDD (14.483, negatif), ISCEDO (13.816, negatif), ST092Q02TA (3.940, pozitif), ST063Q04NA (3.552, pozitif), GRADE (2.963, pozitif), ST001D01T (2.460, pozitif), BSMJ (2.091, pozitif), ST013Q01TA (2.009, pozitif), ST131Q06NA (1.848, pozitif), ST092Q01TA (1.830, pozitif). Şekil 4’teki Kırmızı ve mavi renkler, sırasıyla pozitif ve negatif yönlü etkileri temsil etmektedir. Bu, modelin tahminlerinin hangi özellikler tarafından ne kadar ve hangi yönde etkilendiğini daha net bir şekilde anlaşılmasına yardımcı olmaktadır. Buna göre negatif özellikteki değişkenlerin (ISCEDD, ISCEDO) değerlerindeki artış öğrencilerin Fen başarı skorlarını düşürmekte buna karşı pozitif özellikteki değişkenlerin değerlerindeki artış ise öğrencilerin Fen başarı skorlarını arttırıcı yönde etki etmektedir.

Tablo 16: TR4 (Doğu Marmara) verileri için oluşturulan En yüksek performanslı algoritmaya (RF) göre değişkenlerin önem seviyelerinin sıralanması.

ID	Değişken	SHAP değeri	Önem Grubu	ID	Değişken	SHAP değeri	Önem Grubu
1	ISCEDD	14.483	Grup 1	55	TEACHSUP	0.318	Grup 2
2	ISCEDO	13.816	Grup 1	56	MISCED	0.308	Grup 2
3	ST092Q02TA	3.940	Grup 1	57	ST063Q01NA	0.291	Grup 2
4	ST063Q04NA	3.552	Grup 1	58	ST095Q13NA	0.285	Grup 2
5	GRADE	2.963	Grup 1	59	ST095Q07NA	0.280	Grup 2
6	ST001D01T	2.460	Grup 1	60	ST129Q08TA	0.279	Grup 2
7	BSMJ	2.091	Grup 1	61	REPEAT	0.268	Grup 2
8	ST013Q01TA	2.009	Grup 1	62	ST005Q01TA	0.267	Grup 2
9	ST131Q06NA	1.848	Grup 1	63	ST123Q02NA	0.260	Grup 2
10	ST092Q01TA	1.830	Grup 1	64	IBTEACH	0.250	Grup 2
11	ST071Q03NA	1.547	Grup 1	65	ST082Q12NA	0.240	Grup 2
12	ST093Q01TA	1.460	Grup 1	66	ST103Q01NA	0.237	Grup 2
13	ST059Q03TA	1.408	Grup 1	67	ST082Q13NA	0.213	Grup 3
14	ST100Q01TA	1.368	Grup 1	68	EMOSUPS	0.211	Grup 3
15	SMINS	1.176	Grup 1	69	ST094Q01NA	0.200	Grup 3
16	ENVOPT	1.102	Grup 1	70	ST060Q01NA	0.195	Grup 3
17	ST063Q02NB	1.094	Grup 1	71	ST097Q04TA	0.194	Grup 3
18	ST064Q02NA	1.065	Grup 1	72	ST034Q01TA	0.190	Grup 3
19	ST121Q03NA	0.963	Grup 1	73	ST039Q02NA	0.188	Grup 3
20	EPIST	0.956	Grup 1	74	ST123Q04NA	0.186	Grup 3
21	ST076Q10NA	0.948	Grup 1	75	HISCED	0.185	Grup 3
22	ST071Q04NA	0.921	Grup 1	76	ST011Q07TA	0.182	Grup 3
23	ST032Q01NA	0.920	Grup 1	77	SCIEEFF	0.175	Grup 3
24	ST011Q05TA	0.858	Grup 1	78	ST146Q03TA	0.171	Grup 3
25	ST098Q07TA	0.852	Grup 1	79	ST103Q08NA	0.167	Grup 3
26	ST131Q03NA	0.827	Grup 1	80	ST097Q01TA	0.166	Grup 3
27	ST063Q06NA	0.762	Grup 1	81	ST076Q08NA	0.166	Grup 3
28	ST104Q03NA	0.736	Grup 1	82	ST082Q03NA	0.164	Grup 3
29	ST059Q01TA	0.733	Grup 1	83	ST071Q02NA	0.157	Grup 3
30	ANXTEST	0.722	Grup 1	84	ST098Q01TA	0.156	Grup 3
31	ST098Q10NA	0.718	Grup 1	85	ST129Q01TA	0.149	Grup 3
32	ST064Q01NA	0.716	Grup 1	86	ST078Q08NA	0.146	Grup 3
33	ST098Q02TA	0.688	Grup 1	87	ST034Q06TA	0.141	Grup 3
34	ST093Q05TA	0.676	Grup 2	88	ST007Q01TA	0.137	Grup 3
35	PERFEED	0.645	Grup 2	89	ESCS	0.135	Grup 3
36	ST146Q05TA	0.572	Grup 2	90	BELONG	0.129	Grup 3
37	ST121Q02NA	0.565	Grup 2	91	DISCLISCI	0.127	Grup 3
38	ST098Q09TA	0.534	Grup 2	92	ST104Q02NA	0.122	Grup 3
39	ST098Q06TA	0.532	Grup 2	93	ST082Q02NA	0.116	Grup 3
40	BFMJ2	0.527	Grup 2	94	ST062Q03TA	0.103	Grup 3
41	ST097Q03TA	0.520	Grup 2	95	FISCED	0.097	Grup 3
42	TMINS	0.506	Grup 2	96	ST082Q08NA	0.096	Grup 3
43	ST093Q03TA	0.500	Grup 2	97	ST123Q01NA	0.093	Grup 3
44	ST076Q02NA	0.482	Grup 2	98	ST078Q05NA	0.092	Grup 3
45	ST064Q03NA	0.481	Grup 2	99	ST061Q01NA	0.021	Grup 3
46	ST071Q05NA	0.461	Grup 2				
47	ENVAWARE	0.444	Grup 2				
48	ST104Q05NA	0.409	Grup 2				
49	ST003D02T	0.406	Grup 2				
50	ICTRES	0.402	Grup 2				
51	ST012Q07NA	0.388	Grup 2				
52	ST146Q07NA	0.388	Grup 2				
53	ST146Q01TA	0.358	Grup 2				
54	ST100Q03TA	0.322	Grup 2				



Şekil 22: TR4 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.



Şekil 23: TR4 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.

4.5 TR5 Bölgesi PISA-2015 Bulguları

4.5.1 TR5 Bölgesi Hakkında Temel Bilgiler

2015 PISA uygulamasının Türkiye örnekleminin beşinci bölgesi TR5 olarak adlandırılmıştır. TR5 bölgesi 2015 PISA uygulamasında ayrılan bölgelerden Batı Anadolu'yu temsil etmektedir. Batı Anadolu olarak ayrılan bölge Konya, Karaman ve Ankara'yı kapsamaktadır. TR5 bölgesi (Batı Anadolu) bölgesinden 7. Sınıf ve üzeri düzeyde eğitim gören 553 öğrencinin 2015 PISA sınavına katıldığı görülmektedir. 2015 PISA sınavına katılan tüm Türkiye örnekleminin %9.38'inin Batı Anadolu bölgesinden katıldığını görmekteyiz.

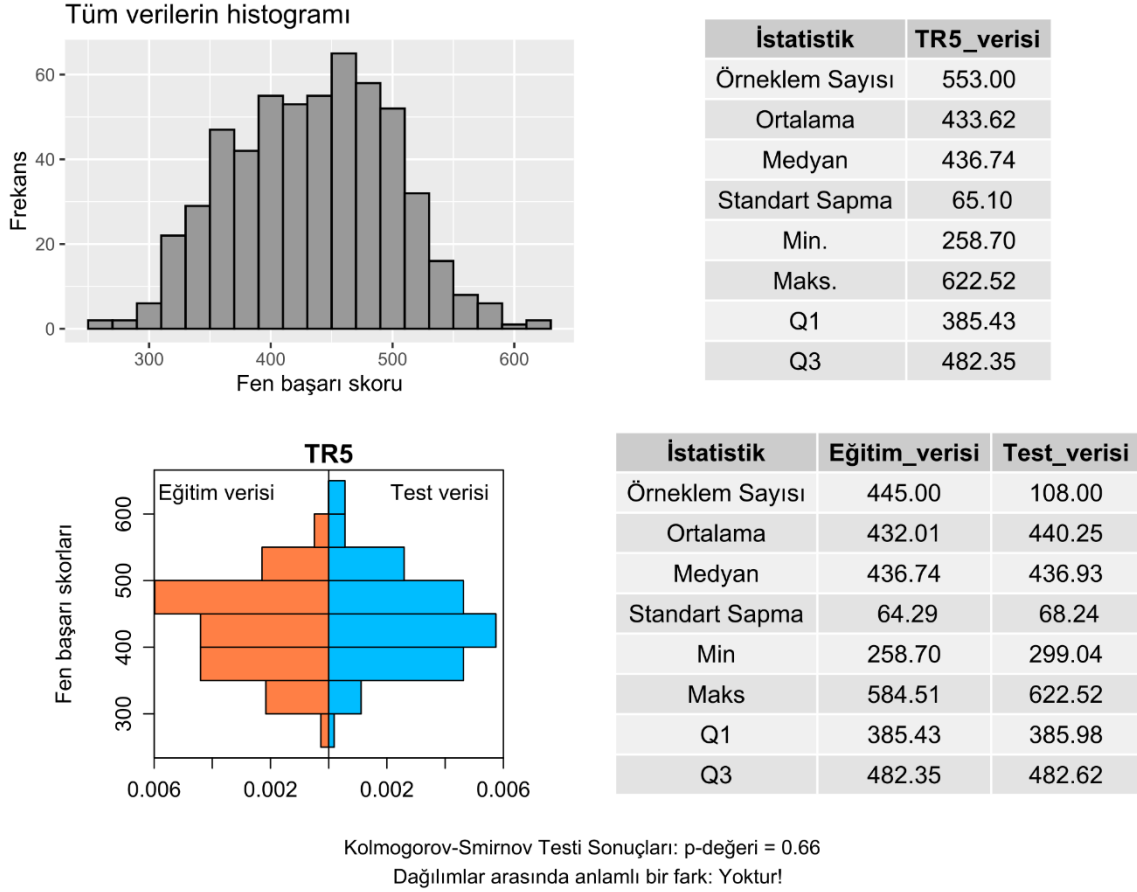
4.5.2 TR5 Bölgesinin Veri Yapısı

Şekil 24'te, Türkiye'nin TR5 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özeti sunulmaktadır. Şekil 24'ün üst panelinde, tüm veri setini temsil eden bir histogram ve istatistiksel özet bilgileri gösterilmektedir. Aynı şeklin alt panelinde ise, makine öğrenmesi modelinin oluşturulması ve modelin performansının değerlendirilmesi süreçlerinde kullanılacak olan eğitim ve test alt veri setlerine ait istatistiksel bilgiler bulunmaktadır. Bu alt veri setleri ana veri setinden rastgele bir şekilde ve literatüre uygun olarak %80 (Eğitim verisi) / %20 (Test verisi) oranında örneklenerek oluşturulmuştur (Gholamy ve diğerleri, 2018).

Şekil 24'ün üst panelindeki histograma göre, fen puanlarının ortalama puan etrafında merkezi bir eğilimle dağıldığı görülmektedir. Bu TR5 bölgesindeki öğrencilerin (553 kişi) Fen başarı skorlarının normal dağılıma uyduğunu belirtmektedir. Ayrıca bu bölge için tüm veri setine ait ortalama puan 433.62, medyan 436.74, standart sapma 65.10 olarak hesaplanmıştır. Bunlara ek olarak minimum ve maksimum puanlar sırasıyla 258.70 ve 622.52 olup, birinci çeyrek (Q1) 385.43 ve üçüncü çeyrek (Q3) 482.35 olarak belirlenmiştir.

Şekil 24'ün alt panelinde, eğitim veri setinin (445 kişi) ortalama puanı 432.01, medyanı 436.74 ve standart sapması 64.29 olarak tespit edilmiştir. Bunun yanında eğitim verisi için minimum ve maksimum puanlar sırasıyla 258.70 ve 584.51 olup birinci çeyrek (Q1) 385.43 ve üçüncü çeyrek (Q3) 482.35 olarak belirlenmiştir. Bunların yanında test veri setinin (108) ortalama puanı 440.25, medyanı 436.93 ve standart sapması 68.24 olarak tespit edilmiştir. Test verisi için minimum ve maksimum puanlar sırasıyla 299.04 ve 622.52 olup, birinci çeyrek (Q1) 385.98 ve üçüncü çeyrek (Q3) 482.62 olarak belirlenmiştir. Bu veri seti

için rasgele bir şekilde oluşturulan eğitim ve test alt veri setlerine ait dağılımlarının birbirinden anlamlı bir şekilde farklı olmadığı Kolmogorov-Smirnov testi sonucuna göre ($p > 0.05$) tespit edilmiştir.



Şekil 24: Türkiye'nin TR5 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)

4.5.3 TR5 Bölgesinde Makine Öğrenme Modeli

Tablo 17 TR5 (Batı Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı ve 10 tekrarlı çapraz doğrulama performans metriklerini ve optimize edilmiş Hiperparametre değerlerini göstermektedir. Bu tabloda elde edilen bulgular, XGBoost, RF, SVR ve BRNN algoritmaları ile Kısım 3.3 oluşturulan makine öğrenme modellerin 10 tekrarlı ve 10 katlı çapraz doğrulama tahmin performansları elde edilmiştir. Bu tablonun ayrıntılı sonuçları ve kısaltmalar için gerekli açıklamalar EK-1 içerisinde sunulmuştur. İncelenen ve oluşturulan makine öğrenmesi modellerin başarı

performansları RMSE, R^2 ve MAE metrikleri üzerinden değerlendirilmiş ve SVR modelinin en düşük RMSE (46.63), en yüksek R^2 (0.47) ve en düşük MAE (37.29) değerleri ile eğitim verileri için en iyi performansı gösterdiği tespit edilmiştir. Buna karşın BRNN, XGBoost ve RF modelleri sırasıyla 47.83, 48.37 ve 50.11 RMSE değerleri ile SVR algoritmasının ardından sıralanmışlardır. Bu veriler, TR5 bölgesinin eğitim verileri için SVR modelinin Fen okuryazarlık skorlarının tahmininde diğer modellere göre daha iyi bir öğrenme ortaya koyduğunu göstermektedir.

Tablo 17: TR5 (Batı Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri

Model	RMSE	R^2	MAE	Hiperparametre değerleri
SVR	46.63	0.47	37.29	sigma 0.004120901 C 2.023479867
BRNN	47.83	0.45	37.86	neurons 1
XGBoost	48.37	0.43	38.85	nrounds 511.00000000 max_depth 2.00000000 eta 0.04248939 gamma 3.73233878 colsample_bytree 0.65200718 min_child_weight 0.00000000 subsample 0.96298925
RF	50.11	0.43	40.62	mtry "51" splitrule "maxstat" min.node.size "4"

Tablo 18'de TR5 bölgesine ait verileri için oluşturulan makine öğrenme algoritmalarının test verisi performans göstergeleri detaylı bir şekilde sunulmaktadır. Bu tabloya göre, XGBoost modelinin RMSE değeri 52.04, R^2 değeri 0.42 ve MAE değeri 41.71 olarak belirlenmiş ve diğer modellerden daha üstün bir performans sergilediği tespit edilmiştir. Ayrıca, XGBoost modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.585, güven aralığının alt sınırı 0.470 ve üst sınırı 0.681 ile yüksek bir uyumu gösterdiği belirlenmiştir. Buna karşın ikinci en iyi performans gösteren RF modelinde, RMSE değeri 54.56, R^2 değeri 0.42 ve MAE değeri 43.12 ile XGBoost'a göre daha düşük olup, uyum korelasyon katsayısı merkezi tahmini 0.443 (lwr.ci: 0.351 - upr.ci: 0.527) olarak hesaplanmıştır. SVR ve BRNN modellerinde ise sırasıyla RMSE değerleri 54.58 ve 55.64, R^2 değerleri 0.37 ve 0.33, MAE değerleri 43.57 ve 43.85 olarak belirlenmiştir. SVR modelinin uyum korelasyon katsayısı merkezi tahmini 0.565 (lwr.ci: 0.437- upr.ci: 0.670)

iken, BRNN modelinin uyum korelasyon katsayısı merkezi tahmini 0.526 (lwr.ci: 0.396-upr.ci: 0.636) olarak hesaplanmıştır. Bu sonuçlar, XGBoost modelinin en düşük hata oranları ve en yüksek açıklama gücü ile diğer modellerden daha üstün performans sergilediğini ortaya koymaktadır.

Tablo 18: TR5 (Batı Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri

Model	RMSE	R^2	MAE	ρ_c
XGBoost	52.04	0.42	41.71	XGBoost.est 0.5853299 XGBoost.lwr.ci 0.4696117 XGBoost.upr.ci 0.6812766
RF	54.56	0.42	43.12	RF.est 0.4434270 RF.lwr.ci 0.3511378 RF.upr.ci 0.5271832
SVR	54.58	0.37	43.57	SVR.est 0.5648426 SVR.lwr.ci 0.4369505 SVR.upr.ci 0.6703494
BRNN	55.64	0.33	43.85	BRNN.est 0.5265765 BRNN.lwr.ci 0.3958956 BRNN.upr.ci 0.6363527

ρ_c = Uyum Korelasyon Katsayısı (Concordance Correlation Coefficient)

est = ρ_c için merkezi tahmin edilen değeri

lwr.ci = güven aralığının alt sınırını (lower confidence interval)

upr.ci = güven aralığının üst sınırını (upper confidence interval)

Tablo 17 ve Tablo 18 verilerine göre yapılan genel değerlendirmede, SVR modeli eğitim verilerinde en iyi sonucu vermiş olsa da XGBoost modelinin test verilerinde en yüksek performansı sergilediği görülmektedir. Bu durum, modelin genelleştirme ve tahmin kabiliyetinin güçlü olduğunu göstermektedir. Ayrıca test verilerinin uyum korelasyon katsayısının (ρ_c) yüksek olması, modelin tahminlerinin gerçek verilerle yüksek uyum gösterdiğini belirtmektedir. Bu, XGBoost modelinin hem eğitim hem de test verilerinde tutarlı ve güvenilir bir performans sergilediğinin kanıtıdır.

Diğer yandan RF, SVR ve BRNN modelleri ise eğitim ve test verileri arasında benzer hata oranlarına sahip olsalar da performansları XGBoost modellerine kıyasla daha düşük seviyelerde kalmıştır. Bu durum SVR, BRNN ve RF modellerinin genelleştirme kabiliyetinin sınırlı olduğunu ve yeni veriler üzerinde daha düşük performans sergilediğini göstermektedir. Her üç modelin de ρ_c değerleri, tahminlerinin gerçek verilerle makul bir uyum gösterdiğini belirtmektedir.

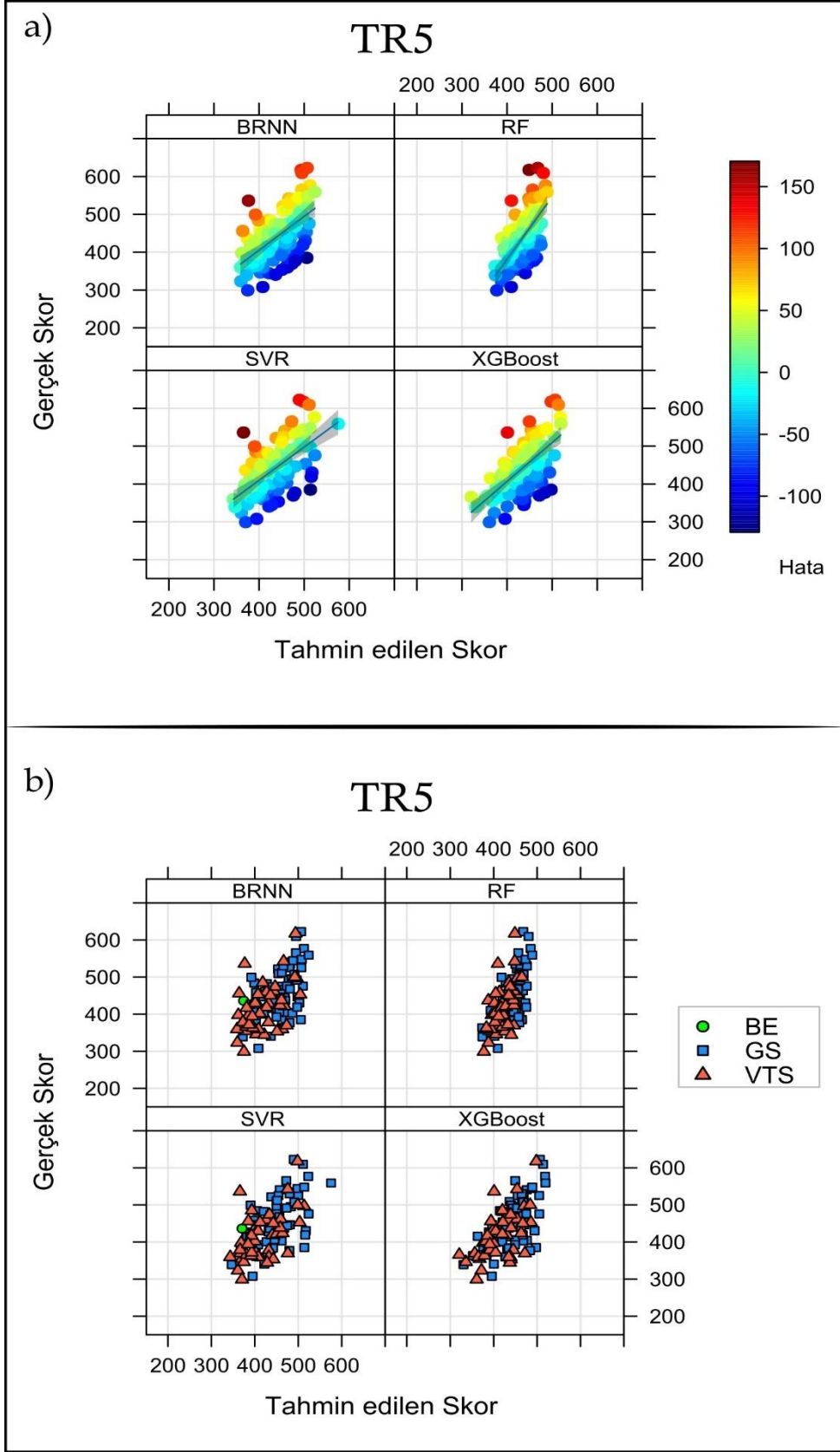
Şekil 25, Tablo 18’de elde edilen performans metriklerine ait çapraz doğrulama diyagramlarını tanımlamaktadır. Şekil 25a’da tüm modellerin hata dağılımlarının regresyon çizgisi boyunca ve yakınlarında dağıldığı görülmektedir. Bu durum tahmin hatalarının dağılımının normal yapıda olduğunu ve hata ortalamasının da sıfıra yakın olduğunu göstermektedir. Bu sonuçlar Tablo 18 de elde edilen bulgular ile uyumludur. Bunlarla birlikte XGBoost modelinin yüksek tahmin performansı sergilediği, Şekil 25a’daki çapraz doğrulama diyagramının hata dağılımlarının regresyon çizgisi etrafında sıkı şekilde toplanmasından dolayı daha açık bir şekilde görülmektedir.

Şekil 25b, aynı test verileri için modellerin farklı okul türlerine (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim) ait tahmin edilen ve gerçek skorlarının karşılaştırmasını göstermektedir. Burada özellikle tanımlayıcı regresyon çizgisinden sapan ve modellerin tahmin etme performansının düşmesine neden olan okul grubunun Mesleki ve Teknik Ortaöğretim (VTS) okul türü olduğu dikkat çekmektedir. Bununla birlikte XGBoost, her tüm okul türlerinde de tahmin edilen skorların gerçek skorlarla en uyumlu olduğu model olarak öne çıkmaktadır. Ayrıca TR5 bölgesindeki Fen başarı skorları için genel ortaokul (GS) başarı düzeyinin VTS okul türüne göre belirgin şekilde artış göstermiş olması dikkat çekicidir. Özellikle okul türünün modellerin oluşturulmasında bir değişken olarak kullanılmamış olmasına rağmen, makine öğrenme algoritmasının bu başarı farklılıklarını diğer değişkenlerin yardımı ile tespit edebilmesi ve tahminlerini buna göre gerçekleştirmesi önemli bir başarı olduğu düşünülmektedir. Bu bulgular, XGBoost modelinin eğitim verileri üzerindeki güçlü performansını ve genelleştirme kabiliyetini destekler niteliktedir. XGBoost’un, farklı okul türlerindeki öğrencilerin performansını doğru bir şekilde tahmin edebilmesi, eğitim politikaları ve müdahalelerinin planlanmasında doğru tahmin skorları önerebileceğini ve bu konuda değerli bir araç olarak kullanılabileceğini göstermektedir. Özellikle Mesleki ve Teknik Ortaöğretim (VTS) öğrencileri için tahminlerdeki sapmalara rağmen, XGBoost’un genel olarak diğer modellere göre daha yüksek doğrulukla sonuçlar vermesi, bu modelin TR5 bölgesi için geniş bir yelpazede güvenilir tahminler oluşturabileceğini göstermektedir.

Şekil 26, farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost) tahmin ettiği başarı skorları ile gerçek başarı skorlarının karşılaştırmalı kutu grafiğini sunmaktadır. Şekil 25 de sunulan çapraz doğrulama bulgularına göre oluşturulmuş bu grafikler, modellerin tahmin performansını ve gerçek verilerle uyumunu değerlendirmek için önemli

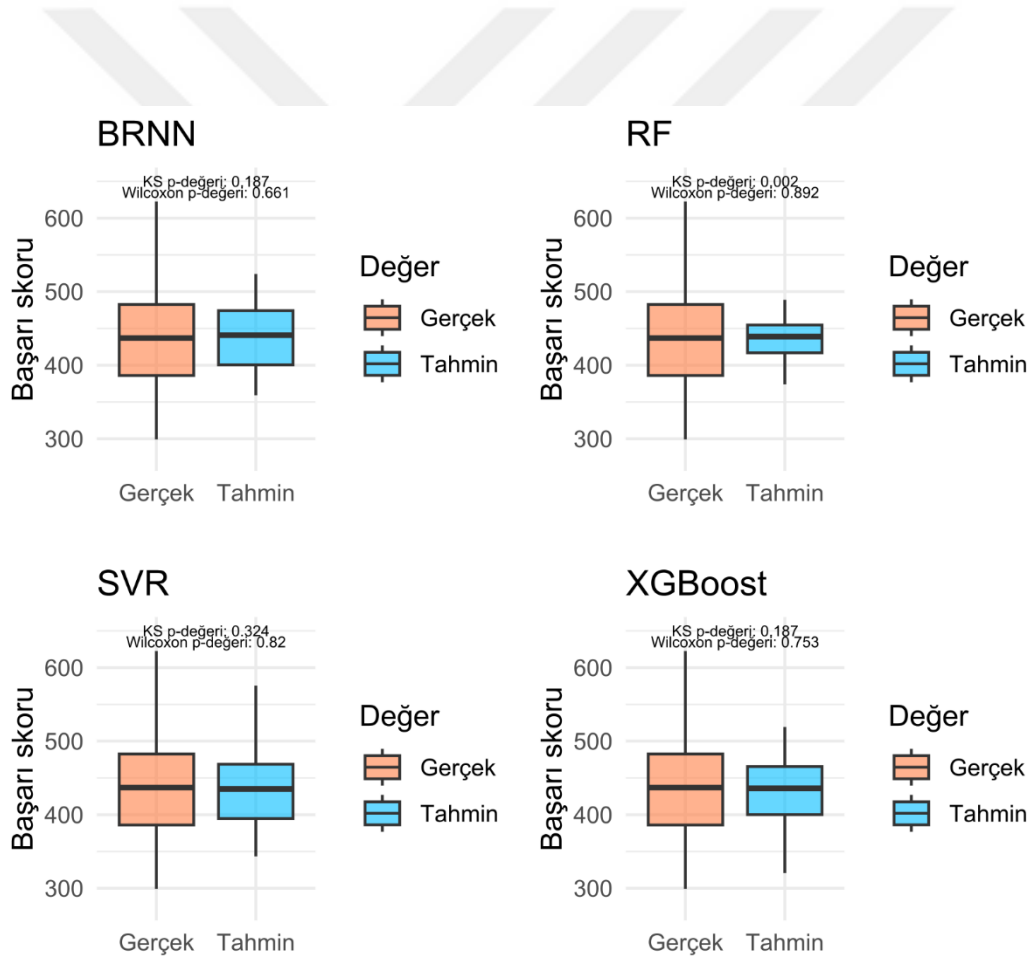
bir yaklaşımla sağlamaktadır. Buna göre RF hariç tüm modellerin tahmin skorlarının hem Kolmogorov -Simirnov testi (KS) hem de Wilcoxon sıralama toplamı testine göre gerçek skorların genel dağılımlarından ve merkezi eğilimlerinden anlamlı ölçüde bir farklılığı ($p > 0.05$) olmadığı tespit edilmiştir. Bu durum makine öğrenmesi modellerinin tahmin edilen ve gerçek başarı skorları arasındaki uyumun yüksek olduğunu ortaya koymaktadır.





Şekil 25: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)

Şekil 26 en yüksek performans seviyesi gösteren XGBoost algoritması açısından değerlendirildiğinde, Kutu grafiklerinin gerçek ve tahmin skorları için temsil ettiği ortalama ve medyan değerlerinin neredeyse örtüştüğü ve dağılımların (kutu dışındaki uç örnekler için yeterli seviyelerde olmasa da) Q1-Q3 aralığı için oldukça yakın olduğunu görülmektedir. Şekil 26'daki bulgulara göre tahmin edilen ve gerçek skorların genel dağılımlarının (KS p-değeri 0.187) ve merkezi eğilimlerinin (Wilcoxon p-değeri 0.763), istatistiksel olarak birbirlerinden farklı olmadığı belirlenmiştir. Bu durum, XGBoost algoritmasının TR5 verileri üzerinde en yüksek doğruluk ve güvenilirliği sunmakla birlikte hem bireysel gerçek skorları hem de gerçek skorlara ait dağılım yapısını doğru şekilde tahmin edebilme yeteneğini ortaya koymaktadır.



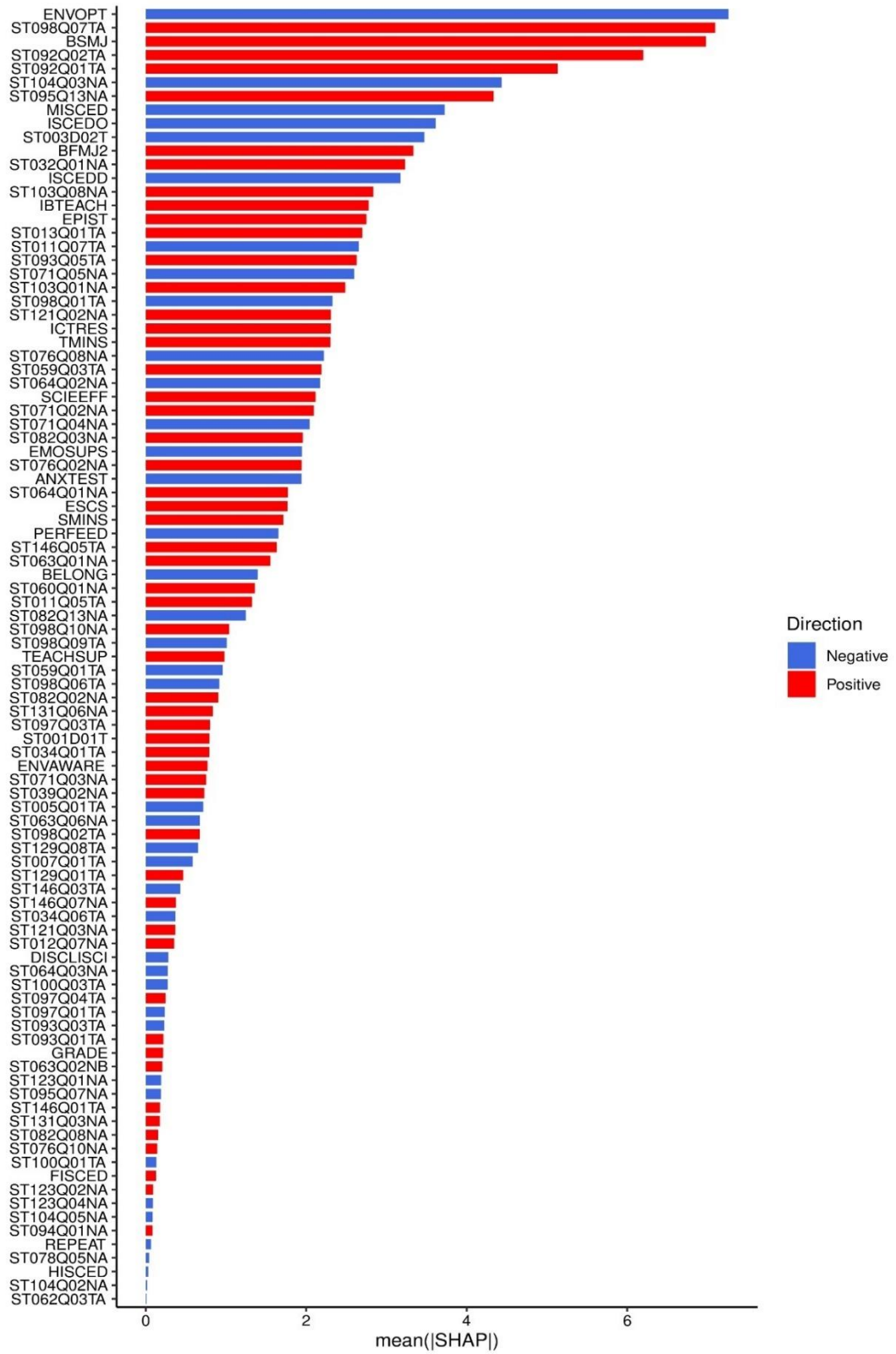
Şekil 26: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.

4.5.4 TR5 Bölgesi İçin Değişkenlerin Önem Seviyeleri

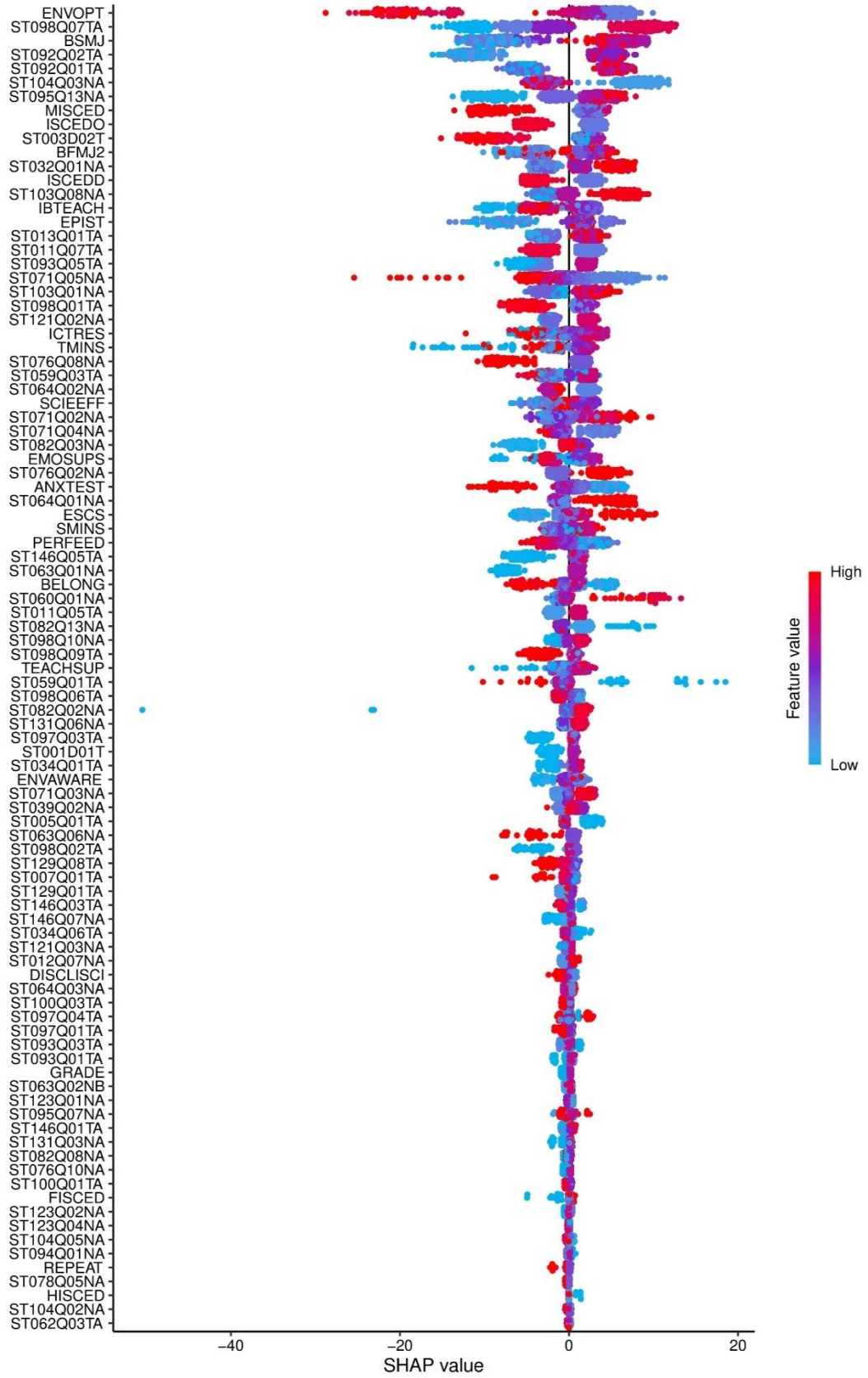
PISA 2015 Türkiye örnekleme TR5 bölgesi Fen okuryazarlık alanı için en iyi tahmin performansın gösteren XGBoost algoritmasına göre SHAP (SHapley Additive exPlanations) değerleri kullanılarak değişkenlerin önem parametreleri belirlenmiştir. Sonuçlar her bir değişkenin ortalama SHAP değerlerinin mutlak değerine göre azalan sırada Tablo 19’da, sıralanmıştır. Bu sıralamaya göre diğer bölgelerdeki değişkenlerin karşılaştırması ise Kısım 4.13’te Jaccard indeksine göre yapılacaktır. Burada model için önemli olan sıralı değişkenler üç önem grubuna ayrılmış ve bu işlem model için kullanılan toplam değişkenleri 3 eşit parçaya ayırarak gerçekleştirilmiştir. Burada yüksek derecede önemli değişkenler 1. Grup (ilk 33 tanesi), orta derecede önemli değişkenler 2. Grup (34 - 66) ve düşük derecede önemli değişkenler 3. Grup (67 - 99) olarak etiketlenmiştir. Ayrıca Şekil 27 ile Tablo 19 da özetlenen değişkenlerin ortalama SHAP değerleri görselleştirilmiştir. Bu şekil her bir tahmin edici değişkenin model tahminlerine olan ortalama etkisini ve yönünü (pozitif ve negatif katkıları) vurgulamaktadır. Şekil 28 her bir değişkenin test veri setindeki her bir gözlemdeki etkisini göstermektedir. Bu şekil içerisinde her bir nokta, bir veri noktasının bir özelliği için SHAP değerini temsil eder. Bu şeklin sağ tarafında pozitif SHAP değerleri, sol tarafında ise negatif SHAP değerleri bulunmaktadır. Tahmin edici değişkenlerin model çıktısına olan etkisi, bu SHAP değerlerinin büyüklüğüne göre belirlenir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Tablo19, Şekil 27 ve 28’deki bulgulara göre, TR5 bölgesi için model tahmininde önem seviyesi yüksek ilk 10 değişken, parantez içleri SHAP değerlerini ve model katkı yönlerini göstermek üzere, şu şekilde sıralanmıştır: ENVOPT (7.261,negatif), ST098Q07TA (7.096,pozitif), BSMJ (6.981,pozitif), ST092Q02TA (6.200,pozitif), ST092Q01TA (5.133,pozitif), ST104Q03NA (4.436,negatif), ST095Q13NA (4.334,pozitif), MISCED (3.726,negatif), ISCEDO (3.614,negatif), ST003D02T (3.472,negatif). Şekil 7 de Kırmızı ve mavi renkler, sırasıyla pozitif ve negatif yönlü etkileri temsil etmektedir. Bu, modelin tahminlerinin hangi özellikler tarafından ne kadar ve hangi yönde etkilendiğini daha net bir şekilde anlaşılmasına yardımcı olmaktadır. Buna göre negatif özellikteki değişkenlerin (ENVOPT, ST104Q03NA, MISCED, ISCEDO, ST003D02T) değerlerindeki artış öğrencilerin Fen başarı skorlarını düşürmekte buna karşı pozitif özellikteki değişkenlerin değerlerindeki artış ise öğrencilerin Fen başarı skorlarını artırıcı yönde etki etmektedir.

Tablo 19: TR5 (Batı Anadolu) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.

ID	Değişken	SHAP değeri	Önem Grubu	ID	Değişken	SHAP değeri	Önem Grubu
1	ENVOPT	7.261	Grup 1	55	ST034Q01TA	0.793	Grup 2
2	ST098Q07TA	7.096	Grup 1	56	ENVAWARE	0.770	Grup 2
3	BSMJ	6.981	Grup 1	57	ST071Q03NA	0.752	Grup 2
4	ST092Q02TA	6.200	Grup 1	58	ST039Q02NA	0.728	Grup 2
5	ST092Q01TA	5.133	Grup 1	59	ST005Q01TA	0.715	Grup 2
6	ST104Q03NA	4.436	Grup 1	60	ST063Q06NA	0.674	Grup 2
7	ST095Q13NA	4.334	Grup 1	61	ST098Q02TA	0.673	Grup 2
8	MISCED	3.726	Grup 1	62	ST129Q08TA	0.653	Grup 2
9	ISCEDO	3.614	Grup 1	63	ST007Q01TA	0.586	Grup 2
10	ST003D02T	3.472	Grup 1	64	ST129Q01TA	0.467	Grup 2
11	BFMJ2	3.336	Grup 1	65	ST146Q03TA	0.429	Grup 2
12	ST032Q01NA	3.231	Grup 1	66	ST146Q07NA	0.375	Grup 2
13	ISCEDD	3.174	Grup 1	67	ST034Q06TA	0.371	Grup 3
14	ST103Q08NA	2.836	Grup 1	68	ST121Q03NA	0.369	Grup 3
15	IBTEACH	2.777	Grup 1	69	ST012Q07NA	0.356	Grup 3
16	EPIST	2.749	Grup 1	70	DISCLSCI	0.281	Grup 3
17	ST013Q01TA	2.699	Grup 1	71	ST064Q03NA	0.275	Grup 3
18	ST011Q07TA	2.654	Grup 1	72	ST100Q03TA	0.275	Grup 3
19	ST093Q05TA	2.628	Grup 1	73	ST097Q04TA	0.248	Grup 3
20	ST071Q05NA	2.598	Grup 1	74	ST097Q01TA	0.237	Grup 3
21	ST103Q01NA	2.487	Grup 1	75	ST093Q03TA	0.230	Grup 3
22	ST098Q01TA	2.326	Grup 1	76	ST093Q01TA	0.220	Grup 3
23	ST121Q02NA	2.307	Grup 1	77	GRADE	0.216	Grup 3
24	ICTRES	2.306	Grup 1	78	ST063Q02NB	0.208	Grup 3
25	TMINS	2.301	Grup 1	79	ST123Q01NA	0.192	Grup 3
26	ST076Q08NA	2.220	Grup 1	80	ST095Q07NA	0.189	Grup 3
27	ST059Q03TA	2.190	Grup 1	81	ST146Q01TA	0.178	Grup 3
28	ST064Q02NA	2.174	Grup 1	82	ST131Q03NA	0.176	Grup 3
29	SCIEEFF	2.116	Grup 1	83	ST082Q08NA	0.156	Grup 3
30	ST071Q02NA	2.095	Grup 1	84	ST076Q10NA	0.143	Grup 3
31	ST071Q04NA	2.043	Grup 1	85	ST100Q01TA	0.132	Grup 3
32	ST082Q03NA	1.958	Grup 1	86	FISCED	0.129	Grup 3
33	EMOSUPS	1.946	Grup 1	87	ST123Q02NA	0.091	Grup 3
34	ST076Q02NA	1.943	Grup 2	88	ST123Q04NA	0.090	Grup 3
35	ANXTEST	1.941	Grup 2	89	ST104Q05NA	0.087	Grup 3
36	ST064Q01NA	1.772	Grup 2	90	ST094Q01NA	0.085	Grup 3
37	ESCS	1.767	Grup 2	91	REPEAT	0.066	Grup 3
38	SMINS	1.716	Grup 2	92	ST078Q05NA	0.041	Grup 3
39	PERFEED	1.654	Grup 2	93	HISCED	0.032	Grup 3
40	ST146Q05TA	1.632	Grup 2	94	ST104Q02NA	0.019	Grup 3
41	ST063Q01NA	1.554	Grup 2	95	ST062Q03TA	0.010	Grup 3
42	BELONG	1.396	Grup 2	96	ST063Q04NA	0	Grup 3
43	ST060Q01NA	1.358	Grup 2	97	ST078Q08NA	0	Grup 3
44	ST011Q05TA	1.325	Grup 2	98	ST082Q12NA	0	Grup 3
45	ST082Q13NA	1.247	Grup 2	99	ST061Q01NA	0	Grup 3
46	ST098Q10NA	1.039	Grup 2				
47	ST098Q09TA	1.008	Grup 2				
48	TEACHSUP	0.980	Grup 2				
49	ST059Q01TA	0.958	Grup 2				
50	ST098Q06TA	0.917	Grup 2				
51	ST082Q02NA	0.904	Grup 2				
52	ST131Q06NA	0.834	Grup 2				
53	ST097Q03TA	0.802	Grup 2				
54	ST001D01T	0.793	Grup 2				



Şekil 27: TR5 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.



Şekil 28: TR5 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.

4.6 TR6 Bölgesi PISA-2015 Bulguları

4.6.1 TR6 Bölgesi Hakkında Temel Bilgiler

2015 PISA uygulamasının Türkiye örnekleminin altıncı bölgesi TR6 olarak adlandırılmıştır. TR6 bölgesi Türkiye'nin Akdeniz Bölgesinde bulunan illeri kapsamaktadır. Akdeniz bölgesi Isparta, Burdur, Antalya, Mersin, Kahramanmaraş, Osmaniye ve Hatay illerinden oluşmaktadır. TR6 bölgesinden 7.sınıf ve üzeri düzeyde eğitim gören 817 öğrenci 2015 PISA sınavına katıldığı görülmektedir. 2015 PISA sınavına katılan tüm Türkiye örnekleminin %13.86'sının Akdeniz Bölgesinden katıldığını görmekteyiz.

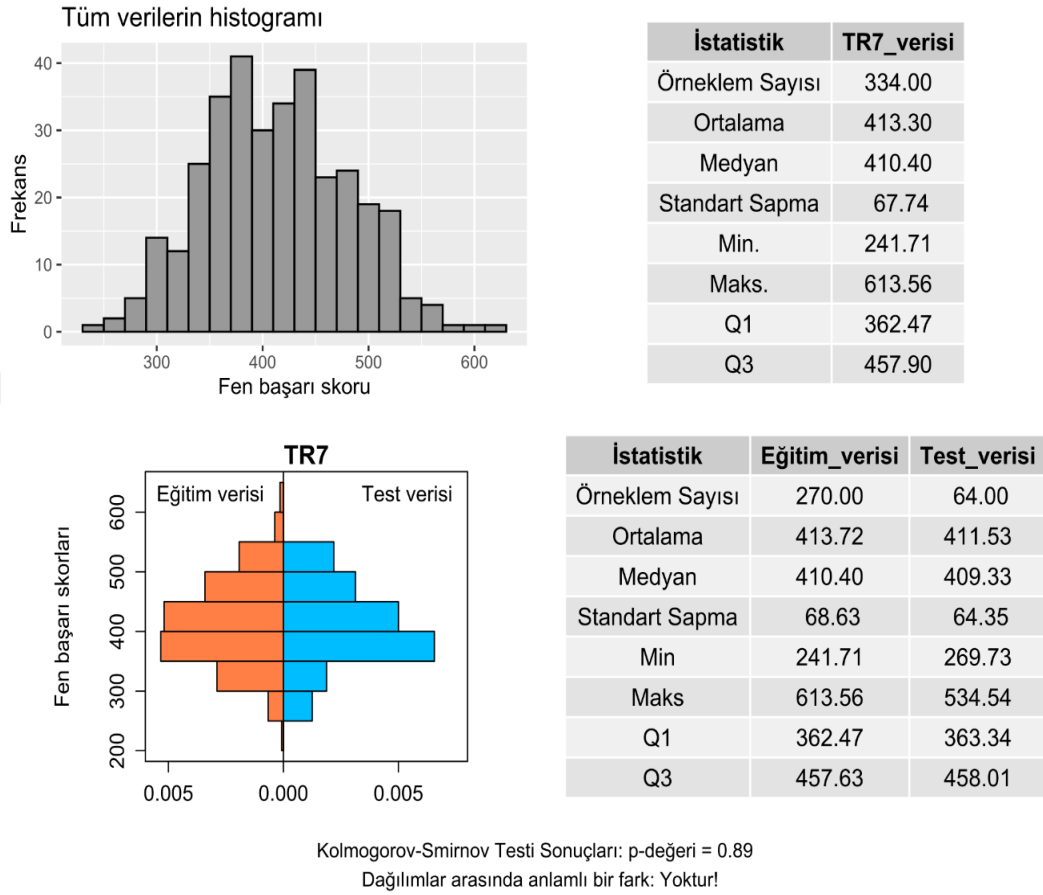
4.6.2 TR6 Bölgesinin Veri Yapısı

Şekil 29'da Türkiye'nin TR6 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özeti sunulmaktadır. Şekil 29'un üst panelinde, tüm veri setini temsil eden bir histogram ve istatistiksel özet bilgileri gösterilmektedir. Aynı şeklin alt panelinde ise, makine öğrenmesi modelinin oluşturulması ve modelin performansının değerlendirilmesi süreçlerinde kullanılacak olan eğitim ve test alt veri setlerine ait istatistiksel bilgiler bulunmaktadır. Bu alt veri setleri ana veri setinden rastgele bir şekilde ve literatüre uygun olarak %80 (Eğitim verisi) / %20 (Test verisi) oranında örneklenerek oluşturulmuştur (Gholamy ve diğerleri, 2018).

Şekil 29'un üst panelindeki histograma göre, fen puanlarının ortalama puan etrafında merkezi bir eğilimle dağıldığı görülmektedir. Bu TR6 bölgesindeki öğrencilerin (817 kişi) Fen başarı puanlarının normal dağılıma uyduğunu belirtmektedir. Ayrıca bu bölge için tüm veri setine ait ortalama puan 434.14, medyan 421.71, standart sapma 76.95 olarak hesaplanmıştır. Bunlara ek olarak minimum ve maksimum puanlar sırasıyla 268.96 ve 681.43 olup, birinci çeyrek (Q1) 376.03 ve üçüncü çeyrek (Q3) 486.63 olarak belirlenmiştir.

Şekil 29'un alt panelinde, eğitim veri setinin (270 kişi) ortalama puanı 413.72, medyanı 410.40 ve standart sapması 68.63 olarak tespit edilmiştir. Eğitim verisi için minimum ve maksimum puanlar sırasıyla 241.71 ve 613.56 olup, birinci çeyrek (Q1) 362.47 ve üçüncü çeyrek (Q3) 457.63 olarak belirlenmiştir. Bunların yanında Test veri setinin (64 kişi), ortalama puanı 411.53, medyanı 409.33 ve standart sapması 64.35 olarak tespit edilmiştir. Test verisi için minimum ve maksimum puanlar sırasıyla 269.73 ve 534.54 olup, birinci çeyrek (Q1) 363.34 ve üçüncü çeyrek (Q3) 458.01 olarak belirlenmiştir. Bu veri seti için rasgele bir şekilde oluşturulan eğitim ve test alt veri setlerine ait dağılımlarının

birbirlerinden anlamlı bir şekilde farklı olmadığı Kolmogorov-Smirnov testi sonucuna göre ($p > 0.05$) tespit edilmiştir.



Şekil 29: Türkiye'nin TR6 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)

4.6.3 TR6 Bölgesinde Makine Öğrenme Modeli

Tablo 20, TR6 (Akdeniz) verileri için oluşturulan makine öğrenme algoritmalarının eğitim verilerinin 10 katlı ve 10 tekrarlı çapraz doğrulama performans metriklerini ve optimize edilmiş hiperparametre değerlerini göstermektedir. Bu tabloda elde edilen bulgular; XGBoost, RF, SVR ve BRNN algoritmaları ile Kısım 3.3'te oluşturulan makine öğrenme modellerin 10 tekrarlı ve 10 katlı çapraz doğrulama tahmin performansları elde edilmiştir. Bu tablonun ayrıntılı sonuçları ve kısaltmalar için gerekli açıklamalar EK-1 içerisinde sunulmuştur. İncelenen ve oluşturulan makine öğrenmesi modellerin başarı performansları RMSE, R^2 ve MAE metrikleri üzerinden değerlendirilmiş ve XGBoost modelinin en düşük

RMSE (47.28), en yüksek R^2 (0.63) ve düşük MAE (37.35) değerleri ile eğitim verileri için en iyi performansı gösterdiği tespit edilmiştir. Buna karşın SVR, BRNN ve RF modelleri sırasıyla 47.62, 47.68 ve 50.36 RMSE değerleri ile XGBoost algoritmasının ardından sıralanmışlardır. Bu veriler, TR6 bölgesinin eğitim verileri için XGBoost modelinin Fen okuryazarlık puanlarının tahmininde diğer modellere göre daha iyi bir öğrenme ortaya koyduğunu göstermektedir.

Tablo 20: TR6 (Akdeniz) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri

Model	RMSE	R^2	MAE	Hiperparametre değerleri
XGBoost	47.28	0.63	37.35	nrounds 936.00000000 max_depth 8.00000000 eta 0.02179791 gamma 4.49466069 colsample_bytree 0.61731507 min_child_weight 20.00000000 subsample 0.55839360
SVR	47.62	0.62	37.85	sigma 0.003651681 C 3.069904528
BRNN	47.68	0.62	37.39	neurons 1
RF	50.36	0.59	40.32	mtry "71" splitrule "extratrees" min.node.size "5"

Tablo 21'de, TR6 bölgesine ait verileri için oluşturulan makine öğrenme algoritmalarının test verisi performans göstergeleri detaylı bir şekilde sunulmaktadır. Bu tabloya göre, XGBoost modelinin RMSE değeri 45.19, R^2 değeri 0.65 ve MAE değeri 36.08 olarak belirlenmiş ve diğer modellerden daha üstün bir performans sergilediği tespit edilmiştir. Ayrıca, XGBoost modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.797, güven aralığının alt sınırı 0.735 ve üst sınırı 0.846 ile yüksek bir uyumu gösterdiği belirlenmiştir. Buna karşın ikinci en iyi performans gösteren SVR modelinde, RMSE değeri 48.40, R^2 değeri 0.61 ve MAE değeri 38.71 ile XGBoost'a göre daha düşük olup, uyum korelasyon katsayısı merkezi tahmini 0.776 (lwr.ci: 0.707- upr.ci: 0.830) olarak hesaplanmıştır. BRNN ve RF modellerinde ise sırasıyla RMSE değerleri 49.16 ve 50.07, R^2 değerleri 0.60 ve 0.57, MAE değerleri 38.95 ve 39.01 olarak belirlenmiştir. BRNN modelinin uyum korelasyon katsayısı merkezi tahmini 0.770 (lwr.ci: 0.700- upr.ci: 0.825) iken, RF modelinin uyum korelasyon katsayısı merkezi tahmini 0.711 (lwr.ci: 0.635- upr.ci:

0.773) olarak hesaplanmıştır. Bu sonuçlar, XGBoost modelinin en düşük hata oranları ve en yüksek açıklama gücü ile diğer modellerden daha üstün performans sergilediğini ortaya koymaktadır.

Tablo 21: TR6 (Akdeniz) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri

Model	RMSE	R^2	MAE	ρ_c
XGBoost	45.19	0.65	36.08	XGBoost.est 0.7970342 XGBoost.lwr.ci 0.7347007 XGBoost.upr.ci 0.8460221
SVR	48.40	0.61	38.71	SVR.est 0.7756706 SVR.lwr.ci 0.7070806 SVR.upr.ci 0.8298008
BRNN	49.16	0.60	38.95	BRNN.est 0.7698223 BRNN.lwr.ci 0.6996731 BRNN.upr.ci 0.8252708
RF	50.07	0.57	39.01	RF.est 0.7107067 RF.lwr.ci 0.6352234 RF.upr.ci 0.7727410

ρ_c = Uyum Korelasyon Katsayısı (Concordance Correlation Coefficient)

est = ρ_c için merkezi tahmin edilen değeri

lwr.ci = güven aralığının alt sınırını (lower confidence interval)

upr.ci = güven aralığının üst sınırını (upper confidence interval)

Tablo 20 ve Tablo 21 verilerine göre yapılan genel değerlendirmede, XGBoost modelinin hem eğitim hem de test verilerinde en yüksek performansı sergilediği görülmektedir. Bu durum, genelleştirme ve tahmin kabiliyetinin güçlü olduğunu göstermektedir. Ayrıca test verilerinin uyum korelasyon katsayısının (ρ_c) yüksek olması, modelin tahminlerinin gerçek verilerle yüksek uyum gösterdiğini belirtmektedir. Bu, XGBoost modelinin hem eğitim hem de test verilerinde tutarlı ve güvenilir bir performans sergilediğinin kanıtıdır.

Diğer yandan SVR, BRNN ve RF modelleri ise eğitim ve test verileri arasında benzer hata oranlarına sahip olsalar da performansları XGBoost modellerine kıyasla daha düşük seviyelerde kalmıştır. Bu durum SVR, BRNN ve RF modellerinin genelleştirme kabiliyetinin sınırlı olduğunu ve yeni veriler üzerinde daha düşük performans sergilediğini göstermektedir. Her üç modelin de ρ_c değerleri, tahminlerinin gerçek verilerle makul bir uyum gösterdiğini belirtmektedir.

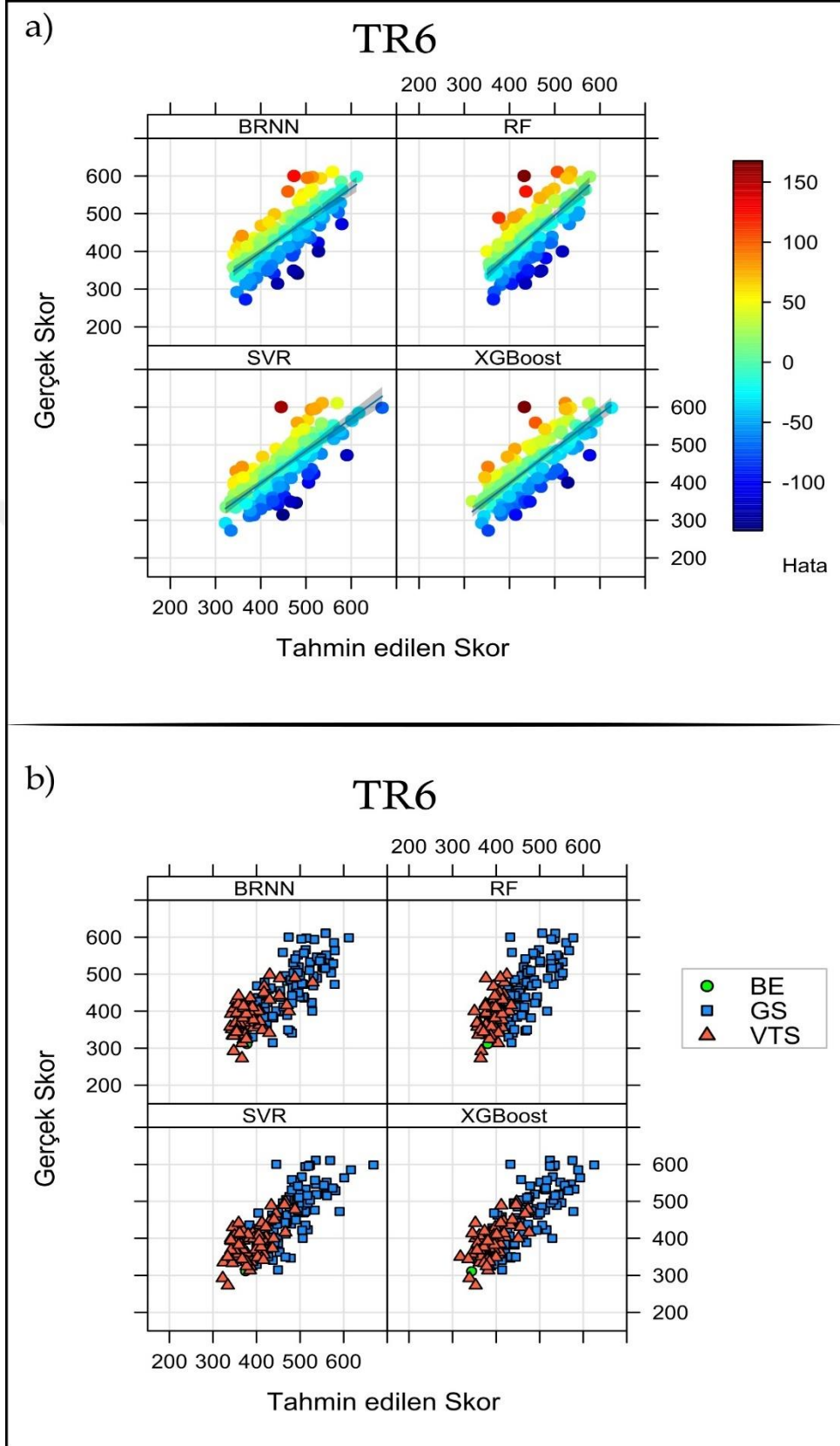
Şekil 30, Tablo 21’de elde edilen performans metriklerine ait çapraz doğrulama diyagramlarını tanımlamaktadır. Şekil 30a’da tüm modellerin hata dağılımlarının regresyon çizgisi boyunca ve yakınlarında dağıldığı görülmektedir. Bu durum tahmin hatalarının dağılımının normal yapıda olduğunu ve hata ortalamasının da sıfıra yakın olduğunu göstermektedir. Bu sonuçlar Tablo 21’de elde edilen bulgular ile uyumludur. Bunlarla birlikte XGBoost modelinin yüksek tahmin performansı sergilediği, Şekil 30a’daki çapraz doğrulama diyagramının hata dağılımlarının regresyon çizgisi etrafında sıkı şekilde toplanmasından dolayı daha açık bir şekilde görülmektedir.

Şekil 30b, aynı test verileri için modellerin farklı okul türlerine (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim) ait tahmin edilen ve gerçek skorlarının karşılaştırmasını göstermektedir. Burada özellikle tanımlayıcı regresyon çizgisinden sapan ve modellerin tahmin etme performansının düşmesine neden olan okul grubunun Mesleki ve Teknik Ortaöğretim (VTS) okul türü olduğu dikkat çekmektedir. Bununla birlikte XGBoost, her tüm okul türlerinde de tahmin edilen skorların gerçek skorlarla en uyumlu olduğu model olarak öne çıkmaktadır. Ayrıca TR6 bölgesindeki Fen başarı skorları için genel ortaokul (GS) başarı düzeyinin VTS okul türüne göre belirgin şekilde artış göstermiş olması dikkat çekicidir. Özellikle okul türünün modellerin oluşturulmasında bir değişken olarak kullanılmamış olmasına rağmen, makine öğrenme algoritmasının bu başarı farklılıklarını diğer değişkenlerin yardımı ile tespit edebilmesi ve tahminlerini buna göre gerçekleştirmesi önemli bir başarı olduğu düşünülmektedir. Bu bulgular, XGBoost modelinin eğitim verileri üzerindeki güçlü performansını ve genelleştirme kabiliyetini destekler niteliktedir. XGBoost’un, farklı okul türlerindeki öğrencilerin performansını doğru bir şekilde tahmin edebilmesi, eğitim politikaları ve müdahalelerinin planlanmasında doğru tahmin skorları önerebileceğini ve bu konuda değerli bir araç olarak kullanılabileceğini göstermektedir. Özellikle Mesleki ve Teknik Ortaöğretim (VTS) öğrencileri için tahminlerdeki sapmalara rağmen, XGBoost’un genel olarak diğer modellere göre daha yüksek doğrulukla sonuçlar vermesi, bu modelin TR6 bölgesi için geniş bir yelpazede güvenilir tahminler oluşturabileceğini göstermektedir.

Şekil 31, farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost) tahmin ettiği başarı skorları ile gerçek başarı skorlarının karşılaştırmalı kutu grafiğini sunmaktadır. Şekil 30’da sunulan çapraz doğrulama bulgularına göre oluşturulmuş bu grafikler, modellerin tahmin performansını ve gerçek verilerle uyumunu değerlendirmek için önemli

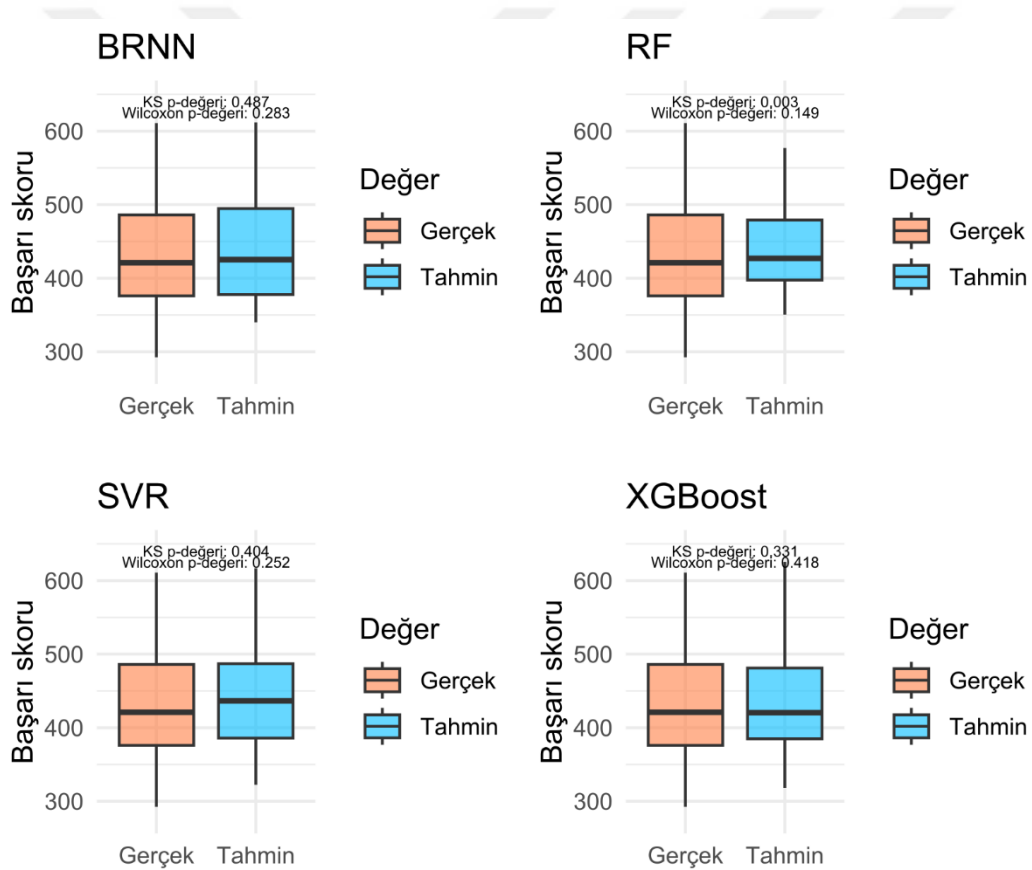
bir yaklaşımlı sağlamaktadır. Buna göre RF hariç tüm modellerin tahmin skorlarının hem Kolmogorov -Simirnov testi (KS) hem de Wilcoxon sıralama toplamı testine göre gerek skorların genel daėılımlarından ve merkezi eğilimlerinden anlamlı ölçüde bir farklılığı ($p > 0.05$) olmadığı tespit edilmiştir. Bu durum makine öğrenmesi modellerinin tahmin edilen ve gerek başarı skorları arasındaki uyumun yüksek olduğunu ortaya koymaktadır.





Şekil 30: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)

Şekil 31 en yüksek performans seviyesi gösteren XGBoost algoritması açısından değerlendirildiğinde, Kutu grafiklerinin gerçek ve tahmin skorları için temsil ettiği ortalama ve medyan değerlerinin neredeyse örtüştüğü ve dağılımların (kutu dışındaki uç örnekler için yeterli seviyelerde olmasa da) Q1-Q3 aralığı için oldukça yakın olduğunu görülmektedir. Şekil 30'deki bulgulara göre Tahmin edilen ve gerçek skorların genel dağılımlarının (KS p-değeri 0.331) ve merkezi eğilimlerinin (Wilcoxon p-değeri 0.418), istatistiksel olarak birbirlerinden farklı olmadığı belirlenmiştir. Bu durum, XGBoost algoritmasının TR6 verileri üzerinde en yüksek doğruluk ve güvenilirliği sunmakla birlikte hem bireysel gerçek skorları hem de gerçek skorlara ait dağılım yapısını doğru şekilde tahmin edebilme yeteneğini ortaya koymaktadır.



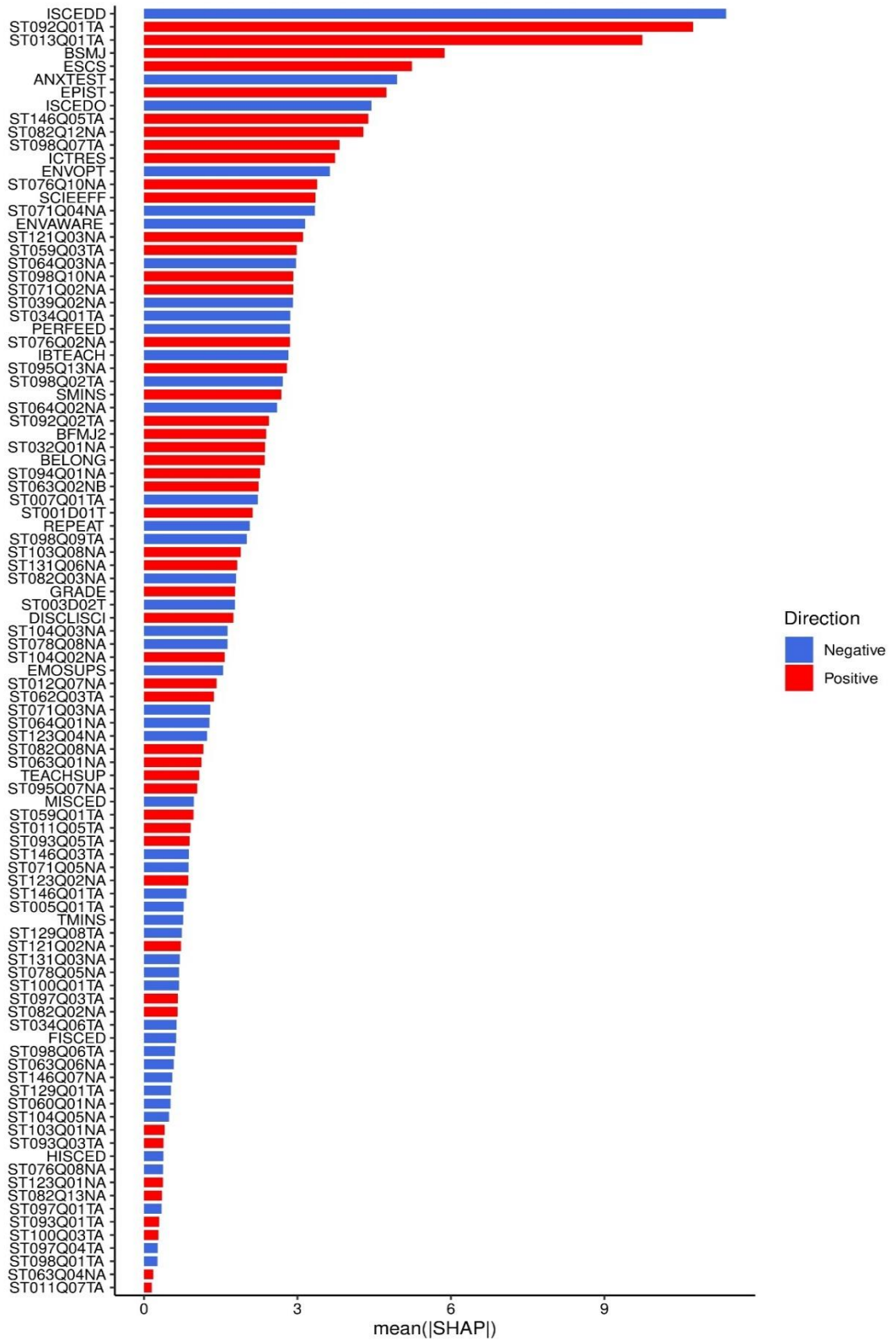
Şekil 31: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.

4.6.4 TR6 Bölgesi İçin Değişkenlerin Önem Seviyeleri

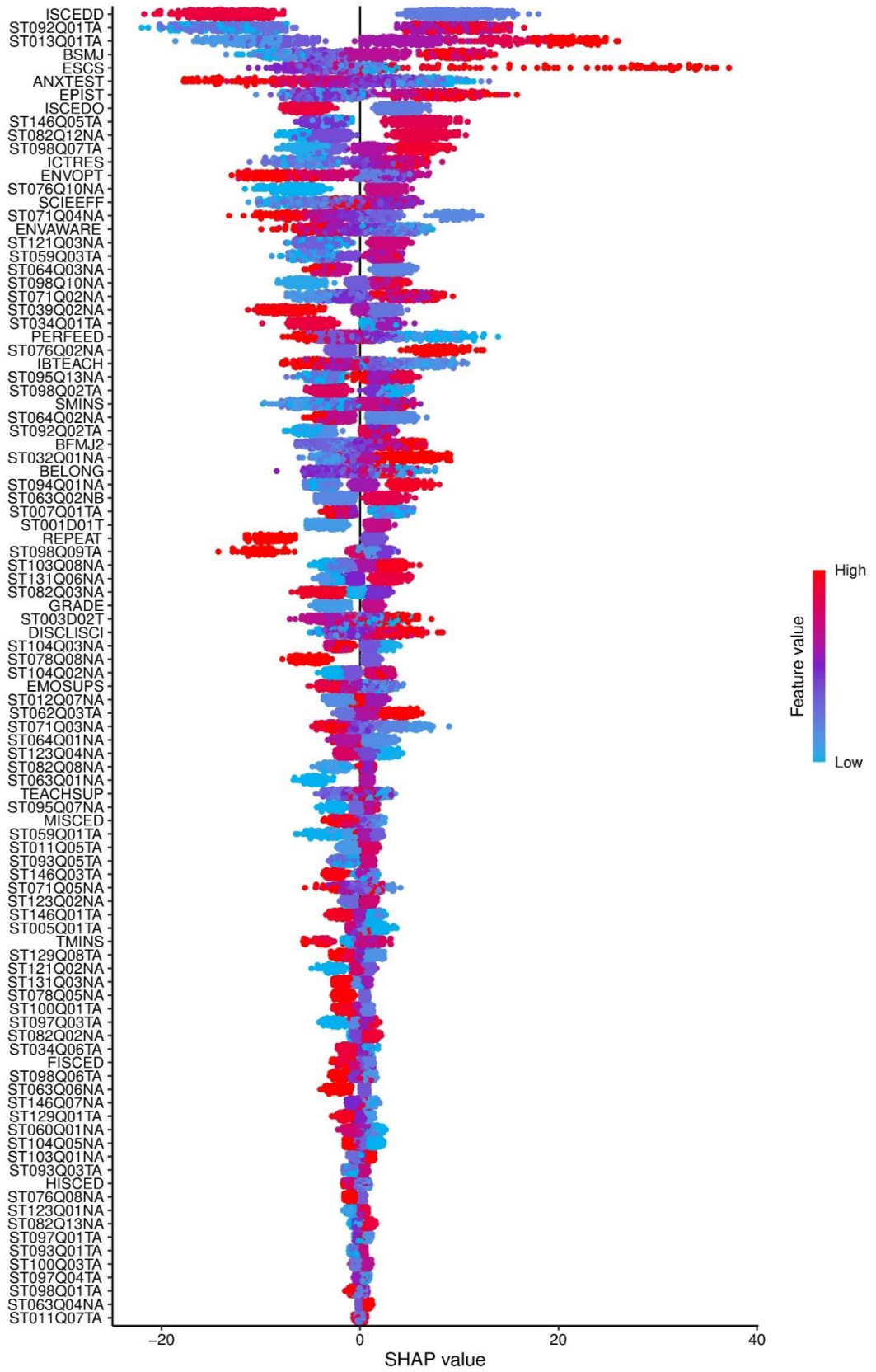
PISA 2015 Türkiye örnekleme TR6 bölgesi Fen okuryazarlık alanı için en iyi tahmin performansı gösteren XGBoost algoritmasına göre SHAP (SHapley Additive exPlanations) değerleri kullanılarak değişkenlerin önem parametreleri belirlenmiştir. Sonuçlar her bir değişkenin ortalama SHAP değerlerinin mutlak değerine göre azalan sırada Tablo 22’de, sıralanmıştır. Bu sıralamaya göre diğer bölgelerdeki değişkenlerin karşılaştırması ise Kısım 4.13’te Jaccard indeksine göre yapılacaktır. Burada model için önemli olan sıralı değişkenler üç önem grubuna ayrılmış ve bu işlem model için kullanılan toplam değişkenleri 3 eşit parçaya ayırarak gerçekleştirilmiştir. Burada yüksek derecede önemli değişkenler 1. Grup (ilk 33 tanesi), orta derecede önemli değişkenler 2. Grup (34- 66) ve düşük derecede önemli değişkenler 3. Grup (67 - 99) olarak etiketlenmiştir . Ayrıca Şekil 32 ile Tablo 22 de özetlenen değişkenlerin ortalama SHAP değerleri görselleştirilmiştir. Bu şekil her bir tahmin edici değişkenin model tahminlerine olan ortalama etkisini ve yönünü (pozitif ve negatif katkıları) vurgulamaktadır. Şekil 33 her bir değişkenin test veri setindeki her bir gözlemdeki etkisini göstermektedir. Bu şekil içerisinde her bir nokta, bir veri noktasının bir özelliği için SHAP değerini temsil eder. Bu şeklin sağ tarafında pozitif SHAP değerleri, sol tarafında ise negatif SHAP değerleri bulunmaktadır. Tahmin edici değişkenlerin model çıktısına olan etkisi, bu SHAP değerlerinin büyüklüğüne göre belirlenir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Tablo22, Şekil 32 ve 33’teki bulgulara göre, TR6 bölgesi için model tahmininde önem seviyesi yüksek ilk 10 değişken, parantez içleri SHAP değerlerini ve model katkı yönlerini göstermek üzere, şu şekilde sıralanmıştır: ISCEDD (11.374, negatif), ST092Q01TA (10.732, pozitif), ST013Q01TA (9.740, pozitif), BSMJ (5.876, pozitif), ESCS (5.236, pozitif), ANXTEST (4.949, negatif), EPIST (4.738, pozitif), ISCEDO (4.445, negatif), ST146Q05TA (4.384, pozitif), ST082Q12NA (4.286, pozitif). Şekil 9 de Kırmızı ve mavi renkler, sırasıyla pozitif ve negatif yönlü etkileri temsil etmektedir. Bu, modelin tahminlerinin hangi özellikler tarafından ne kadar ve hangi yönde etkilendiğini daha net bir şekilde anlaşılmasına yardımcı olmaktadır. Buna göre negatif özellikteki değişkenlerin (ISCEDD, ANXTEST, ISCEDO) değerlerindeki artış öğrencilerin Fen başarı skorlarını düşürmekte buna karşı pozitif özellikteki değişkenlerin değerlerindeki artış ise öğrencilerin Fen başarı skorlarını artırıcı yönde etki etmektedir.

Tablo 22: TR6 (Akdeniz) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.

ID	Değişken	SHAP değeri	Önem Grubu	ID	Değişken	SHAP değeri	Önem Grubu
1	ISCEDD	11.374	Grup 1	55	ST064Q01NA	1.280	Grup 2
2	ST092Q01TA	10.732	Grup 1	56	ST123Q04NA	1.235	Grup 2
3	ST013Q01TA	9.740	Grup 1	57	ST082Q08NA	1.159	Grup 2
4	BSMJ	5.876	Grup 1	58	ST063Q01NA	1.123	Grup 2
5	ESCS	5.236	Grup 1	59	TEACHSUP	1.080	Grup 2
6	ANXTEST	4.949	Grup 1	60	ST095Q07NA	1.043	Grup 2
7	EPIST	4.738	Grup 1	61	MISCED	0.976	Grup 2
8	ISCEDO	4.445	Grup 1	62	ST059Q01TA	0.966	Grup 2
9	ST146Q05TA	4.384	Grup 1	63	ST011Q05TA	0.912	Grup 2
10	ST082Q12NA	4.286	Grup 1	64	ST093Q05TA	0.893	Grup 2
11	ST098Q07TA	3.820	Grup 1	65	ST146Q03TA	0.879	Grup 2
12	ICTRES	3.736	Grup 1	66	ST071Q05NA	0.872	Grup 2
13	ENVOPT	3.632	Grup 1	67	ST123Q02NA	0.867	Grup 3
14	ST076Q10NA	3.384	Grup 1	68	ST146Q01TA	0.833	Grup 3
15	SCIEEFF	3.349	Grup 1	69	ST005Q01TA	0.775	Grup 3
16	ST071Q04NA	3.342	Grup 1	70	TMINS	0.768	Grup 3
17	ENVAWARE	3.149	Grup 1	71	ST129Q08TA	0.739	Grup 3
18	ST121Q03NA	3.111	Grup 1	72	ST121Q02NA	0.723	Grup 3
19	ST059Q03TA	2.985	Grup 1	73	ST131Q03NA	0.699	Grup 3
20	ST064Q03NA	2.971	Grup 1	74	ST078Q05NA	0.687	Grup 3
21	ST098Q10NA	2.919	Grup 1	75	ST100Q01TA	0.686	Grup 3
22	ST071Q02NA	2.919	Grup 1	76	ST097Q03TA	0.665	Grup 3
23	ST039Q02NA	2.911	Grup 1	77	ST082Q02NA	0.656	Grup 3
24	ST034Q01TA	2.860	Grup 1	78	ST034Q06TA	0.637	Grup 3
25	PERFEED	2.855	Grup 1	79	FISCED	0.630	Grup 3
26	ST076Q02NA	2.855	Grup 1	80	ST098Q06TA	0.607	Grup 3
27	IBTEACH	2.823	Grup 1	81	ST063Q06NA	0.584	Grup 3
28	ST095Q13NA	2.793	Grup 1	82	ST146Q07NA	0.556	Grup 3
29	ST098Q02TA	2.714	Grup 1	83	ST129Q01TA	0.525	Grup 3
30	SMINS	2.684	Grup 1	84	ST060Q01NA	0.519	Grup 3
31	ST064Q02NA	2.602	Grup 1	85	ST104Q05NA	0.492	Grup 3
32	ST092Q02TA	2.441	Grup 1	86	ST103Q01NA	0.405	Grup 3
33	BFMJ2	2.389	Grup 1	87	ST093Q03TA	0.382	Grup 3
34	ST032Q01NA	2.365	Grup 2	88	HISCED	0.380	Grup 3
35	BELONG	2.359	Grup 2	89	ST076Q08NA	0.374	Grup 3
36	ST094Q01NA	2.271	Grup 2	90	ST123Q01NA	0.372	Grup 3
37	ST063Q02NB	2.238	Grup 2	91	ST082Q13NA	0.353	Grup 3
38	ST007Q01TA	2.226	Grup 2	92	ST097Q01TA	0.346	Grup 3
39	ST001D01T	2.124	Grup 2	93	ST093Q01TA	0.295	Grup 3
40	REPEAT	2.067	Grup 2	94	ST100Q03TA	0.286	Grup 3
41	ST098Q09TA	2.009	Grup 2	95	ST097Q04TA	0.270	Grup 3
42	ST103Q08NA	1.892	Grup 2	96	ST098Q01TA	0.265	Grup 3
43	ST131Q06NA	1.823	Grup 2	97	ST063Q04NA	0.182	Grup 3
44	ST082Q03NA	1.802	Grup 2	98	ST011Q07TA	0.153	Grup 3
45	GRADE	1.781	Grup 2	99	ST061Q01NA	0	Grup 3
46	ST003D02T	1.778	Grup 2				
47	DISCLISCI	1.747	Grup 2				
48	ST104Q03NA	1.637	Grup 2				
49	ST078Q08NA	1.633	Grup 2				
50	ST104Q02NA	1.576	Grup 2				
51	EMOSUPS	1.549	Grup 2				
52	ST012Q07NA	1.420	Grup 2				
53	ST062Q03TA	1.367	Grup 2				
54	ST071Q03NA	1.298	Grup 2				



Şekil 32: TR6 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.



Şekil 33: TR6 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.

4.7 TR7 Bölgesi PISA-2015 Bulguları

4.7.1 TR7 Bölgesi Hakkında Temel Bilgiler

2015 PISA uygulamasının Türkiye örnekleminin yedinci bölgesi TR7 olarak adlandırılmıştır. TR7 bölgesi 2015 PISA uygulamasında ayrılan bölgelerden Orta Anadolu'yu temsil etmektedir. Orta Anadolu olarak ayrılan bölge Kırıkkale, Kırşehir, Aksaray, Niğde, Nevşehir, Kayseri, Yozgat, Sivas illerinin kapsamaktadır. TR7 bölgesinden (Orta Anadolu) 7. Sınıf ve üzeri düzeyde eğitim gören 334 öğrencinin 2015 PISA sınavına katıldığı görülmektedir. 2015 PISA sınavına katılan tüm Türkiye örnekleminin %5,67'sinin Orta Anadolu bölgesinden katıldığını görmekteyiz.

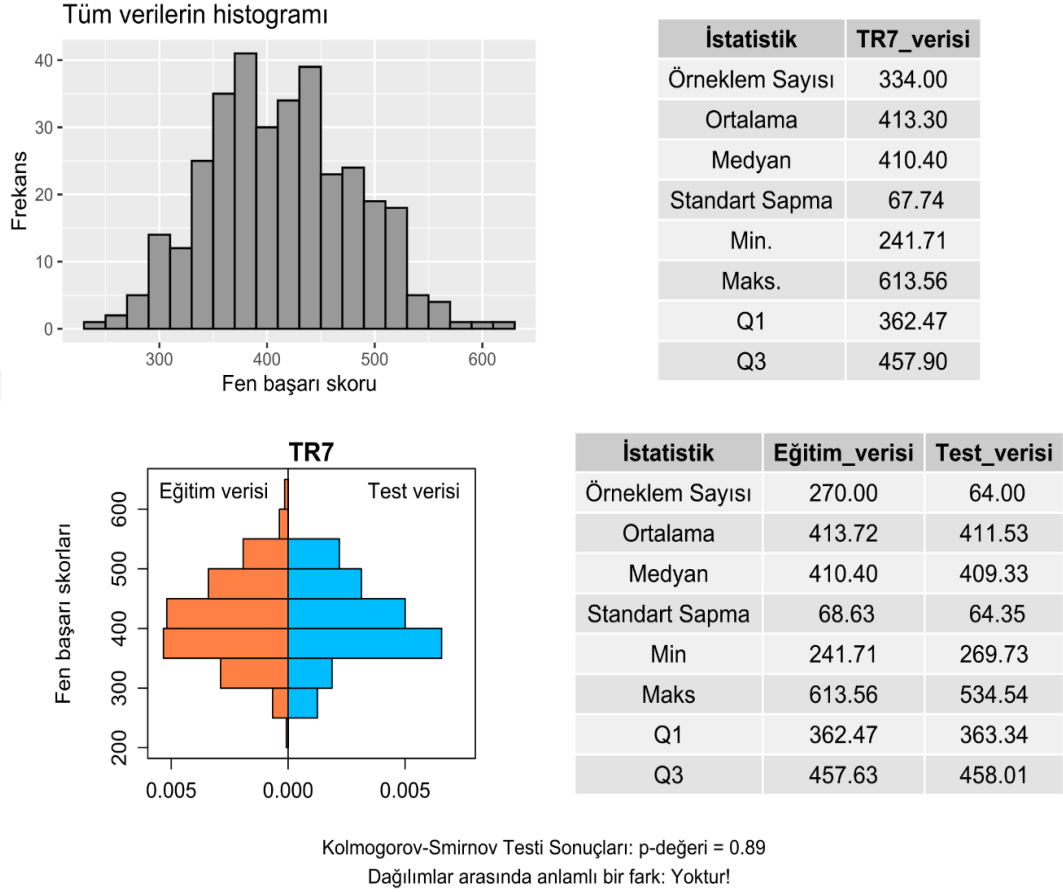
4.7.2 TR7 Bölgesinin Veri Yapısı

Şekil 38'de, Türkiye'nin TR7 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özeti sunulmaktadır. Şekil 1'in üst panelinde, tüm veri setini temsil eden bir histogram ve istatistiksel özet bilgileri gösterilmektedir. Aynı şeklin alt panelinde ise, makine öğrenmesi modelinin oluşturulması ve modelin performansının değerlendirilmesi süreçlerinde kullanılacak olan eğitim ve test alt veri setlerine ait istatistiksel bilgiler bulunmaktadır. Bu alt veri setleri ana veri setinden rastgele bir şekilde ve literatüre uygun olarak %80 (Eğitim verisi) / %20 (Test verisi) oranında örneklenerek oluşturulmuştur (Gholamy ve diğerleri, 2018).

Şekil 34'ün üst panelindeki histograma göre, fen puanlarının ortalama puan etrafında merkezi bir eğilimle dağıldığı görülmektedir. Bu TR7 bölgesindeki öğrencilerin (334 kişi) Fen başarı puanlarının normal dağılıma uyduğunu belirtmektedir. Ayrıca bu bölge için tüm veri setine ait ortalama puan 413.30, medyan 410.40, standart sapma 67.74 olarak hesaplanmıştır. Bunlara ek olarak minimum ve maksimum puanlar sırasıyla 241.71 ve 613.56 olup, birinci çeyrek (Q1) 362.47 ve üçüncü çeyrek (Q3) 457.90 olarak belirlenmiştir.

Şekil 34'ün alt panelinde, Eğitim veri setinin (158 kişi) ortalama puanı 408.27, medyanı 403.97 ve standart sapması 54.66 olarak tespit edilmiştir. Eğitim verisi için minimum ve maksimum puanlar sırasıyla 250.05 ve 565.70 olup, birinci çeyrek (Q1) 376.95 ve üçüncü çeyrek (Q3) 441.08 olarak belirlenmiştir. Bunların yanında Test veri setinin (36 kişi), ortalama puanı 408.09, medyanı 406.31 ve standart sapması 50.79 olarak tespit edilmiştir. Test verisi için minimum ve maksimum puanlar sırasıyla 320.52 ve 507.81 olup, birinci çeyrek (Q1) 376.04 ve üçüncü çeyrek (Q3) 440.18 olarak belirlenmiştir. Bu veri seti

için rasgele bir şekilde oluşturulan eğitim ve test alt veri setlerine ait dağılımlarının birbirlerinden anlamlı bir şekilde farklı olmadığı Kolmogorov-Smirnov testi sonucuna göre ($p > 0.05$) tespit edilmiştir.



Şekil 34: Türkiye'nin TR7 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)

4.7.3 TR7 Bölgesinde Makine Öğrenme Modeli

Tablo 23 TR7 (Orta Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı ve 10 tekrarlı çapraz doğrulama performans metriklerini ve optimize edilmiş Hiperparametre değerlerini göstermektedir. Bu tabloda elde edilen bulgular, XGBoost, RF, SVR ve BRNN algoritmaları ile Kısım 3.3 oluşturulan makine öğrenme modellerin 10 tekrarlı ve 10 katlı çapraz doğrulama tahmin performansları elde edilmiştir. Bu tablonun ayrıntılı sonuçları ve kısaltmalar için gerekli açıklamalar EK-1 içerisinde sunulmuştur. İncelenen ve oluşturulan makine öğrenmesi modellerin başarı

performansları RMSE, R^2 ve MAE metrikleri üzerinden değerlendirilmiş ve SVR modelinin en düşük RMSE (44.47), en yüksek R^2 (0.59) ve düşük MAE (36.24) değerleri ile eğitim verileri için en iyi performansı gösterdiği tespit edilmiştir. Buna karşın XGBoost, RF ve BRNN modelleri sırasıyla 44.80, 47.66 ve 48.70 RMSE değerleri ile SVR algoritmasının ardından sıralanmışlardır. Bu veriler, TR7 bölgesinin eğitim verileri için SVR modelinin Fen okuryazarlık skorlarının tahmininde diğer modellere göre daha iyi bir öğrenme ortaya koyduğunu göstermektedir.

Tablo 23: TR7 (Orta Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri

Model	RMSE	R^2	MAE	Hiperparametre değerleri
SVR	44.47	0.59	36.24	sigma 0.006365391 C 0.972266366
XGBoost	44.80	0.57	35.64	nrounds 929.00000000 max_depth 3.00000000 eta 0.05587832 gamma 2.15745360 colsample_bytree 0.41682794 min_child_weight 17.00000000 subsample 0.81535185
RF	47.66	0.56	38.47	mtry "77" splitrule "maxstat" min.node.size "11"
BRNN	48.70	0.51	39.86	neurons 1

Tablo 24'te TR7 bölgesine ait verileri için oluşturulan makine öğrenme algoritmalarının test verisi performans göstergeleri detaylı bir şekilde sunulmaktadır. Bu tabloya göre, SVR modelinin RMSE değeri 42.28, R^2 değeri 0.57 ve MAE değeri 32.80 olarak belirlenmiş ve diğer modellerden daha üstün bir performans sergilediği tespit edilmiştir. Ayrıca, SVR modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.712, güven aralığının alt sınırı 0.585 ve üst sınırı 0.805 ile yüksek bir uyumu gösterdiği belirlenmiştir. Buna karşın ikinci en iyi performans gösteren XGBoost modelinde, RMSE değeri 43.24, R^2 değeri 0.56 ve MAE değeri 34.12 ile SVR modeline göre daha düşük olup, uyum korelasyon katsayısı merkezi tahmini 0.736 (lwr.ci: 0.604- upr.ci: 0.829) olarak hesaplanmıştır. BRNN ve RF modellerinde ise sırasıyla RMSE değerleri 46.60 ve 47.75, R^2 değerleri 0.50 ve 0.44, MAE değerleri 39.36 ve 36.61 olarak belirlenmiştir.

BRNN modelinin uyum korelasyon katsayısı merkezi tahmini 0.701 (lwr.ci: 0.554- upr.ci: 0.805) iken, RF modelinin uyum korelasyon katsayısı merkezi tahmini 0.593 (lwr.ci: 0.444- upr.ci: 0.710) olarak hesaplanmıştır. Bu sonuçlar, TR7 bölgesi için SVR modelinin en düşük hata oranları ve en yüksek açıklama gücü ile diğer modellerden daha üstün performans sergilediğini ortaya koymaktadır.

Tablo 24: TR7 (Orta Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri

Model	RMSE	R^2	MAE	ρ_c
SVR	42.28	0.57	32.80	SVR.est 0.7124192 SVR.lwr.ci 0.5852522 SVR.upr.ci 0.8053820
XGBoost	43.24	0.56	34.12	XGBoost.est 0.7361119 XGBoost.lwr.ci 0.6044504 XGBoost.upr.ci 0.8286298
BRNN	46.60	0.50	39.36	BRNN.est 0.7012453 BRNN.lwr.ci 0.5542531 BRNN.upr.ci 0.8058204
RF	47.75	0.44	36.61	RF.est 0.5927578 RF.lwr.ci 0.4436694 RF.upr.ci 0.7099207

Tablo 23 ve Tablo 24 verilerine göre yapılan genel değerlendirmede, SVR modelinin hem eğitim hem de test verilerinde en yüksek performansı sergilediği görülmektedir. Bu durum, modelin genelleştirme kabiliyetinin güçlü olduğunu göstermektedir. Ayrıca test verilerinin uyum korelasyon katsayısının (ρ_c) yüksek olması, modelin tahminlerinin gerçek verilerle yüksek uyum gösterdiğini belirtmektedir. Bu, SVR modelinin TR7 bölgesinin hem eğitim hem de test verilerinde tutarlı ve güvenilir bir performans sergilediğinin kanıtıdır.

Diğer yandan XGBoost, BRNN ve RF modelleri ise eğitim ve test verileri arasında benzer hata oranlarına sahip olsalar da performansları SVR modellerine kıyasla daha düşük seviyelerde kalmıştır. Bu durum XGBoost, BRNN ve RF modellerinin TR7 bölgesi için genelleştirme kabiliyetinin sınırlı olduğunu ve yeni veriler üzerinde daha düşük performans sergilediğini göstermektedir. Her üç modelin de ρ_c değerleri, tahminlerinin gerçek verilerle makul bir uyum gösterdiğini belirtmektedir.

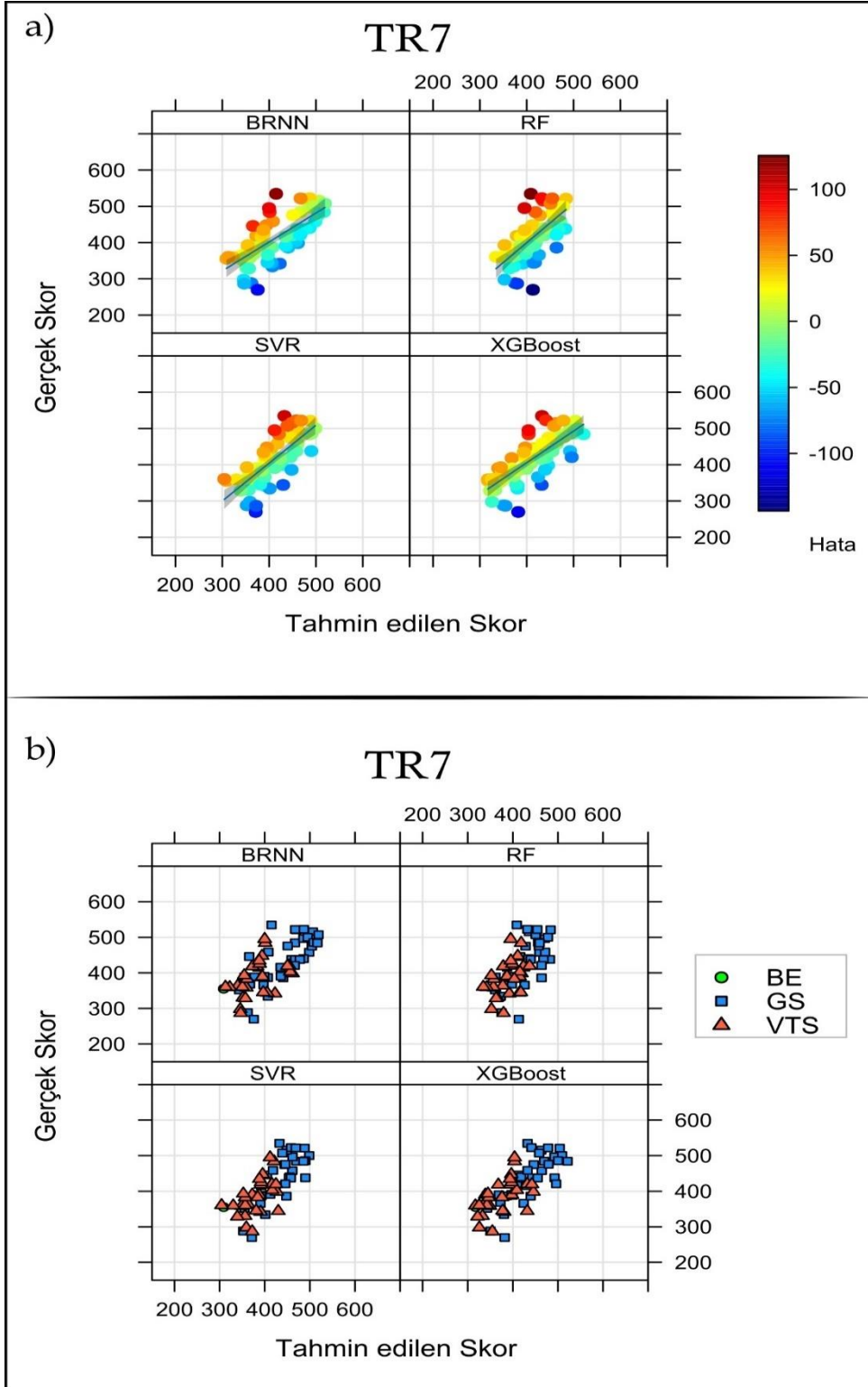
Şekil 35, Tablo 24'te elde edilen performans metriklerine ait çapraz doğrulama diyagramlarını tanımlamaktadır. Şekil 35a'da tüm modellerin hata dağılımlarının regresyon

çizgisi boyunca ve yakınlarında dağıldığı görülmektedir. Bu durum tahmin hatalarının dağılımının normal yapıda olduğunu ve hata ortalamasının da sıfıra yakın olduğunu göstermektedir. Bu sonuçlar Tablo 24 de elde edilen bulgular ile uyumludur. Bunlarla birlikte SVR modelinin yüksek tahmin performansı sergilediği, Şekil 35a'daki çapraz doğrulama diyagramının hata dağılımlarının regresyon çizgisi etrafında sıkı şekilde toplanmasından dolayı daha açık bir şekilde görülmektedir.

Şekil 35b, aynı test verileri için modellerin farklı okul türlerine (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim) ait tahmin edilen ve gerçek skorlarının karşılaştırmasını göstermektedir. Burada özellikle tanımlayıcı regresyon çizgisinden sapan ve modellerin tahmin etme performansının düşmesine neden olan okul grubunun Mesleki ve Teknik Ortaöğretim (VTS) okul türü olduğu dikkat çekmektedir. Bununla birlikte XGBoost, her tüm okul türlerinde de tahmin edilen skorların gerçek skorlarla en uyumlu olduğu model olarak öne çıkmaktadır. Ayrıca TR7 bölgesindeki Fen başarı skorları için genel ortaokul (GS) başarı düzeyinin VTS okul türüne göre belirgin şekilde artış göstermiş olması dikkat çekicidir. Özellikle okul türünün modellerin oluşturulmasında bir değişken olarak kullanılmamış olmasına rağmen, makine öğrenme algoritmasının bu başarı farklılıklarını diğer değişkenlerin yardımı ile tespit edebilmesi ve tahminlerini buna göre gerçekleştirmesi önemli bir başarı olduğu düşünülmektedir. Bu bulgular, SVR modelinin eğitim verileri üzerindeki güçlü performansını ve genelleştirme kabiliyetini destekler niteliktedir. SVR'nin, farklı okul türlerindeki öğrencilerin performansını doğru bir şekilde tahmin edebilmesi, eğitim politikaları ve müdahalelerinin planlanmasında doğru tahmin skorları önerebileceğini ve bu konuda değerli bir araç olarak kullanılabilirliğini göstermektedir. Özellikle Mesleki ve Teknik Ortaöğretim (VTS) öğrencileri için tahminlerdeki sapmalara rağmen, SVR modelinin genel olarak diğer modellere göre daha yüksek doğrulukla sonuçlar vermesi, bu modelin TR7 bölgesi için geniş bir yelpazede güvenilir tahminler oluşturabileceğini göstermektedir.

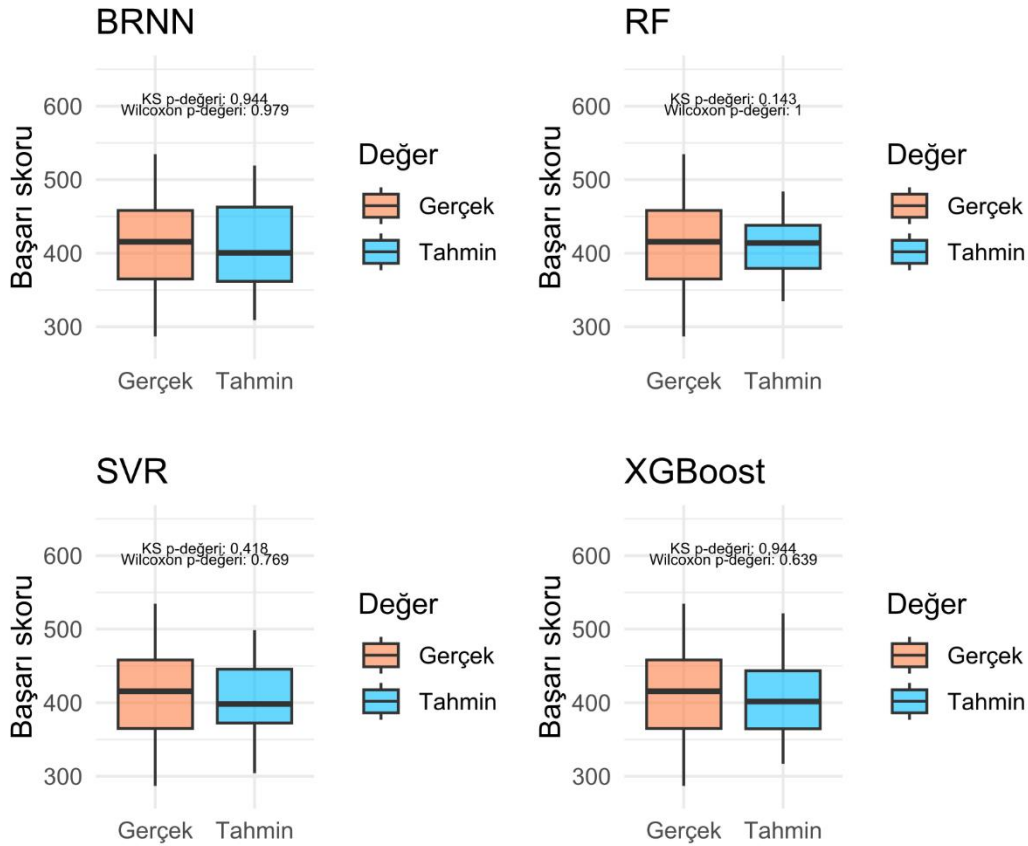
Şekil 36, farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost) tahmin ettiği başarı skorları ile gerçek başarı skorlarının karşılaştırmalı kutu grafiğini sunmaktadır. Şekil 35 de sunulan çapraz doğrulama bulgularına göre oluşturulmuş bu grafikler, modellerin tahmin performansını ve gerçek verilerle uyumunu değerlendirmek için önemli bir yaklaşım sağlamaktadır. Buna göre RF hariç tüm modellerin tahmin skorlarının hem Kolmogorov -Simirnov testi (KS) hem de Wilcoxon sıralama toplamı testine göre gerçek

skorların genel dağılımlarından ve merkezi eğilimlerinden anlamlı ölçüde bir farklılığı ($p > 0.05$) olmadığı tespit edilmiştir. Bu durum makine öğrenmesi modellerinin tahmin edilen ve gerçek başarı skorları arasındaki uyumun yüksek olduğunu ortaya koymaktadır.



Şekil 35: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)

Şekil 36, en yüksek performans seviyesi gösteren SVR algoritması açısından değerlendirildiğinde, Kutu grafiklerinin gerçek ve tahmin skorları için temsil ettiği ortalama ve medyan değerlerinin neredeyse örtüştüğü ve dağılımların (kutu dışındaki uç örnekler için yeterli seviyelerde olmasa da) Q1-Q3 aralığı için oldukça yakın olduğunu görülmektedir. Şekil 9'daki bulgulara göre Tahmin edilen ve gerçek skorların genel dağılımlarının (KS p-değeri 0.418) ve merkezi eğilimlerinin (Wilcoxon p-değeri 0.769), istatistiksel olarak birbirlerinden farklı olmadığı belirlenmiştir. Bu durum, SVR algoritmasının TR7 verileri üzerinde en yüksek doğruluk ve güvenilirliği sunmakla birlikte hem bireysel gerçek skorları hem de gerçek skorlara ait dağılım yapısını doğru şekilde tahmin edebilme yeteneğini ortaya koymaktadır.



Şekil 36: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.

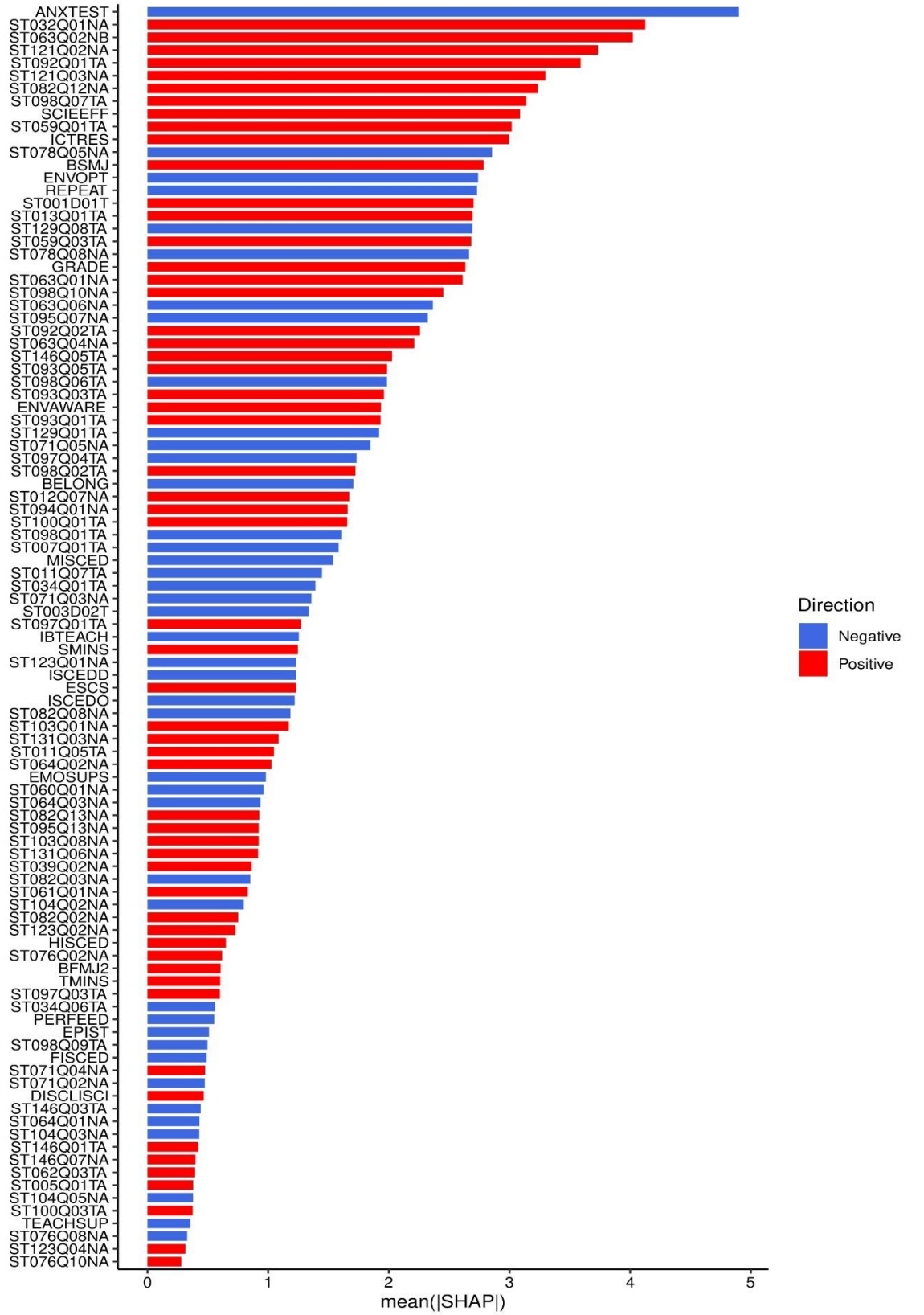
4.7.4 TR7 Bölgesi İçin Değişkenlerin Önem Seviyeleri

PISA 2015 Türkiye örnekleme TR7 bölgesi Fen okuryazarlık alanı için en iyi tahmin performansı gösteren SVR algoritmasına göre SHAP (SHapley Additive exPlanations) değerleri kullanılarak değişkenlerin önem parametreleri belirlenmiştir. Sonuçlar her bir değişkenin ortalama SHAP değerlerinin mutlak değerine göre azalan sırada Tablo 25'te, sıralanmıştır. Bu sıralamaya göre diğer bölgelerdeki değişkenlerin karşılaştırması ise Kısım 4.13'te Jaccard indeksine göre yapılacaktır. Burada model için önemli olan sıralı değişkenler üç önem grubuna ayrılmış ve bu işlem model için kullanılan toplam değişkenleri 3 eşit parçaya ayırarak gerçekleştirilmiştir. Burada yüksek derecede önemli değişkenler 1. Grup (ilk 33 tanesi), orta derecede önemli değişkenler 2. Grup (34 - 66) ve düşük derecede önemli değişkenler 3. Grup (67 - 99) olarak etiketlenmiştir. Ayrıca Şekil 37 ile Tablo 25 de özetlenen değişkenlerin ortalama SHAP değerleri görselleştirilmiştir. Bu şekil her bir tahmin edici değişkenin model tahminlerine olan ortalama etkisini ve yönünü (pozitif ve negatif katkıları) vurgulamaktadır. Şekil 38 her bir değişkenin test veri setindeki her bir gözlemdeki etkisini göstermektedir. Bu şekil içerisinde her bir nokta, bir veri noktasının bir özelliği için SHAP değerini temsil eder. Bu şeklin sağ tarafında pozitif SHAP değerleri, sol tarafında ise negatif SHAP değerleri bulunmaktadır. Tahmin edici değişkenlerin model çıktısına olan etkisi, bu SHAP değerlerinin büyüklüğüne göre belirlenir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Tablo 25, Şekil 37 ve 38'deki bulgulara göre, TR7 bölgesi için model tahmininde önem seviyesi yüksek ilk 10 değişken, parantez içleri SHAP değerlerini ve model katkı yönlerini göstermek üzere, şu şekilde sıralanmıştır: ANXTEST (4.901,negatif), ST032Q01NA (4.125,pozitif), ST063Q02NB (4.022,pozitif), ST121Q02NA (3.733, pozitif), ST092Q01TA (3.589, pozitif), ST121Q03NA (3.298, pozitif), ST082Q12NA (3.234, pozitif), ST098Q07TA (3.140 pozitif), SCIEEFF (3.088, pozitif), ST059Q01TA (3.019, pozitif). Şekil 9 da Kırmızı ve mavi renkler, sırasıyla pozitif ve negatif yönlü etkileri temsil etmektedir. Bu, modelin tahminlerinin hangi özellikler tarafından ne kadar ve hangi yönde etkilendiğini daha net bir şekilde anlaşılmasına yardımcı olmaktadır. Buna göre negatif özellikteki değişkenin (ANXTEST) değerlerindeki artış öğrencilerin Fen başarı skorlarını

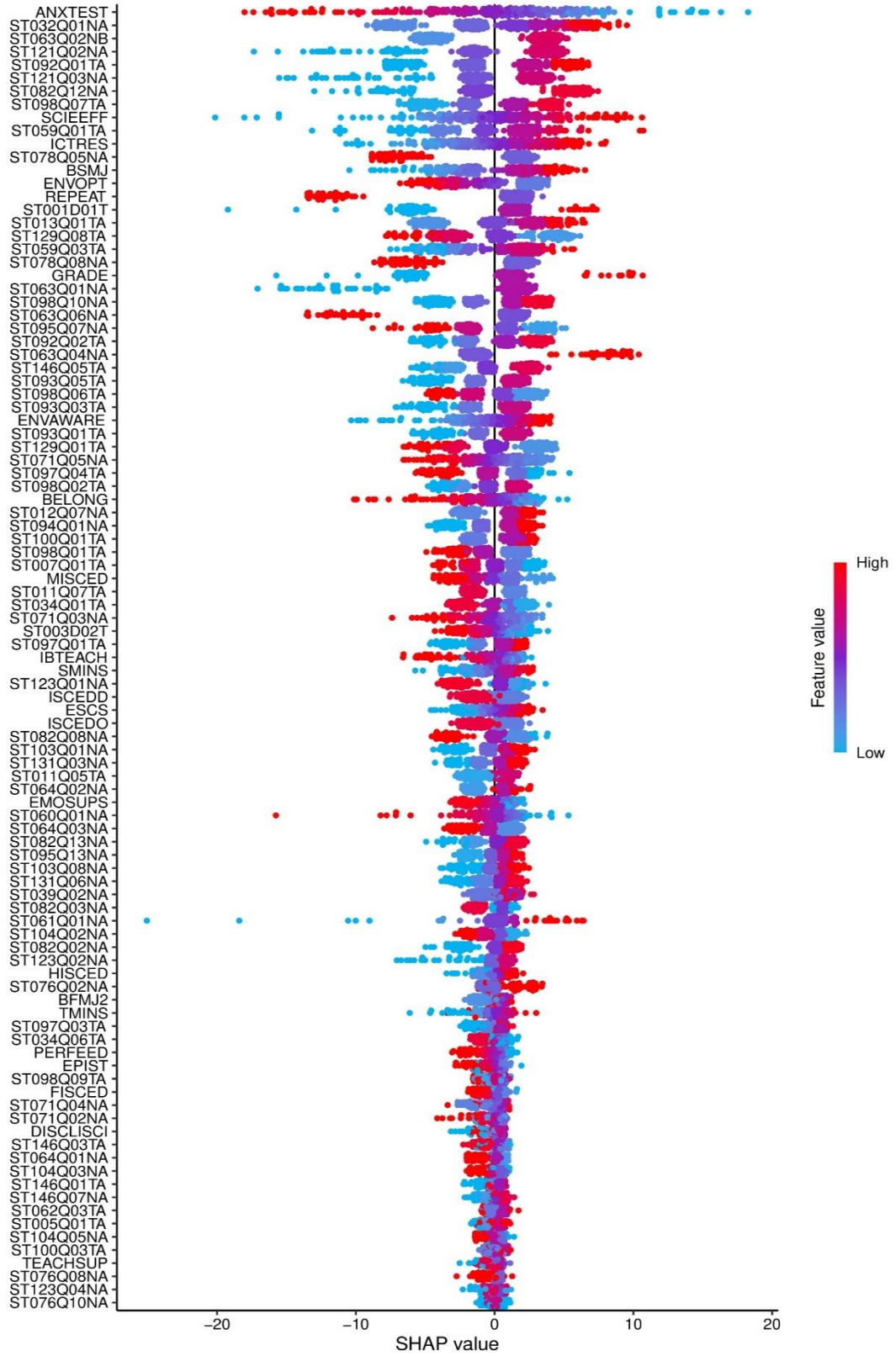
düşürmekte buna karşı pozitif özellikteki değişkenlerin değerlerindeki artış ise öğrencilerin Fen başarı skorlarını arttırıcı yönde etki etmektedir.

Tablo 25: TR7 (Orta Anadolu) verileri için oluşturulan En yüksek performanslı algoritmaya (SVR) göre değişkenlerin önem seviyelerinin sıralanması.

ID	Değişken	SHAP değeri	Önem Grubu	ID	Değişken	SHAP değeri	Önem Grubu
1	ANXTEST	4.901	Grup 1	55	ISCEDO	1.220	Grup 2
2	ST032Q01NA	4.125	Grup 1	56	ST082Q08NA	1.185	Grup 2
3	ST063Q02NB	4.022	Grup 1	57	ST103Q01NA	1.171	Grup 2
4	ST121Q02NA	3.733	Grup 1	58	ST131Q03NA	1.087	Grup 2
5	ST092Q01TA	3.589	Grup 1	59	ST011Q05TA	1.049	Grup 2
6	ST121Q03NA	3.298	Grup 1	60	ST064Q02NA	1.028	Grup 2
7	ST082Q12NA	3.234	Grup 1	61	EMOSUPS	0.981	Grup 2
8	ST098Q07TA	3.140	Grup 1	62	ST060Q01NA	0.962	Grup 2
9	SCIEEFF	3.088	Grup 1	63	ST064Q03NA	0.936	Grup 2
10	ST059Q01TA	3.019	Grup 1	64	ST082Q13NA	0.927	Grup 2
11	ICTRES	2.996	Grup 1	65	ST095Q13NA	0.923	Grup 2
12	ST078Q05NA	2.855	Grup 1	66	ST103Q08NA	0.923	Grup 2
13	BSMJ	2.787	Grup 1	67	ST131Q06NA	0.916	Grup 3
14	ENVOPT	2.740	Grup 1	68	ST039Q02NA	0.863	Grup 3
15	REPEAT	2.732	Grup 1	69	ST082Q03NA	0.852	Grup 3
16	ST001D01T	2.702	Grup 1	70	ST061Q01NA	0.831	Grup 3
17	ST013Q01TA	2.693	Grup 1	71	ST104Q02NA	0.798	Grup 3
18	ST129Q08TA	2.692	Grup 1	72	ST082Q02NA	0.752	Grup 3
19	ST059Q03TA	2.683	Grup 1	73	ST123Q02NA	0.729	Grup 3
20	ST078Q08NA	2.664	Grup 1	74	HIISCED	0.650	Grup 3
21	GRADE	2.635	Grup 1	75	ST076Q02NA	0.619	Grup 3
22	ST063Q01NA	2.613	Grup 1	76	BFMJ2	0.607	Grup 3
23	ST098Q10NA	2.452	Grup 1	77	TMINs	0.603	Grup 3
24	ST063Q06NA	2.364	Grup 1	78	ST097Q03TA	0.600	Grup 3
25	ST095Q07NA	2.323	Grup 1	79	ST034Q06TA	0.561	Grup 3
26	ST092Q02TA	2.258	Grup 1	80	PERFEED	0.553	Grup 3
27	ST063Q04NA	2.211	Grup 1	81	EPIST	0.510	Grup 3
28	ST146Q05TA	2.026	Grup 1	82	ST098Q09TA	0.498	Grup 3
29	ST093Q05TA	1.986	Grup 1	83	FISCED	0.490	Grup 3
30	ST098Q06TA	1.985	Grup 1	84	ST071Q04NA	0.479	Grup 3
31	ST093Q03TA	1.960	Grup 1	85	ST071Q02NA	0.476	Grup 3
32	ENVAWARE	1.935	Grup 1	86	DISCLISCI	0.466	Grup 3
33	ST093Q01TA	1.931	Grup 1	87	ST146Q03TA	0.442	Grup 3
34	ST129Q01TA	1.919	Grup 2	88	ST064Q01NA	0.432	Grup 3
35	ST071Q05NA	1.847	Grup 2	89	ST104Q03NA	0.429	Grup 3
36	ST097Q04TA	1.733	Grup 2	90	ST146Q01TA	0.421	Grup 3
37	ST098Q02TA	1.724	Grup 2	91	ST146Q07NA	0.398	Grup 3
38	BELONG	1.706	Grup 2	92	ST062Q03TA	0.396	Grup 3
39	ST012Q07NA	1.672	Grup 2	93	ST005Q01TA	0.379	Grup 3
40	ST094Q01NA	1.659	Grup 2	94	ST104Q05NA	0.379	Grup 3
41	ST100Q01TA	1.656	Grup 2	95	ST100Q03TA	0.375	Grup 3
42	ST098Q01TA	1.613	Grup 2	96	TEACHSUP	0.355	Grup 3
43	ST007Q01TA	1.584	Grup 2	97	ST076Q08NA	0.328	Grup 3
44	MISCED	1.538	Grup 2	98	ST123Q04NA	0.315	Grup 3
45	ST011Q07TA	1.446	Grup 2	99	ST076Q10NA	0.281	Grup 3
46	ST034Q01TA	1.391	Grup 2				
47	ST071Q03NA	1.359	Grup 2				
48	ST003D02T	1.337	Grup 2				
49	ST097Q01TA	1.272	Grup 2				
50	IBTEACH	1.255	Grup 2				
51	SMINS	1.246	Grup 2				
52	ST123Q01NA	1.232	Grup 2				
53	ISCEDD	1.232	Grup 2				
54	ESCS	1.231	Grup 2				



Şekil 37: TR7 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.



Şekil 38: TR7 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı

4.8 TR8 Bölgesi PISA-2015 Bulguları

4.8.1 TR8 Bölgesi Hakkında Temel Bilgiler

2015 PISA uygulamasının Türkiye örnekleminin sekizinci bölgesi TR8 olarak adlandırılmıştır. TR8 bölgesi 2015 PISA uygulamasında ayrılan bölgelerden Batı Karadeniz'i temsil etmektedir. Batı Karadeniz olarak ayrılan bölge Zonguldak, Bartın, Karabük, Kastamonu, Çankırı, Sinop, Samsun, Amasya, Tokat illerini kapsamaktadır. TR8 bölgesinden (Batı Karadeniz) 7. Sınıf ve üzeri düzeyde eğitim gören 303 öğrencinin 2015 PISA sınavına katıldığı görülmektedir. 2015 PISA sınavına katılan tüm Türkiye örnekleminin %5.14'ünün Orta Anadolu bölgesinden katıldığını görmekteyiz.

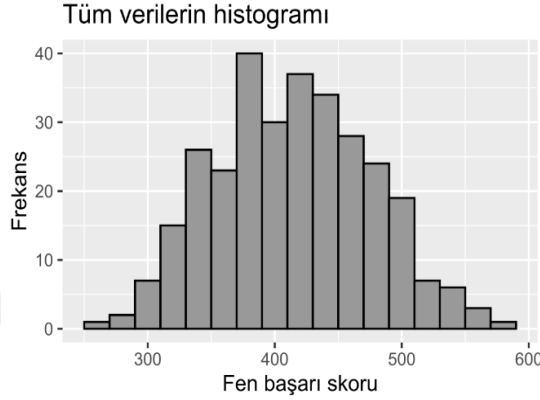
4.8.2 TR8 Bölgesinin Veri Yapısı

Şekil 39'da Türkiye'nin TR8 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı skorlarına ait istatistiksel özeti sunulmaktadır. Şekil 39'un üst panelinde, tüm veri setini temsil eden bir histogram ve istatistiksel özet bilgileri gösterilmektedir. Aynı şeklin alt panelinde ise, makine öğrenmesi modelinin oluşturulması ve modelin performansının değerlendirilmesi süreçlerinde kullanılacak olan eğitim ve test alt veri setlerine ait istatistiksel bilgiler bulunmaktadır. Bu alt veri setleri ana veri setinden rastgele bir şekilde ve literatüre uygun olarak %80 (Eğitim verisi) / %20 (Test verisi) oranında örneklenerek oluşturulmuştur (Gholamy ve diğerleri, 2018).

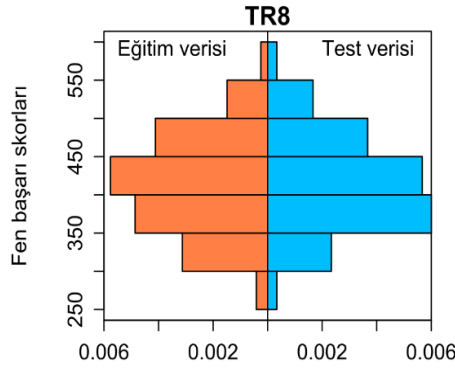
Şekil 39'un üst panelindeki histograma göre, fen puanlarının ortalama puan etrafında merkezi bir eğilimle dağıldığı görülmektedir. Bu TR8 bölgesindeki öğrencilerin (303 kişi) Fen başarı skorlarının normal dağılıma uyduğunu belirtmektedir. Ayrıca bu bölge için tüm veri setine ait ortalama puan 414.09, medyan 412.76, standart sapma 61.50 olarak hesaplanmıştır. Bunlara ek olarak minimum ve maksimum puanlar sırasıyla 266.75 ve 570.54 olup, birinci çeyrek (Q1) 370.86 ve üçüncü çeyrek (Q3) 456.58 olarak belirlenmiştir.

Şekil 39'un alt panelinde, eğitim veri setinin (243 kişi) ortalama puanı 413.56, medyanı 412.32 ve standart sapması 61.82 olarak tespit edilmiştir. Eğitim verisi için minimum ve maksimum puanlar sırasıyla 266.75 ve 570 olup, birinci çeyrek (Q1) 370.86 ve üçüncü çeyrek (Q3) 457.09 olarak belirlenmiştir. Bunların yanında Test veri setinin (60 kişi), ortalama puanı 416.17, medyanı 412.79 ve standart sapması 60.65 olarak tespit edilmiştir. Test verisi için minimum ve maksimum puanlar sırasıyla 278.51 ve 556.47 olup, birinci çeyrek (Q1) 371.45 ve üçüncü çeyrek (Q3) 454.40 olarak belirlenmiştir. Bu veri seti için

rasgele bir şekilde oluşturulan eğitim ve test alt veri setlerine ait dağılımlarının birbirlerinden anlamlı bir şekilde farklı olmadığı Kolmogorov-Smirnov testi sonucuna göre ($p > 0.05$) tespit edilmiştir.



İstatistik	TR8_verisi
Örneklem Sayısı	303.00
Ortalama	414.09
Medyan	412.76
Standart Sapma	61.50
Min.	266.75
Maks.	570.54
Q1	370.86
Q3	456.58



İstatistik	Eğitim_verisi	Test_verisi
Örneklem Sayısı	243.00	60.00
Ortalama	413.58	416.17
Medyan	412.32	412.79
Standart Sapma	61.82	60.65
Min	266.75	278.51
Maks	570.54	556.47
Q1	370.86	371.45
Q3	457.09	454.40

Kolmogorov-Smirnov Testi Sonuçları: p-değeri = 0.99
Dağılımlar arasında anlamlı bir fark: Yoktur!

Şekil 39: Türkiye'nin TR8 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenі olasılık yoğunluklarını göstermektedir)

4.8.3 TR8 Bölgesinde Makine Öğrenme Modeli

Tablo 26 TR8 (Batı Karadeniz) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı ve 10 tekrarlı çapraz doğrulama performans metriklerini ve optimize edilmiş hiperparametre değerlerini göstermektedir. Bu tabloda elde edilen bulgular, XGBoost, RF, SVR ve BRNN algoritmaları ile Kısım 3.3'te oluşturulan makine öğrenme modellerin 10 tekrarlı ve 10 katlı çapraz doğrulama tahmin performansları elde edilmiştir. Bu tablonun ayrıntılı sonuçları ve kısaltmalar için gerekli açıklamalar EK-1 içerisinde sunulmuştur. İncelenen ve oluşturulan makine öğrenmesi modellerin başarı

performansları RMSE, R^2 ve MAE metrikleri üzerinden değerlendirilmiş ve XGBoost modelinin en düşük RMSE (42.63), en yüksek R^2 (0.54) ve düşük MAE (34.90) değerleri ile eğitim verileri için en iyi performansı gösterdiği tespit edilmiştir. Buna karşın SVR, RF ve BRNN modelleri sırasıyla 44.16, 45.18 ve 46.52 RMSE değerleri ile XGBoost algoritmasının ardından sıralanmışlardır. Bu veriler, TR8 bölgesinin eğitim verileri için XGBoost modelinin Fen okuryazarlık skorlarının tahmininde diğer modellere göre daha iyi bir öğrenme ortaya koyduğunu göstermektedir.

Tablo 26: TR8 (Batı Karadeniz) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri

Model	RMSE	R^2	MAE	Hiperparametre değerleri
XGBoost	42.63	0.54	34.90	nrounds 209.00000000 max_depth 5.00000000 eta 0.05929089 gamma 2.59145337 colsample_bytree 0.32752042 min_child_weight 6.00000000 subsample 0.51286911
SVR	44.16	0.51	35.39	sigma 0.007016197 C 0.679956811
RF	45.18	0.49	36.58	mtry "65" splitrule "maxstat" min.node.size "15"
BRNN	46.52	0.46	38.13	neurons 16

Tablo 27'de TR8 bölgesine ait verileri için oluşturulan makine öğrenme algoritmalarının test verisi performans göstergeleri detaylı bir şekilde sunulmaktadır. Bu tabloya göre, XGBoost modelinin RMSE değeri 45.72, R^2 değeri 0.43 ve MAE değeri 37.08 olarak belirlenmiş ve diğer modellerden daha üstün bir performans sergilediği tespit edilmiştir. Ayrıca, XGBoost modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.562, güven aralığının alt sınırı 0.409 ve üst sınırı 0.685 ile yüksek bir uyumu gösterdiği belirlenmiştir. Buna karşın ikinci en iyi performans gösteren SVR modelinde, RMSE değeri 45.87, R^2 değeri 0.44 ve MAE değeri 37.73 ile XGBoost'a göre daha düşük olup, uyum korelasyon katsayısı merkezi tahmini 0.544 (lwr.ci: 0.398- upr.ci: 0.663) olarak hesaplanmıştır. BRNN ve RF modellerinde ise sırasıyla RMSE değerleri 46.84 ve 48.30, R^2 değerleri 0.39 ve 0.40, MAE değerleri 37.33 ve 38.94 olarak belirlenmiştir. Modelininlinin

uyum korelasyon katsayısı merkezi tahmini 0.571 (lwr.ci: 0.403- upr.ci: 0.702) iken, RF modelinin uyum korelasyon katsayısı merkezi tahmini 0.456 (lwr.ci: 0.318- upr.ci: 0.575) olarak hesaplanmıştır. Bu sonuçlar, XGBoost modelinin en düşük hata oranları ve en yüksek açıklama gücü ile diğer modellerden daha üstün performans sergilediğini ortaya koymaktadır.

Tablo 27: TR8 (Batı Karadeniz) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri

Model	RMSE	R^2	MAE	ρ_c
XGBoost	45.72	0.43	37.08	XGBoost.est 0.5624673 XGBoost.lwr.ci 0.4095537 XGBoost.upr.ci 0.6846395
SVR	45.87	0.44	37.73	SVR.est 0.5441168 SVR.lwr.ci 0.3982394 SVR.upr.ci 0.6631503
BRNN	46.84	0.39	37.33	BRNN.est 0.5714077 BRNN.lwr.ci 0.4031653 BRNN.upr.ci 0.7022866
RF	48.30	0.40	38.94	RF.est 0.4563993 RF.lwr.ci 0.3184991 RF.upr.ci 0.5753820

ρ_c = Uyum Korelasyon Katsayısı (Concordance Correlation Coefficient)

est = ρ_c için merkezi tahmin edilen değeri

lwr.ci = güven aralığının alt sınırını (lower confidence interval)

upr.ci = güven aralığının üst sınırını (upper confidence interval)

Tablo 26 ve Tablo 27 verilerine göre yapılan genel değerlendirmede, XGBoost modelinin hem eğitim hem de test verilerinde en yüksek performansı sergilediği görülmektedir. Bu durum, modelin genelleştirme kabiliyetinin güçlü olduğunu göstermektedir. Ayrıca test verilerinin uyum korelasyon katsayısının (ρ_c) yüksek olması, modelin tahminlerinin gerçek verilerle yüksek uyum gösterdiğini belirtmektedir. Bu, XGBoost modelinin hem eğitim hem de test verilerinde tutarlı ve güvenilir bir performans sergilediğinin kanıtıdır.

Diğer yandan SVR, BRNN ve RF modelleri ise eğitim ve test verileri arasında benzer hata oranlarına sahip olsalar da performansları XGBoost modellerine kıyasla daha düşük seviyelerde kalmıştır. Bu durum SVR, BRNN ve RF modellerinin genelleştirme kabiliyetinin sınırlı olduğunu ve yeni veriler üzerinde daha düşük performans sergilediğini

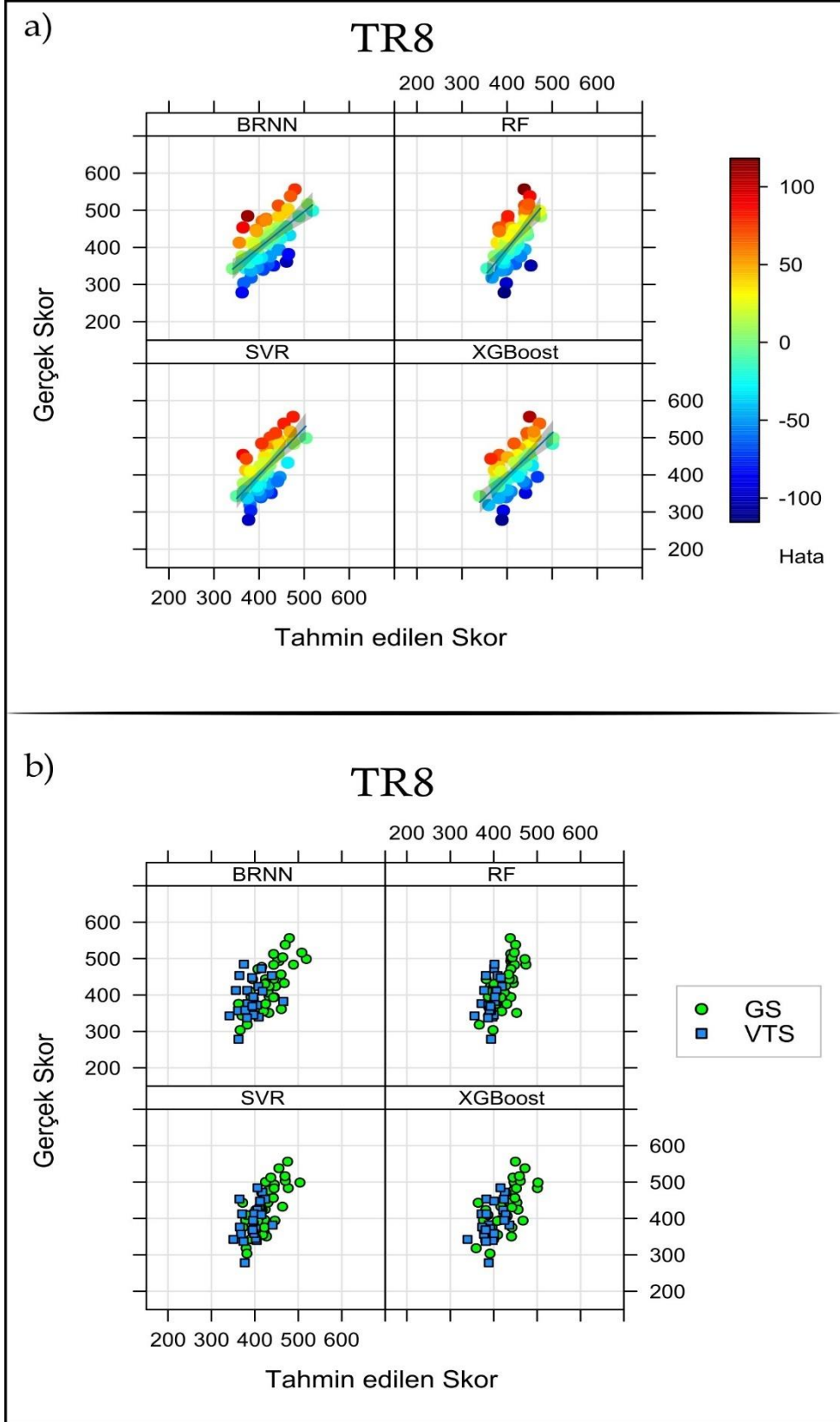
göstermektedir. Her üç modelin de ρ_c değerleri, tahminlerinin gerçek verilerle makul bir uyum gösterdiğini belirtmektedir.

Şekil 40, Tablo 27’de elde edilen performans metriklerine ait çapraz doğrulama diyagramlarını tanımlamaktadır. Şekil 40a’da tüm modellerin hata dağılımlarının regresyon çizgisi boyunca ve yakınlarında dağıldığı görülmektedir. Bu durum tahmin hatalarının dağılımının normal yapıda olduğunu ve hata ortalamasının da sıfıra yakın olduğunu göstermektedir. Bu sonuçlar Tablo 27 de elde edilen bulgular ile uyumludur. Bunlarla birlikte XGBoost modelinin yüksek tahmin performansı sergilediği, Şekil 40a’ daki çapraz doğrulama diyagramının hata dağılımlarının regresyon çizgisi etrafında sıkı şekilde toplanmasından dolayı daha açık bir şekilde görülmektedir.

Şekil 40b, aynı test verileri için modellerin farklı okul türlerine (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim) ait tahmin edilen ve gerçek skorlarının karşılaştırmasını göstermektedir. Burada özellikle tanımlayıcı regresyon çizgisinden sapan ve modellerin tahmin etme performansının düşmesine neden olan okul grubunun Mesleki ve Teknik Ortaöğretim (VTS) okul türü olduğu dikkat çekmektedir. Bununla birlikte XGBoost, her tüm okul türlerinde de tahmin edilen skorların gerçek skorlarla en uyumlu olduğu model olarak öne çıkmaktadır. Ayrıca TR8 bölgesindeki Fen başarı skorları için genel ortaokul (GS) başarı düzeyinin VTS okul türüne göre belirgin şekilde artış göstermiş olması dikkat çekicidir. Özellikle okul türünün modellerin oluşturulmasında bir değişken olarak kullanılmamış olmasına rağmen, makine öğrenme algoritmasının bu başarı farklılıklarını diğer değişkenlerin yardımı ile tespit edebilmesi ve tahminlerini buna göre gerçekleştirmesi önemli bir başarı olduğu düşünülmektedir. Bu bulgular, XGBoost modelinin eğitim verileri üzerindeki güçlü performansını ve genelleştirme kabiliyetini destekler niteliktedir. XGBoost'un, farklı okul türlerindeki öğrencilerin performansını doğru bir şekilde tahmin edebilmesi, eğitim politikaları ve müdahalelerinin planlanmasında doğru tahmin skorları önerebileceğini ve bu konuda değerli bir araç olarak kullanılabileceğini göstermektedir. Özellikle Mesleki ve Teknik Ortaöğretim (VTS) öğrencileri için tahminlerdeki sapmalara rağmen, XGBoost'un genel olarak diğer modellere göre daha yüksek doğrulukla sonuçlar vermesi, bu modelin TR8 bölgesi için geniş bir yelpazede güvenilir tahminler oluşturabileceğini göstermektedir.

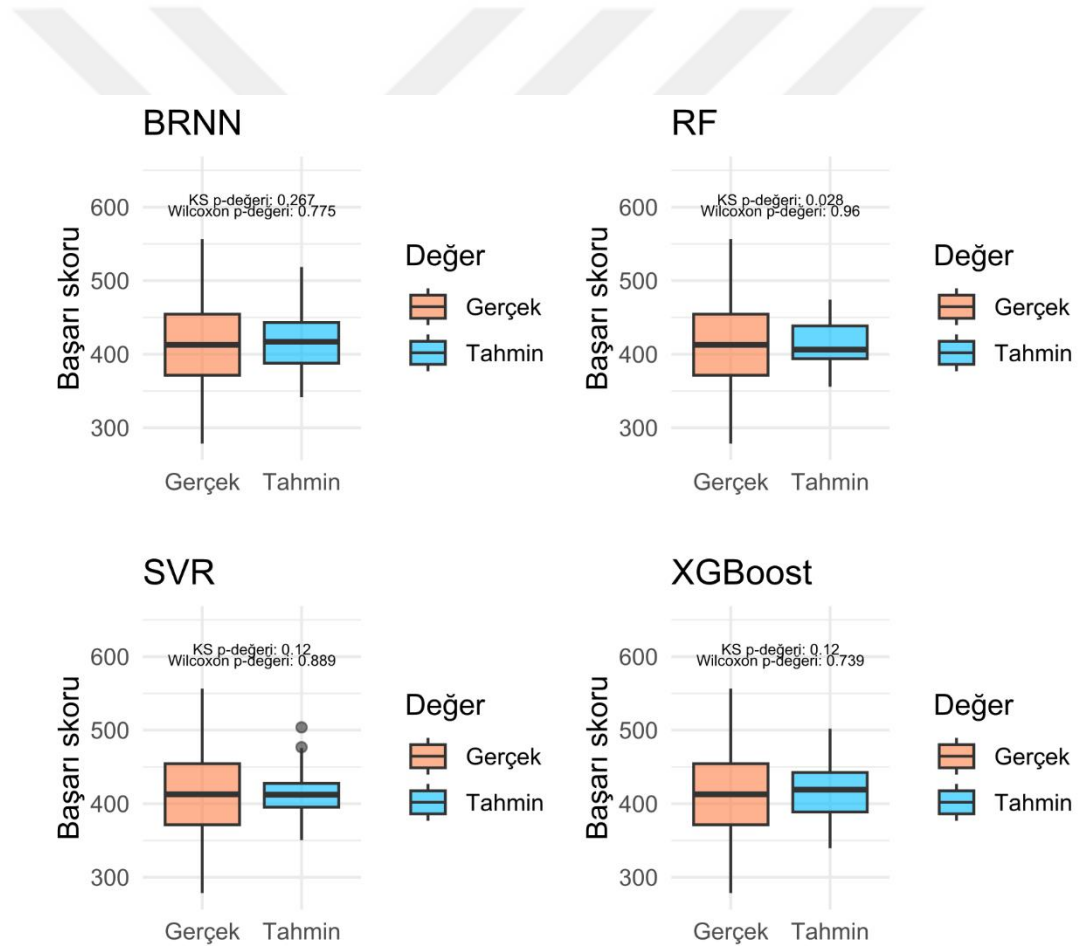
Şekil 41, farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost) tahmin ettiği başarı skorları ile gerçek başarı skorlarının karşılaştırmalı kutu grafiğini sunmaktadır. Şekil 40’ta sunulan çapraz doğrulama bulgularına göre oluşturulmuş bu grafikler, modellerin tahmin performansını ve gerçek verilerle uyumunu değerlendirmek için önemli bir yaklaşım sağlamaktadır. Buna göre tüm modellerin tahmin skorlarının hem Kolmogorov -Simirnov testi (KS) hem de Wilcoxon sıralama toplamı testine göre gerçek skorların genel dağılımlarından ve merkezi eğilimlerinden anlamlı ölçüde bir farklılığı ($p > 0.05$) olmadığı tespit edilmiştir. Bu durum makine öğrenmesi modellerinin tahmin edilen ve gerçek başarı skorları arasındaki uyumun yüksek olduğunu ortaya koymaktadır.





Şekil 40: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)

Şekil 41, en yüksek performans seviyesi gösteren XGBoost algoritması açısından değerlendirildiğinde, Kutu grafiklerinin gerçek ve tahmin skorları için temsil ettiği ortalama ve medyan değerlerinin neredeyse örtüştüğü ve dağılımların (kutu dışındaki uç örnekler için yeterli seviyelerde olmasa da) Q1-Q3 aralığı için oldukça yakın olduğunu görülmektedir. Şekil 41’deki bulgulara göre Tahmin edilen ve gerçek skorların genel dağılımlarının (KS p-değeri 0.12) ve merkezi eğilimlerinin (Wilcoxon p-değeri 0.739), istatistiksel olarak birbirlerinden farklı olmadığı belirlenmiştir. Bu durum, XGBoost algoritmasının TR8 verileri üzerinde en yüksek doğruluk ve güvenilirliği sunmakla birlikte hem bireysel gerçek skorları hem de gerçek skorlara ait dağılım yapısını doğru şekilde tahmin edebilme yeteneğini ortaya koymaktadır.



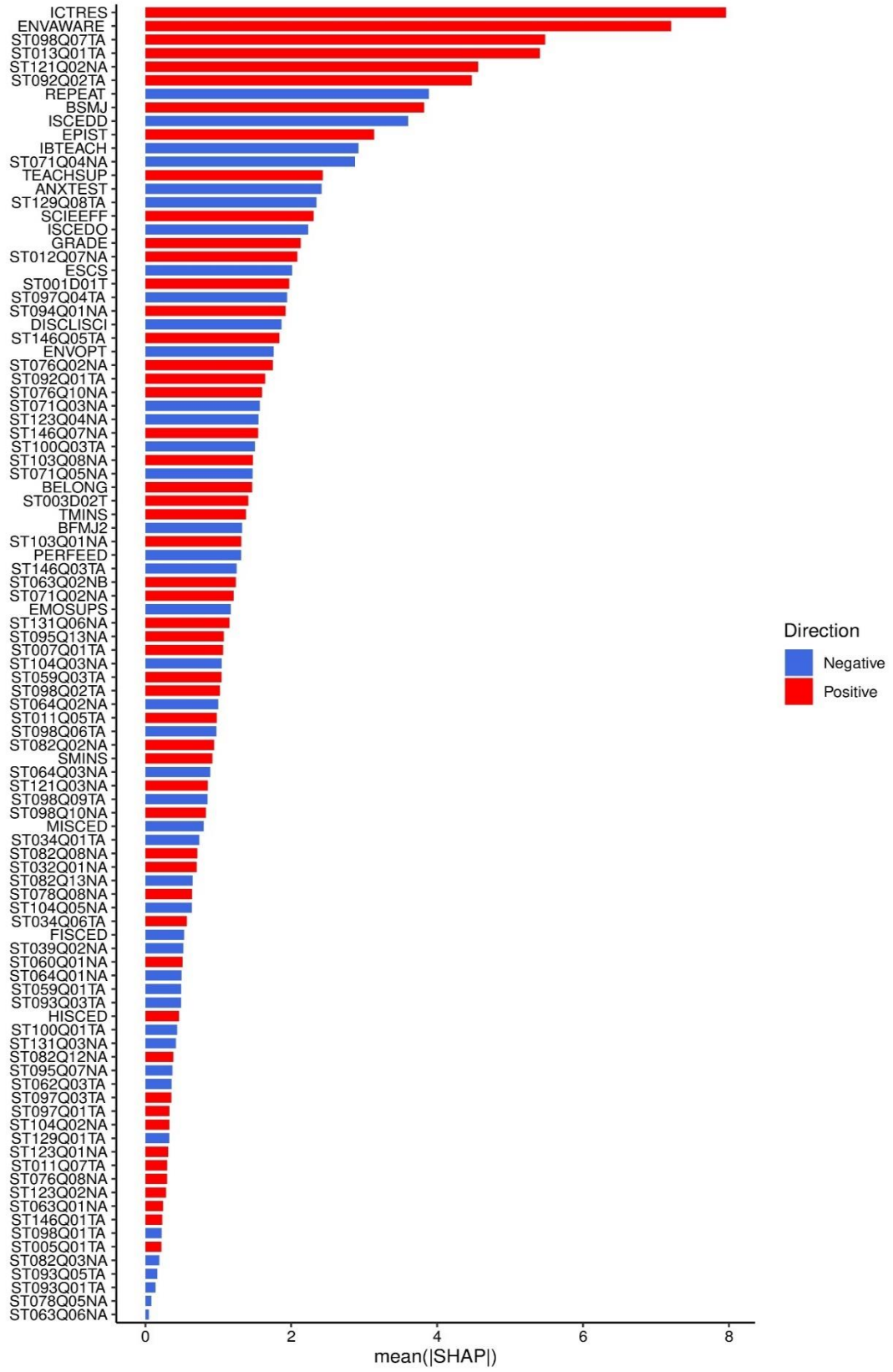
Şekil 41: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.

4.8.4 TR8 Bölgesi İçin Değişkenlerin Önem Seviyeleri

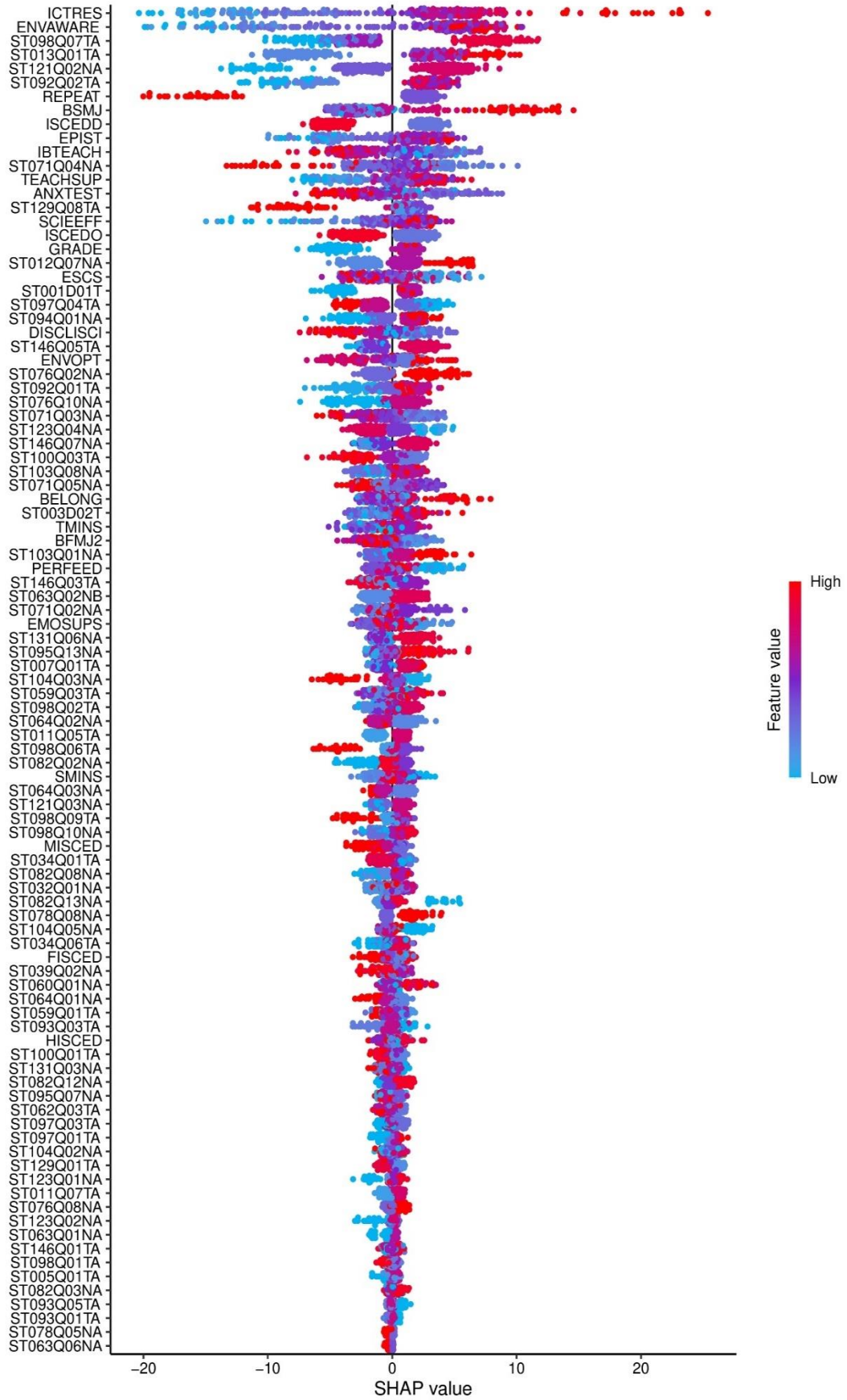
PISA 2015 Türkiye örnekleme TR8 bölgesi Fen okuryazarlık alanı için en iyi tahmin performansı gösteren XGBoost algoritmasına göre SHAP (SHapley Additive exPlanations) değerleri kullanılarak) değişkenlerin önem parametreleri belirlenmiştir. Sonuçlar her bir değişkenin ortalama SHAP değerlerinin mutlak değerine göre azalan sırada Tablo 28’de, sıralanmıştır. Bu sıralamaya göre diğer bölgelerdeki değişkenlerin karşılaştırması ise Kısım 4.13’te Jaccard indeksine göre yapılacaktır. Burada model için önemli olan sıralı değişkenler üç önem grubuna ayrılmış ve bu işlem model için kullanılan toplam değişkenleri 3 eşit parçaya ayırarak gerçekleştirilmiştir. Burada yüksek derecede önemli değişkenler 1. Grup (ilk 33 tanesi), orta derecede önemli değişkenler 2. Grup (34- 66) ve düşük derecede önemli değişkenler 3. Grup (67 - 99) olarak etiketlenmiştir. Ayrıca Şekil 42 ile Tablo 28’de özetlenen değişkenlerin ortalama SHAP değerleri görselleştirilmiştir. Bu şekil her bir tahmin edici değişkenin model tahminlerine olan ortalama etkisini ve yönünü (pozitif ve negatif katkıları) vurgulamaktadır. Şekil 43 her bir değişkenin test veri setindeki her bir gözlemdeki etkisini göstermektedir. Bu şekil içerisinde her bir nokta, bir veri noktasının bir özelliği için SHAP değerini temsil eder. Bu şeklin sağ tarafında pozitif SHAP değerleri, sol tarafında ise negatif SHAP değerleri bulunmaktadır. Tahmin edici değişkenlerin model çıktısına olan etkisi, bu SHAP değerlerinin büyüklüğüne göre belirlenir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Tablo 28, Şekil 42 ve 43’teki bulgulara göre, TR8 bölgesi için model tahmininde önem seviyesi yüksek ilk 10 değişken, parantez içleri SHAP değerlerini ve model katkı yönlerini göstermek üzere, şu şekilde sıralanmıştır: ICTRES (7.959, pozitif), ENVWARE (7.207, pozitif), ST098Q07TA (5.480, pozitif), ST013Q01TA (5.410, pozitif), ST121Q02NA (4.562, pozitif), ST092Q02TA (4.474, pozitif), REPEAT (3.888, negatif), BSMJ (3.821, pozitif), ISCEDD (3.601, negatif), EPIST (3.136, pozitif). Şekil 4 de Kırmızı ve mavi renkler, sırasıyla pozitif ve negatif yönlü etkileri temsil etmektedir. Bu, modelin tahminlerinin hangi özellikler tarafından ne kadar ve hangi yönde etkilendiğini daha net bir şekilde anlaşılmasına yardımcı olmaktadır. Buna göre negatif özellikteki değişkenlerin (REPEAT, ISCEDD) değerlerindeki artış öğrencilerin Fen başarı skorlarını düşürmekte buna karşı pozitif özellikteki değişkenlerin değerlerindeki artış ise öğrencilerin Fen başarı skorlarını arttırıcı yönde etki etmektedir.

Tablo 28: TR8 (Batı Karadeniz) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.

ID	Değişken	SHAP değeri	Önem Grubu	ID	Değişken	SHAP değeri	Önem Grubu
1	ICTRES	7.959	Grup 1	55	ST082Q02NA	0.941	Grup 2
2	ENVAWARE	7.207	Grup 1	56	SMINS	0.919	Grup 2
3	ST098Q07TA	5.480	Grup 1	57	ST064Q03NA	0.890	Grup 2
4	ST013Q01TA	5.410	Grup 1	58	ST121Q03NA	0.855	Grup 2
5	ST121Q02NA	4.562	Grup 1	59	ST098Q09TA	0.852	Grup 2
6	ST092Q02TA	4.474	Grup 1	60	ST098Q10NA	0.829	Grup 2
7	REPEAT	3.888	Grup 1	61	MISCED	0.799	Grup 2
8	BSMJ	3.821	Grup 1	62	ST034Q01TA	0.738	Grup 2
9	ISCEDD	3.601	Grup 1	63	ST082Q08NA	0.713	Grup 2
10	EPIST	3.136	Grup 1	64	ST032Q01NA	0.704	Grup 2
11	IBTEACH	2.920	Grup 1	65	ST082Q13NA	0.648	Grup 2
12	ST071Q04NA	2.874	Grup 1	66	ST078Q08NA	0.640	Grup 2
13	TEACHSUP	2.432	Grup 1	67	ST104Q05NA	0.637	Grup 3
14	ANXTEST	2.416	Grup 1	68	ST034Q06TA	0.570	Grup 3
15	ST129Q08TA	2.345	Grup 1	69	FISCED	0.530	Grup 3
16	SCIEEFF	2.306	Grup 1	70	ST039Q02NA	0.519	Grup 3
17	ISCEDO	2.232	Grup 1	71	ST060Q01NA	0.511	Grup 3
18	GRADE	2.130	Grup 1	72	ST064Q01NA	0.496	Grup 3
19	ST012Q07NA	2.083	Grup 1	73	ST059Q01TA	0.491	Grup 3
20	ESCS	2.012	Grup 1	74	ST093Q03TA	0.489	Grup 3
21	ST001D01T	1.972	Grup 1	75	HISCED	0.462	Grup 3
22	ST097Q04TA	1.943	Grup 1	76	ST100Q01TA	0.437	Grup 3
23	ST094Q01NA	1.921	Grup 1	77	ST131Q03NA	0.420	Grup 3
24	DISCLSCI	1.868	Grup 1	78	ST082Q12NA	0.382	Grup 3
25	ST146Q05TA	1.838	Grup 1	79	ST095Q07NA	0.371	Grup 3
26	ENVOPT	1.758	Grup 1	80	ST062Q03TA	0.360	Grup 3
27	ST076Q02NA	1.748	Grup 1	81	ST097Q03TA	0.355	Grup 3
28	ST092Q01TA	1.642	Grup 1	82	ST097Q01TA	0.330	Grup 3
29	ST076Q10NA	1.601	Grup 1	83	ST104Q02NA	0.330	Grup 3
30	ST071Q03NA	1.569	Grup 1	84	ST129Q01TA	0.328	Grup 3
31	ST123Q04NA	1.549	Grup 1	85	ST123Q01NA	0.313	Grup 3
32	ST146Q07NA	1.546	Grup 1	86	ST011Q07TA	0.299	Grup 3
33	ST100Q03TA	1.502	Grup 1	87	ST076Q08NA	0.298	Grup 3
34	ST103Q08NA	1.473	Grup 2	88	ST123Q02NA	0.285	Grup 3
35	ST071Q05NA	1.471	Grup 2	89	ST063Q01NA	0.243	Grup 3
36	BELONG	1.463	Grup 2	90	ST146Q01TA	0.232	Grup 3
37	ST003D02T	1.412	Grup 2	91	ST098Q01TA	0.223	Grup 3
38	TMINS	1.380	Grup 2	92	ST005Q01TA	0.219	Grup 3
39	BFMJ2	1.326	Grup 2	93	ST082Q03NA	0.191	Grup 3
40	ST103Q01NA	1.313	Grup 2	94	ST093Q05TA	0.163	Grup 3
41	PERFEED	1.312	Grup 2	95	ST093Q01TA	0.141	Grup 3
42	ST146Q03TA	1.252	Grup 2	96	ST078Q05NA	0.083	Grup 3
43	ST063Q02NB	1.240	Grup 2	97	ST063Q06NA	0.049	Grup 3
44	ST071Q02NA	1.209	Grup 2	98	ST063Q04NA	0	Grup 3
45	EMOSUPS	1.172	Grup 2	99	ST061Q01NA	0	Grup 3
46	ST131Q06NA	1.153	Grup 2				
47	ST095Q13NA	1.075	Grup 2				
48	ST007Q01TA	1.064	Grup 2				
49	ST104Q03NA	1.045	Grup 2				
50	ST059Q03TA	1.044	Grup 2				
51	ST098Q02TA	1.021	Grup 2				
52	ST064Q02NA	0.999	Grup 2				
53	ST011Q05TA	0.979	Grup 2				
54	ST098Q06TA	0.973	Grup 2				



Şekil 42: TR8 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.



Şekil 43: TR8 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.

4.9 TR9 Bölgesi PISA-2015 Bulguları

4.9.1 TR9 Bölgesi Hakkında Temel Bilgiler

2015 PISA uygulamasının Türkiye örnekleminin dokuzuncu bölgesi TR9 olarak adlandırılmıştır. TR9 bölgesi 2015 PISA uygulamasında ayrılan bölgelerden Doğu Karadeniz'i temsil etmektedir. Doğu Karadeniz olarak ayrılan bölge Ordu, Giresun, Gümüşhane, Trabzon, Rize ve Artvin illerini kapsamaktadır. TR9 bölgesinden (Doğu Karadeniz) 7. Sınıf ve üzeri düzeyde eğitim gören 194 öğrencinin 2015 PISA sınavına katıldığı görülmektedir. 2015 PISA sınavına katılan tüm Türkiye örnekleminin %3.29'unun Doğu Karadeniz bölgesinden katıldığını görmekteyiz. Tüm Türkiye örneklem bölgelerinden en az sayıda öğrenci katılımının Doğu Karadeniz olduğunu görmekteyiz.

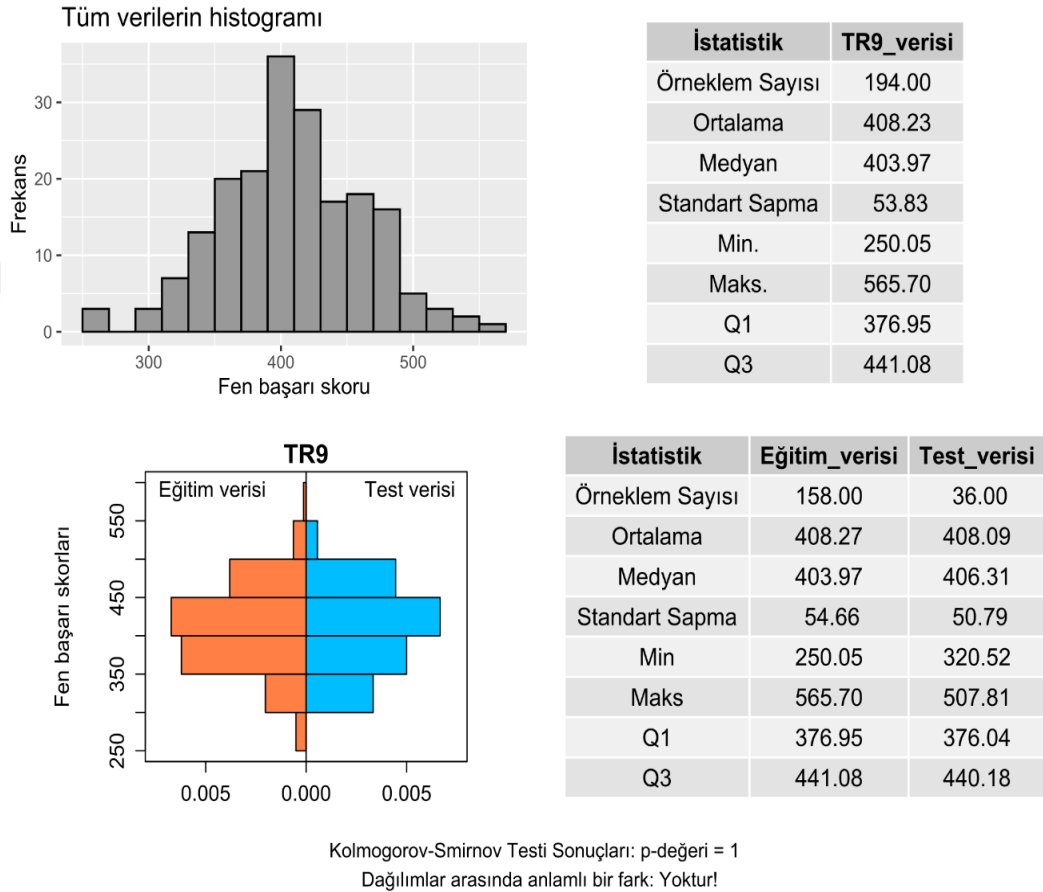
4.9.2 TR9 Bölgesinin Veri Yapısı

Şekil 44'te, Türkiye'nin TR9 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özeti sunulmaktadır. Şekil 44'ün üst panelinde, tüm veri setini temsil eden bir histogram ve istatistiksel özet bilgileri gösterilmektedir. Aynı şeklin alt panelinde ise, makine öğrenmesi modelinin oluşturulması ve modelin performansının değerlendirilmesi süreçlerinde kullanılacak olan eğitim ve test alt veri setlerine ait istatistiksel bilgiler bulunmaktadır. Bu alt veri setleri ana veri setinden rastgele bir şekilde ve literatüre uygun olarak %80 (Eğitim verisi) / %20 (Test verisi) oranında örneklenerek oluşturulmuştur (Gholamy ve diğerleri, 2018).

Şekil 44'ün üst panelindeki histograma göre, fen puanlarının ortalama puan etrafında merkezi bir eğilimle dağıldığı görülmektedir. Bu TR9 bölgesindeki öğrencilerin (194 kişi) Fen başarı puanlarının normal dağılıma uyduğunu belirtmektedir. Ayrıca bu bölge için tüm veri setine ait ortalama puan 408.23, medyan 403.97, standart sapma 53.83 olarak hesaplanmıştır. Bunlara ek olarak minimum ve maksimum puanlar sırasıyla 250.05 ve 565.70 olup, birinci çeyrek (Q1) 376.95 ve üçüncü çeyrek (Q3) 441.08 olarak belirlenmiştir.

Şekil 44'ün alt panelinde, Eğitim veri setinin (158 kişi) ortalama puanı 408.27, medyanı 403.97 ve standart sapması 54.66 olarak tespit edilmiştir. Eğitim verisi için minimum ve maksimum puanlar sırasıyla 250.05 ve 565.70 olup, birinci çeyrek (Q1) 376.95 ve üçüncü çeyrek (Q3) 441.08 olarak belirlenmiştir. Bunların yanında Test veri setinin (36 kişi), ortalama puanı 408.09, medyanı 406.31 ve standart sapması 50.79 olarak tespit edilmiştir. Test verisi için minimum ve maksimum puanlar sırasıyla 250.05 ve 507.81 olup,

birinci çeyrek (Q1) 376.04 ve üçüncü çeyrek (Q3) 440.18 olarak belirlenmiştir. Bu veri seti için rasgele bir şekilde oluşturulan eğitim ve test alt veri setlerine ait dağılımlarının birbirlerinden anlamlı bir şekilde farklı olmadığı Kolmogorov-Smirnov testi sonucuna göre ($p > 0.05$) tespit edilmiştir.



Şekil 44: Türkiye'nin TR8 bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)

4.9.3 TR9 Bölgesinde Makine Öğrenme Modeli

Tablo 29, TR9 (Doğu Karadeniz) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı ve 10 tekrarlı çapraz doğrulama performans metriklerini ve optimize edilmiş Hiperparametre değerlerini göstermektedir. Bu tabloda elde edilen bulgular, XGBoost, RF, SVR ve BRNN algoritmaları ile Kısım 3.3'te oluşturulan makine öğrenme modellerin 10 tekrarlı ve 10 katlı çapraz doğrulama tahmin performansları

elde edilmiştir. Bu tablonun ayrıntılı sonuçları ve kısaltmalar için gerekli açıklamalar EK-1 içerisinde sunulmuştur. İncelenen ve oluşturulan makine öğrenmesi modellerin başarı performansları RMSE, R^2 ve MAE metrikleri üzerinden değerlendirilmiş ve SVR modelinin en düşük RMSE (40.25), en yüksek R^2 (0.48) ve düşük MAE (32.79) değerleri ile eğitim verileri için en iyi performansı gösterdiği tespit edilmiştir. Buna karşın RF, XGBoost ve BRNN modelleri sırasıyla, 41.33, 42.14 ve 47.72 RMSE değerleri ile SVR algoritmasının ardından sıralanmışlardır. Bu veriler, TR9 bölgesinin eğitim verileri için SVR modelinin Fen okuryazarlık puanlarının tahmininde diğer modellere göre daha iyi bir öğrenme ortaya koyduğunu göstermektedir.

Tablo 29: TR9 (Doğu Karadeniz) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri

Model	RMSE	R^2	MAE	Hiperparametre değerleri
SVR	40.25	0.48	32.79	sigma 0.002548514 C 1.756834059
RF	41.33	0.50	33.62	mtry "60" splitrule "extratrees" min.node.size "5"
XGBoost	42.14	0.43	33.25	nrounds 234.00000000 max_depth 1.00000000 eta 0.06910115 gamma 2.40063659 colsample_bytree 0.57295894 min_child_weight 19.00000000 subsample 0.74068099
BRNN	44.71	0.37	36.12	neurons 20

Tablo 30'da, TR9 bölgesine ait verileri için oluşturulan makine öğrenme algoritmalarının test verisi performans göstergeleri detaylı bir şekilde sunulmaktadır. Bu tabloya göre, SVR modelinin RMSE değeri 36.35, R^2 değeri 0.47 ve MAE değeri 28.82 olarak belirlenmiş ve diğer modellerden daha üstün bir performans sergilediği tespit edilmiştir. Ayrıca, SVR modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.651, güven aralığının alt sınırı 0.440 ve üst sınırı 0.793 ile yüksek bir uyumu gösterdiği belirlenmiştir. Buna karşın ikinci en iyi performans gösteren XGBoost modelinde, RMSE değeri 37.10, R^2 değeri 0.46 ve MAE değeri 29.51 ile SVR modeline göre daha düşük olup, uyum korelasyon katsayısı merkezi tahmini 0.665 (lwr.ci: 0.447- upr.ci: 0.809) olarak hesaplanmıştır. RF ve BRNN modellerinde ise sırasıyla RMSE değerleri 37.73

ve 46.01, R^2 değerleri 0.49 ve 0.31, MAE değerleri 31.22 ve 34.89 olarak belirlenmiştir. RF modelinin uyum korelasyon katsayısı merkezi tahmini 0.531 (lwr.ci: 0.361- upr.ci: 0.668) iken, BRNN modelinin uyum korelasyon katsayısı merkezi tahmini 0.542 (lwr.ci: 0.272- upr.ci: 0.733) olarak hesaplanmıştır. Bu sonuçlar, SVR modelinin en düşük hata oranları ve en yüksek açıklama gücü ile diğer modellerden daha üstün performans sergilediğini ortaya koymaktadır.

Tablo 30: TR9 (Doğu Karadeniz) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri

Model	RMSE	R^2	MAE	ρ_c
SVR	36.35	0.47	28.82	SVR.est 0.6506932 SVR.lwr.ci 0.4405586 SVR.upr.ci 0.7932273
XGBoost	37.10	0.46	29.51	XGBoost.est 0.6651506 XGBoost.lwr.ci 0.4470215 XGBoost.upr.ci 0.8086162
RF	37.73	0.49	31.22	RF.est 0.5315184 RF.lwr.ci 0.3606952 RF.upr.ci 0.6678396
BRNN	46.01	0.31	34.89	BRNN.est 0.5421009 BRNN.lwr.ci 0.2717284 BRNN.upr.ci 0.7331586

ρ_c = Uyum Korelasyon Katsayısı (Concordance Correlation Coefficient)

est = ρ_c için merkezi tahmin edilen değeri

lwr.ci = güven aralığının alt sınırını (lower confidence interval)

upr.ci = güven aralığının üst sınırını (upper confidence interval)

Tablo 29 ve Tablo 30 verilerine göre yapılan genel değerlendirmede, SVR modelinin hem eğitim hem de test verilerinde en yüksek performansı sergilediği görülmektedir. Bu durum, modelin genelleştirme kabiliyetinin güçlü olduğunu göstermektedir. Ayrıca test verilerinin uyum korelasyon katsayısının (ρ_c) yüksek olması, modelin tahminlerinin gerçek verilerle yüksek uyum gösterdiğini belirtmektedir. Bu, SVR modelinin TR9 bölgesi için hem eğitim hem de test verilerinde tutarlı ve güvenilir bir performans sergilediğinin kanıtıdır.

Diğer yandan XGBoost, RF ve BRNN modelleri ise eğitim ve test verileri arasında benzer hata oranlarına sahip olsalar da performansları SVR modellerine kıyasla daha düşük seviyelerde kalmıştır. Bu durum XGBoost, RF ve BRNN modellerinin genelleştirme kabiliyetinin sınırlı olduğunu ve yeni veriler üzerinde daha düşük performans sergilediğini

göstermektedir. Her üç modelin de ρ_c değerleri, tahminlerinin gerçek verilerle makul bir uyum gösterdiğini belirtmektedir.

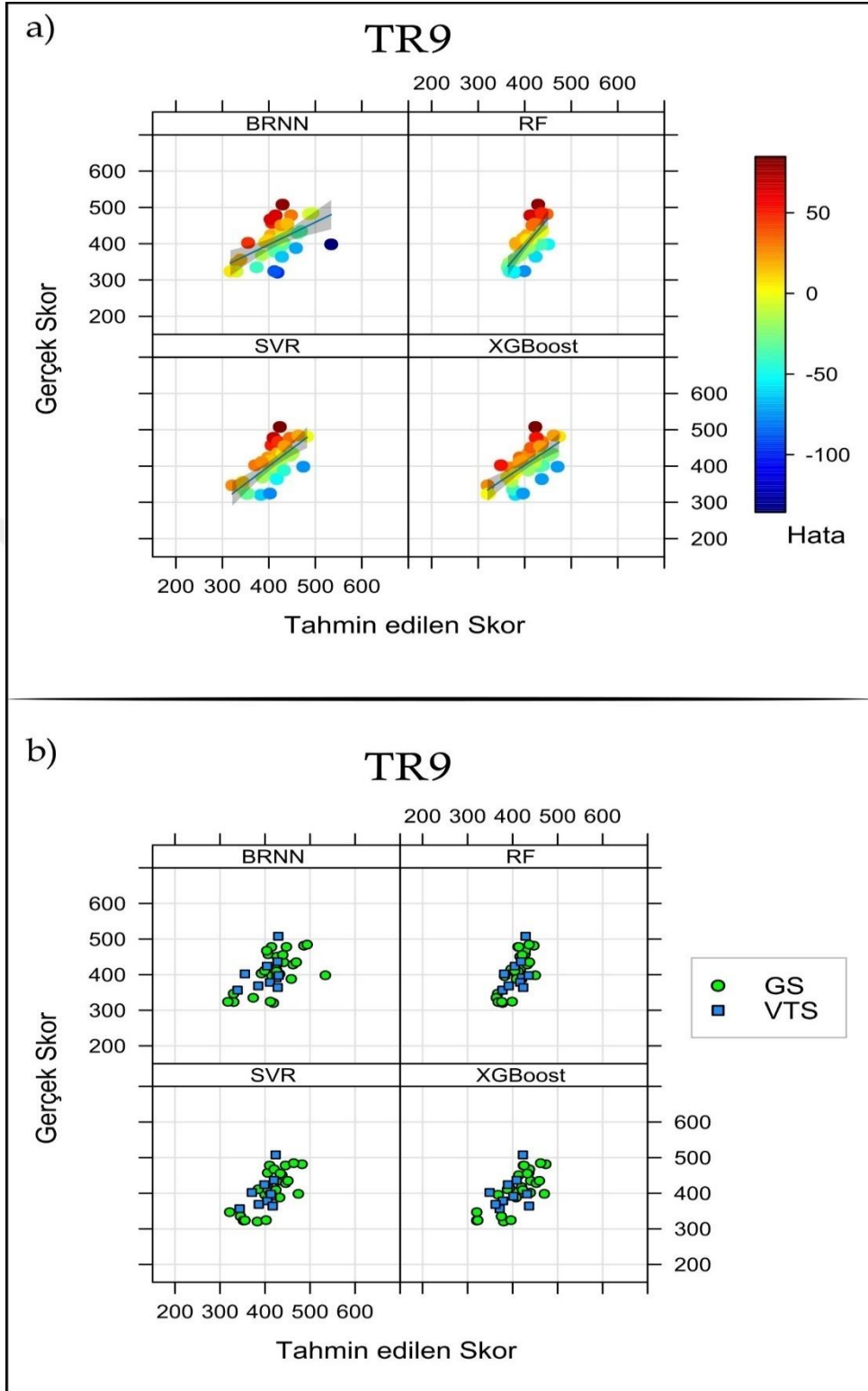
Şekil 45, Tablo 30 da elde edilen performans metriklerine ait çapraz doğrulama diyagramlarını tanımlamaktadır. Şekil 45a'da tüm modellerin hata dağılımlarının regresyon çizgisi boyunca ve yakınlarında dağıldığı görülmektedir. Bu durum tahmin hatalarının dağılımının normal yapıda olduğunu ve hata ortalamasının da sıfıra yakın olduğunu göstermektedir. Bu sonuçlar Tablo 30 da elde edilen bulgular ile uyumludur. Bunlarla birlikte SVR modelinin yüksek tahmin performansı sergilediği, Şekil 45a'daki çapraz doğrulama diyagramının hata dağılımlarının regresyon çizgisi etrafında sıkı şekilde toplanmasından dolayı daha açık bir şekilde görülmektedir.

Şekil 45b, aynı test verileri için modellerin farklı okul türlerine (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim) ait tahmin edilen ve gerçek skorlarının karşılaştırmasını göstermektedir. Burada özellikle tanımlayıcı regresyon çizgisinden sapan ve modellerin tahmin etme performansının düşmesine neden olan okul grubunun Mesleki ve Teknik Ortaöğretim (VTS) okul türü olduğu dikkat çekmektedir. Bununla birlikte SVR, her tüm okul türlerinde de tahmin edilen skorların gerçek skorlarla en uyumlu olduğu model olarak öne çıkmaktadır. Ayrıca TR9 bölgesindeki Fen başarı skorları için genel ortaokul (GS) başarı düzeyinin VTS okul türüne göre belirgin şekilde artış göstermiş olması dikkat çekicidir. Özellikle okul türünün modellerin oluşturulmasında bir değişken olarak kullanılmamış olmasına rağmen, makine öğrenme algoritmasının bu başarı farklılıklarını diğer değişkenlerin yardımı ile tespit edebilmesi ve tahminlerini buna göre gerçekleştirmesi önemli bir başarı olduğu düşünülmektedir. Bu bulgular, SVR modelinin eğitim verileri üzerindeki güçlü performansını ve genelleştirme kabiliyetini destekler niteliktedir. SVR'nin, farklı okul türlerindeki öğrencilerin performansını doğru bir şekilde tahmin edebilmesi, eğitim politikaları ve müdahalelerinin planlanmasında doğru tahmin skorları önerebileceğini ve bu konuda değerli bir araç olarak kullanılabileceğini göstermektedir. Özellikle Mesleki ve Teknik Ortaöğretim (VTS) öğrencileri için tahminlerdeki sapmalara rağmen, SVR'nin genel olarak diğer modellere göre daha yüksek doğrulukla sonuçlar vermesi, bu modelin TR9 bölgesi için geniş bir yelpazede güvenilir tahminler oluşturabileceğini göstermektedir.

Şekil 46, farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost) tahmin ettiği başarı skorları ile gerçek başarı skorlarının karşılaştırmalı kutu grafiğini sunmaktadır.

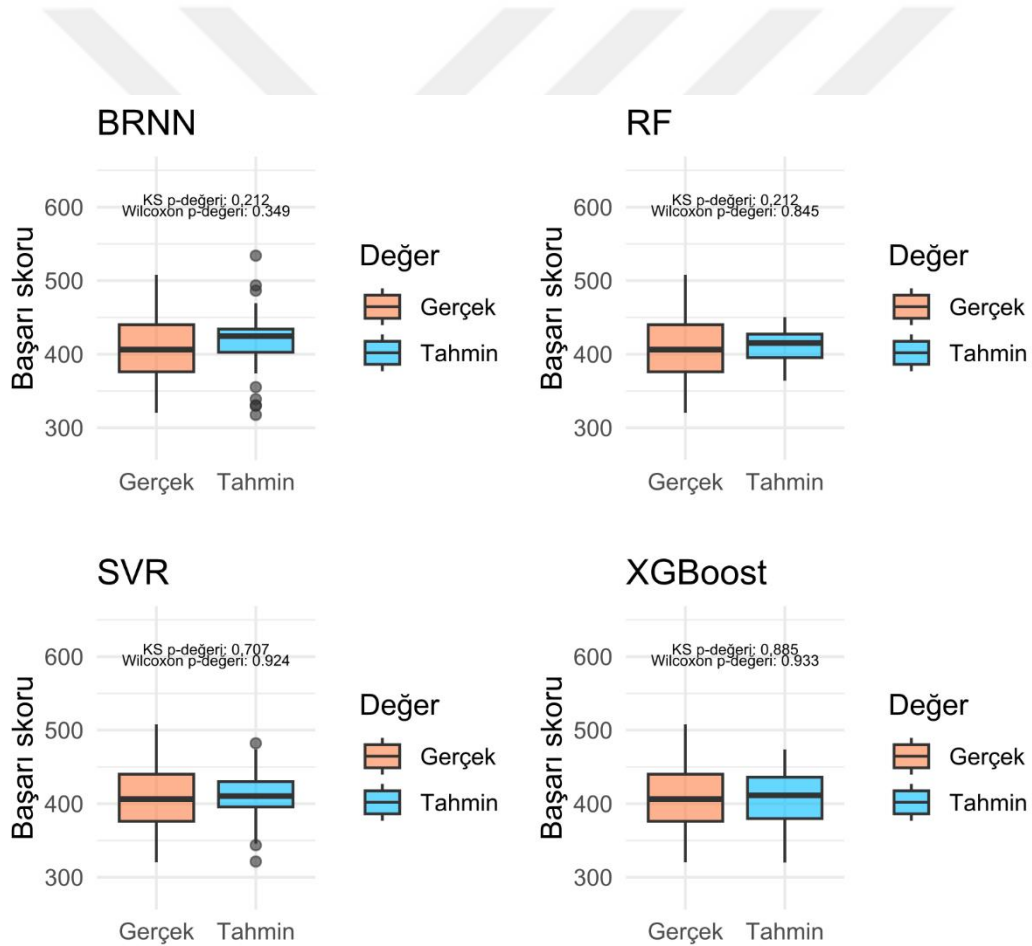
Şekil 45’te sunulan çapraz doğrulama bulgularına göre oluşturulmuş bu grafikler, modellerin tahmin performansını ve gerçek verilerle uyumunu değerlendirmek için önemli bir yaklaşım sağlamaktadır. Buna göre tüm modellerin tahmin skorlarının hem Kolmogorov -Simirnov testi (KS) hem de Wilcoxon sıralama toplamı testine göre gerçek skorların genel dağılımlarından ve merkezi eğilimlerinden anlamlı ölçüde bir farklılığı ($p > 0.05$) olmadığı tespit edilmiştir. Bu durum makine öğrenmesi modellerinin tahmin edilen ve gerçek başarı skorları arasındaki uyumun yüksek olduğunu ortaya koymaktadır.





Şekil 45: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)

Şekil 46, en yüksek performans seviyesi gösteren SVR algoritması açısından değerlendirildiğinde, Kutu grafiklerinin gerçek ve tahmin skorları için temsil ettiği ortalama ve medyan değerlerinin neredeyse örtüştüğü ve dağılımların (kutu dışındaki uç örnekler için yeterli seviyelerde olmasa da) Q1-Q3 aralığı için oldukça yakın olduğunu görülmektedir. Şekil 46'deki bulgulara göre Tahmin edilen ve gerçek skorların genel dağılımlarının (KS p-değeri 0.707) ve merkezi eğilimlerinin (Wilcoxon p-değeri 0.924), istatistiksel olarak birbirlerinden farklı olmadığı belirlenmiştir. Bu durum, SVR algoritmasının TR9 verileri üzerinde en yüksek doğruluk ve güvenilirliği sunmakla birlikte hem bireysel gerçek skorları hem de gerçek skorlara ait dağılım yapısını doğru şekilde tahmin edebilme yeteneğini ortaya koymaktadır.



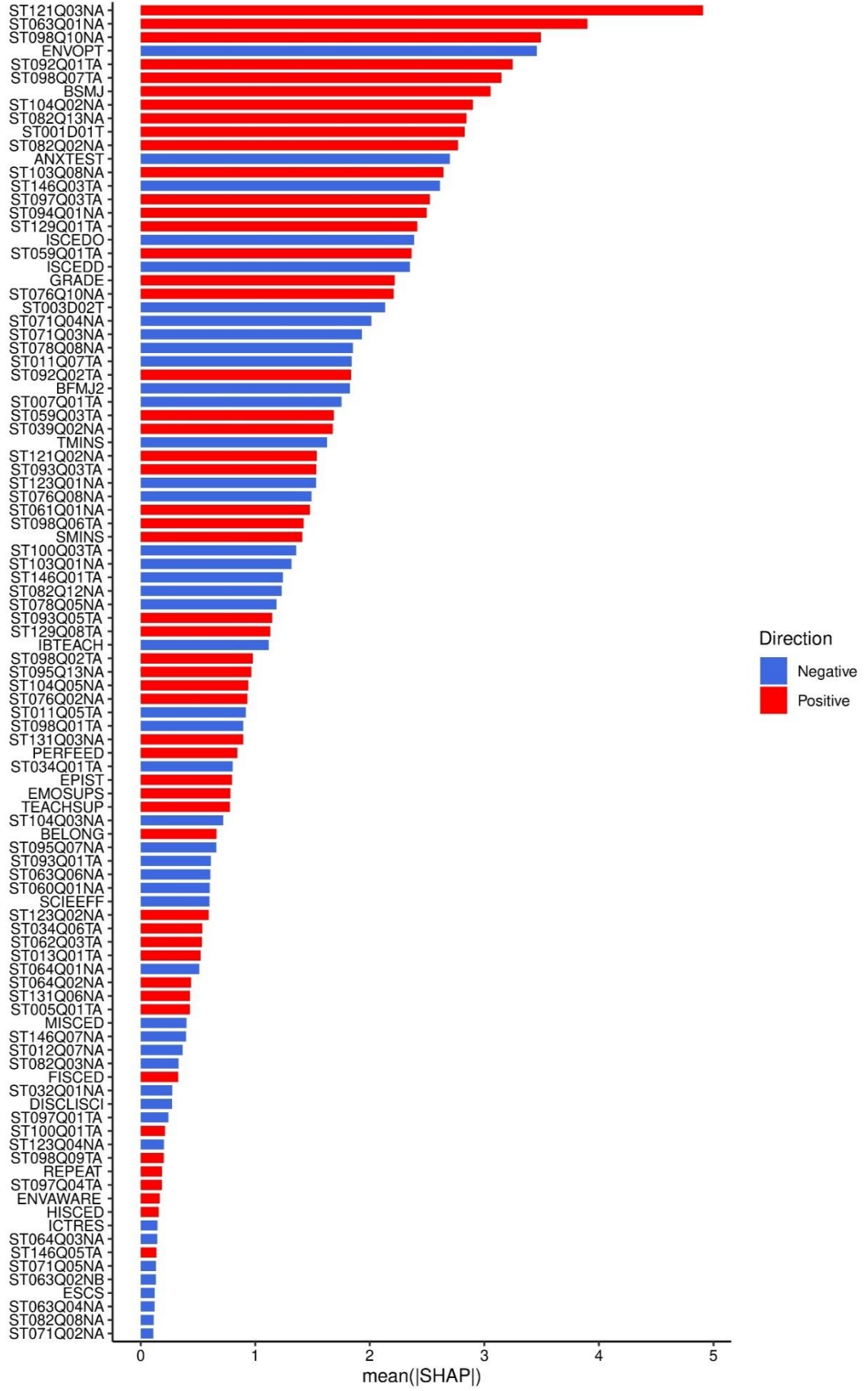
Şekil 46: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.

4.9.4 TR9 Bölgesi İçin Değişkenlerin Önem Seviyeleri

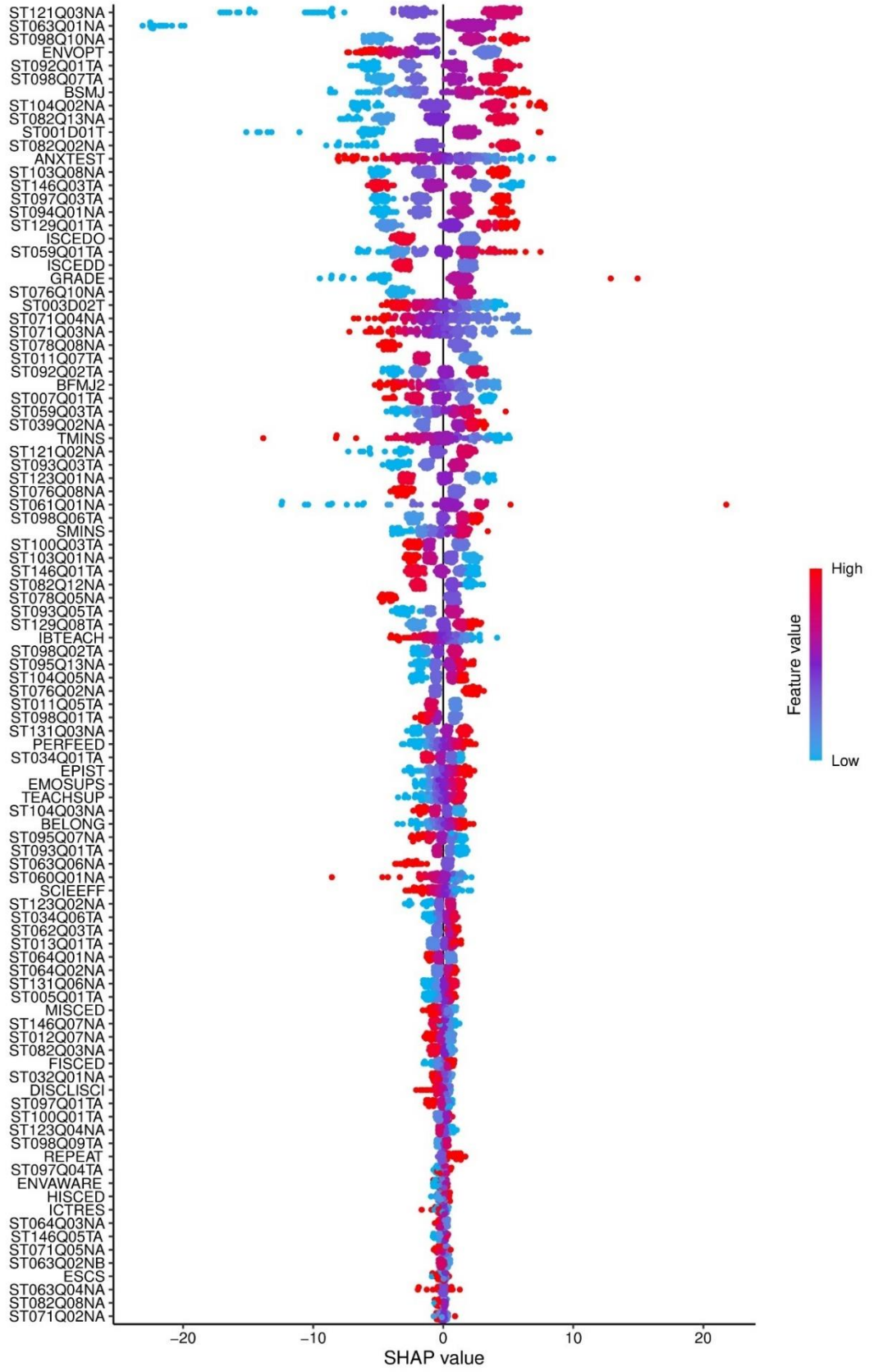
PISA 2015 Türkiye örnekleme TR9 bölgesi Fen okuryazarlık alanı için en iyi tahmin performans gösteren SVR algoritmasına göre SHAP (SHapley Additive exPlanations) değerleri kullanılarak değişkenlerin önem parametreleri belirlenmiştir. Sonuçlar her bir değişkenin ortalama SHAP değerlerinin mutlak değerine göre azalan sırada Tablo 31’de, sıralanmıştır. Bu sıralamaya göre diğer bölgelerdeki değişkenlerin karşılaştırması ise Kısım 4.13’te Jaccard indeksine göre yapılacaktır. Burada model için önemli olan sıralı değişkenler üç önem grubuna ayrılmış ve bu işlem model için kullanılan toplam değişkenleri 3 eşit parçaya ayırarak gerçekleştirilmiştir. Burada yüksek derecede önemli değişkenler 1. Grup (ilk 33 tanesi), orta derecede önemli değişkenler 2. Grup (34 - 66) ve düşük derecede önemli değişkenler 3. Grup (67 - 99) olarak etiketlenmiştir. Ayrıca Şekil 47 ‘ile Tablo 31 de özetlenen değişkenlerin ortalama SHAP değerleri görselleştirilmiştir. Bu şekil her bir tahmin edici değişkenin model tahminlerine olan ortalama etkisini ve yönünü (pozitif ve negatif katkıları) vurgulamaktadır. Şekil 48 her bir değişkenin test veri setindeki her bir gözlemdeki etkisini göstermektedir. Bu şekil içerisinde her bir nokta, bir veri noktasının bir özelliği için SHAP değerini temsil eder. Bu şeklin sağ tarafında pozitif SHAP değerleri, sol tarafında ise negatif SHAP değerleri bulunmaktadır. Tahmin edici değişkenlerin model çıktısına olan etkisi, bu SHAP değerlerinin büyüklüğüne göre belirlenir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Tablo31, Şekil 47 ve 48’teki bulgulara göre, TR9 bölgesi için model tahmininde önem seviyesi yüksek ilk 10 değişken, parantez içleri SHAP değerlerini ve model katkı yönlerini göstermek üzere, şu şekilde sıralanmıştır: ST121Q03NA (4.908, pozitif), ST063Q01NA (3.900, pozitif), ST098Q10NA (3.494, pozitif), ENVOPT (3.458, negatif), ST092Q01TA (3.248, pozitif), ST098Q07TA (3.148, pozitif), BSMJ (3.054, pozitif), ST104Q02NA (2.901, pozitif), ST082Q13NA (2.844, pozitif), ST001D01T (2.828, pozitif). Şekil 4 de Kırmızı ve mavi renkler, sırasıyla pozitif ve negatif yönlü etkileri temsil etmektedir. Bu, modelin tahminlerinin hangi özellikler tarafından ne kadar ve hangi yönde etkilendiğini daha net bir şekilde anlaşılmasına yardımcı olmaktadır. Buna göre negatif özellikteki değişkenlerin (ENVOPT) değerindeki artış öğrencilerin Fen başarı skorlarını düşürmekte buna karşı pozitif özellikteki değişkenlerin değerlerindeki artış ise öğrencilerin Fen başarı skorlarını arttırıcı yönde etki etmektedir.

Tablo 31: TR9 (Doğu Karadeniz) verileri için oluşturulan En yüksek performanslı algoritmaya (SVR) göre değişkenlerin önem seviyelerinin sıralanması.

ID	Değişken	SHAP değeri	Önem Grubu	ID	Değişken	SHAP değeri	Önem Grubu
1	ST121Q03NA	4.908	Grup 1	55	ST131Q03NA	0.895	Grup 2
2	ST063Q01NA	3.900	Grup 1	56	PERFEED	0.844	Grup 2
3	ST098Q10NA	3.494	Grup 1	57	ST034Q01TA	0.803	Grup 2
4	ENVOPT	3.458	Grup 1	58	EPIST	0.798	Grup 2
5	ST092Q01TA	3.248	Grup 1	59	EMOSUPS	0.783	Grup 2
6	ST098Q07TA	3.148	Grup 1	60	TEACHSUP	0.779	Grup 2
7	BSMJ	3.054	Grup 1	61	ST104Q03NA	0.721	Grup 2
8	ST104Q02NA	2.901	Grup 1	62	BELONG	0.662	Grup 2
9	ST082Q13NA	2.844	Grup 1	63	ST095Q07NA	0.661	Grup 2
10	ST001D01T	2.828	Grup 1	64	ST093Q01TA	0.614	Grup 2
11	ST082Q02NA	2.770	Grup 1	65	ST063Q06NA	0.608	Grup 2
12	ANXTEST	2.699	Grup 1	66	ST060Q01NA	0.603	Grup 2
13	ST103Q08NA	2.644	Grup 1	67	SCIEEFF	0.601	Grup 3
14	ST146Q03TA	2.612	Grup 1	68	ST123Q02NA	0.593	Grup 3
15	ST097Q03TA	2.524	Grup 1	69	ST034Q06TA	0.538	Grup 3
16	ST094Q01NA	2.497	Grup 1	70	ST062Q03TA	0.535	Grup 3
17	ST129Q01TA	2.414	Grup 1	71	ST013Q01TA	0.524	Grup 3
18	ISCEDO	2.387	Grup 1	72	ST064Q01NA	0.511	Grup 3
19	ST059Q01TA	2.365	Grup 1	73	ST064Q02NA	0.440	Grup 3
20	ISCEDD	2.350	Grup 1	74	ST131Q06NA	0.431	Grup 3
21	GRADE	2.218	Grup 1	75	ST005Q01TA	0.430	Grup 3
22	ST076Q10NA	2.208	Grup 1	76	MISCED	0.400	Grup 3
23	ST003D02T	2.135	Grup 1	77	ST146Q07NA	0.397	Grup 3
24	ST071Q04NA	2.014	Grup 1	78	ST012Q07NA	0.366	Grup 3
25	ST071Q03NA	1.930	Grup 1	79	ST082Q03NA	0.331	Grup 3
26	ST078Q08NA	1.853	Grup 1	80	FISCED	0.328	Grup 3
27	ST011Q07TA	1.842	Grup 1	81	ST032Q01NA	0.276	Grup 3
28	ST092Q02TA	1.836	Grup 1	82	DISCLISCI	0.273	Grup 3
29	BFMJ2	1.827	Grup 1	83	ST097Q01TA	0.242	Grup 3
30	ST007Q01TA	1.753	Grup 1	84	ST100Q01TA	0.213	Grup 3
31	ST059Q03TA	1.687	Grup 1	85	ST123Q04NA	0.203	Grup 3
32	ST039Q02NA	1.676	Grup 1	86	ST098Q09TA	0.201	Grup 3
33	TMIN	1.627	Grup 1	87	REPEAT	0.187	Grup 3
34	ST121Q02NA	1.537	Grup 2	88	ST097Q04TA	0.186	Grup 3
35	ST093Q03TA	1.534	Grup 2	89	ENVAWARE	0.166	Grup 3
36	ST123Q01NA	1.531	Grup 2	90	HISCED	0.158	Grup 3
37	ST076Q08NA	1.491	Grup 2	91	ICTRES	0.147	Grup 3
38	ST061Q01NA	1.478	Grup 2	92	ST064Q03NA	0.145	Grup 3
39	ST098Q06TA	1.422	Grup 2	93	ST146Q05TA	0.137	Grup 3
40	SMINS	1.410	Grup 2	94	ST071Q05NA	0.134	Grup 3
41	ST100Q03TA	1.356	Grup 2	95	ST063Q02NB	0.133	Grup 3
42	ST103Q01NA	1.317	Grup 2	96	ESCS	0.123	Grup 3
43	ST146Q01TA	1.242	Grup 2	97	ST063Q04NA	0.122	Grup 3
44	ST082Q12NA	1.230	Grup 2	98	ST082Q08NA	0.114	Grup 3
45	ST078Q05NA	1.186	Grup 2	99	ST071Q02NA	0.110	Grup 3
46	ST093Q05TA	1.147	Grup 2				
47	ST129Q08TA	1.132	Grup 2				
48	IBTEACH	1.117	Grup 2				
49	ST098Q02TA	0.980	Grup 2				
50	ST095Q13NA	0.966	Grup 2				
51	ST104Q05NA	0.938	Grup 2				
52	ST076Q02NA	0.931	Grup 2				
53	ST011Q05TA	0.918	Grup 2				
54	ST098Q01TA	0.895	Grup 2				



Şekil 47: TR9 bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.



Şekil 48: TR9 bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.

4.10 TRA Bölgesi PISA-2015 Bulguları

4.10.1 TRA Bölgesi Hakkında Temel Bilgiler

2015 PISA uygulamasının Türkiye örnekleminin onuncu bölgesi TRA olarak adlandırılmıştır. TRA bölgesi 2015 PISA uygulamasında ayrılan bölgelerden Kuzeydoğu Anadolu'yu temsil etmektedir. Kuzeydoğu Anadolu olarak ayrılan bölge Erzincan, Bayburt, Erzurum, Ardahan, Kars, Iğdır, Ağrı illerini kapsamaktadır. TRA bölgesinden (Kuzeydoğu Anadolu) 7. Sınıf ve üzeri düzeyde eğitim gören 199 öğrencinin 2015 PISA sınavına katıldığı görülmektedir. 2015 PISA sınavına katılan tüm Türkiye örnekleminin %3.38'inin Kuzeydoğu Anadolu bölgesinden katıldığını görmekteyiz.

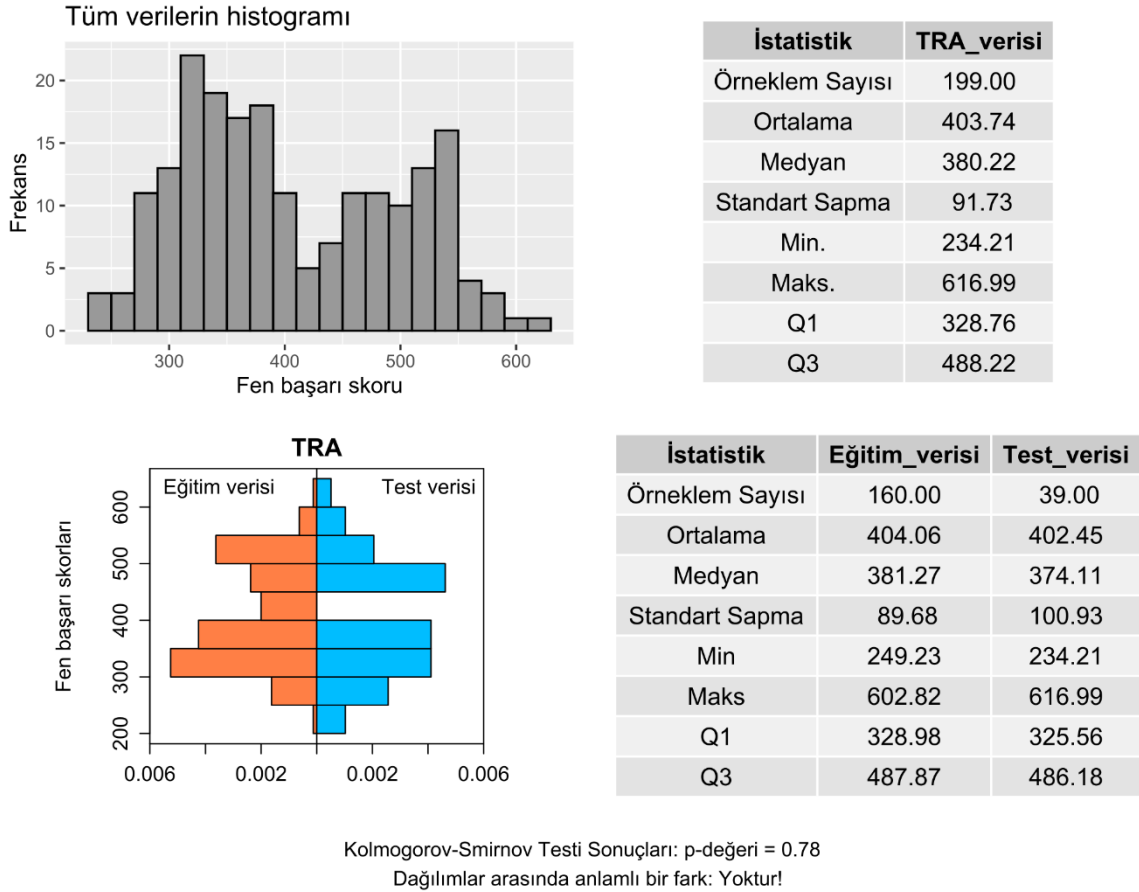
4.10.2 TRA Bölgesinin Veri Yapısı

Şekil 49'da Türkiye'nin TRA bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özeti sunulmaktadır. Şekil 49'un üst panelinde, tüm veri setini temsil eden bir histogram ve istatistiksel özet bilgileri gösterilmektedir. Aynı şeklin alt panelinde ise, makine öğrenmesi modelinin oluşturulması ve modelin performansının değerlendirilmesi süreçlerinde kullanılacak olan eğitim ve test alt veri setlerine ait istatistiksel bilgiler bulunmaktadır. Bu alt veri setleri ana veri setinden rastgele bir şekilde ve literatüre uygun olarak %80 (Eğitim verisi) / %20 (Test verisi) oranında örneklenerek oluşturulmuştur (Gholamy ve diğerleri, 2018).

Şekil 49'un üst panelindeki histograma göre, fen skorlarının ortalama puan etrafında merkezi bir eğilimle dağıldığı görülmektedir. Bu TRA bölgesindeki öğrencilerin (199 kişi) Fen başarı puanlarının normal dağılıma uyduğunu belirtmektedir. Ayrıca bu bölge için tüm veri setine ait ortalama puan 403.74, medyan 380.22, standart sapma 91.73 olarak hesaplanmıştır. Bunlara ek olarak minimum ve maksimum puanlar sırasıyla 234.21 ve 616.99 olup, birinci çeyrek (Q1) 328.76 ve üçüncü çeyrek (Q3) 488.22 olarak belirlenmiştir.

Şekil 49'un alt panelinde, eğitim veri setinin (160 kişi) ortalama puanı 404.06, medyanı 381.27 ve standart sapması 89.68 olarak tespit edilmiştir. Eğitim verisi için minimum ve maksimum puanlar sırasıyla 249.23 ve 602.82 olup, birinci çeyrek (Q1) 328.98 ve üçüncü çeyrek (Q3) 487.87 olarak belirlenmiştir. Bunların yanında Test veri setinin (39 kişi), ortalama puanı 402.45, medyanı 374.11 ve standart sapması 100.93 olarak tespit edilmiştir. Test verisi için minimum ve maksimum puanlar sırasıyla 234.21 ve 616.99 olup, birinci çeyrek (Q1) 325.56 ve üçüncü çeyrek (Q3) 486.18 olarak belirlenmiştir. Bu veri seti

için rasgele bir şekilde oluşturulan eğitim ve test alt veri setlerine ait dağılımlarının birbirlerinden anlamlı bir şekilde farklı olmadığı Kolmogorov-Smirnov testi sonucuna göre ($p > 0.05$) tespit edilmiştir.



Şekil 49: Türkiye'nin TRA bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)

4.10.3 TRA Bölgesinde Makine Öğrenme Modeli

Tablo 32, TRA (Kuzeydoğu Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı ve 10 tekrarlı çapraz doğrulama performans metriklerini ve optimize edilmiş Hiperparametre değerlerini göstermektedir. Bu tabloda elde edilen bulgular, XGBoost, RF, SVR ve BRNN algoritmaları ile Kısım 3.3'te oluşturulan makine öğrenme modellerin 10 tekrarlı ve 10 katlı çapraz doğrulama tahmin performansları elde edilmiştir. Bu tablonun ayrıntılı sonuçları ve kısaltmalar için gerekli açıklamalar EK-1

içerisinde sunulmuştur. İncelenen ve oluşturulan makine öğrenmesi modellerin başarı performansları RMSE, R^2 ve MAE metrikleri üzerinden değerlendirilmiş ve XGBoost modelinin en düşük RMSE (49.21), en yüksek R^2 (0.65) ve düşük MAE (40.14) değerleri ile eğitim verileri için en iyi performansı gösterdiği tespit edilmiştir.

Buna karşın SVR, BRNN ve RF modelleri sırasıyla 50.58, 51.59 ve 51.73 RMSE değerleri ile XGBoost algoritmasının ardından sıralanmışlardır. Bu veriler, TRA bölgesinin eğitim verileri için XGBoost modelinin Fen okuryazarlık puanlarının tahmininde diğer modellere göre daha iyi bir öğrenme ortaya koyduğunu göstermektedir.

Tablo 32: TRA (Kuzeydoğu Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri

Model	RMSE	R^2	MAE	Hiperparametre değerleri
XGBoost	49.21	0.65	40.14	nrounds 359.0000000 max_depth 8.0000000 eta 0.0760793 gamma 6.8846000 colsample_bytree 0.6445984 min_child_weight 10.0000000 subsample 0.8234931
SVR	50.58	0.63	41.15	sigma 0.005086933 C 2.026260064
BRNN	51.59	0.62	42.09	mtry "49" splitrule "extratrees" min.node.size "8"
RF	51.73	0.63	41.96	neurons 2

Tablo 33'te, TRA bölgesine ait verileri için oluşturulan makine öğrenme algoritmalarının test verisi performans göstergeleri detaylı bir şekilde sunulmaktadır. Bu tabloya göre, XGBoost modelinin RMSE değeri 48.30, R^2 değeri 0.78 ve MAE değeri 39.54 olarak belirlenmiş ve diğer modellerden daha üstün bir performans sergilediği tespit edilmiştir. Ayrıca, XGBoost modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.852, güven aralığının alt sınırı 0.755 ve üst sınırı 0.913 ile yüksek bir uyumu gösterdiği belirlenmiştir. Buna karşın ikinci en iyi performans gösteren SVR modelinde, RMSE değeri 52.14, R^2 değeri 0.74 ve MAE değeri 42.30 ile XGBoost'a göre daha düşük olup, uyum korelasyon katsayısı merkezi tahmini 0.824 (lwr.ci: 0.714- upr.ci: 0.895) olarak hesaplanmıştır. BRNN ve RF modellerinde ise sırasıyla RMSE değerleri 56.13 ve 56.61, R^2

değerleri 0.69 ve 0.71, MAE değerleri 47.27 ve 48.23 olarak belirlenmiştir. BRNN modelinin uyum korelasyon katsayısı merkezi tahmini 0.827 (lwr.ci: 0.697- upr.ci: 0.905) iken, RF modelinin uyum korelasyon katsayısı merkezi tahmini 0.779 (lwr.ci: 0.655- upr.ci: 0.863) olarak hesaplanmıştır. Bu sonuçlar, XGBoost modelinin en düşük hata oranları ve en yüksek açıklama gücü ile diğer modellerden daha üstün performans sergilediğini ortaya koymaktadır.

Tablo 33: TRA (Kuzey Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri

Model	RMSE	R^2	MAE	ρ_c
XGBoost	48.30	0.78	39.54	XGBoost.est 0.8521316 XGBoost.lwr.ci 0.7548216 XGBoost.upr.ci 0.9127403
SVR	52.14	0.74	42.30	SVR.est 0.8237772 SVR.lwr.ci 0.7116452 SVR.upr.ci 0.8949801
BRNN	56.13	0.69	47.27	BRNN.est 0.8273921 BRNN.lwr.ci 0.6970451 BRNN.upr.ci 0.9048030
RF	56.61	0.71	48.23	RF.est 0.7794147 RF.lwr.ci 0.6550807 RF.upr.ci 0.8626491

ρ_c = Uyum Korelasyon Katsayısı (Concordance Correlation Coefficient)

est = ρ_c için merkezi tahmin edilen değeri

lwr.ci = güven aralığının alt sınırını (lower confidence interval)

upr.ci = güven aralığının üst sınırını (upper confidence interval)

Tablo 32 ve Tablo 33 verilerine göre yapılan genel değerlendirmede, XGBoost modelinin hem eğitim hem de test verilerinde en yüksek performansı sergilediği görülmektedir. Bu durum genelleştirme kabiliyetinin güçlü olduğunu göstermektedir. Ayrıca test verilerinin uyum korelasyon katsayısının (ρ_c) yüksek olması, modelin tahminlerinin gerçek verilerle yüksek uyum gösterdiğini belirtmektedir. Bu, XGBoost modelinin hem eğitim hem de test verilerinde tutarlı ve güvenilir bir performans sergilediğinin kanıtıdır.

Diğer yandan SVR, BRNN ve RF modelleri ise eğitim ve test verileri arasında benzer hata oranlarına sahip olsalar da performansları XGBoost modellerine kıyasla daha düşük seviyelerde kalmıştır. Bu durum SVR, BRNN ve RF modellerinin genelleştirme kabiliyetinin sınırlı olduğunu ve yeni veriler üzerinde daha düşük performans sergilediğini göstermektedir. Her üç modelin de ρ_c değerleri, tahminlerinin gerçek verilerle makul bir uyum gösterdiğini belirtmektedir.

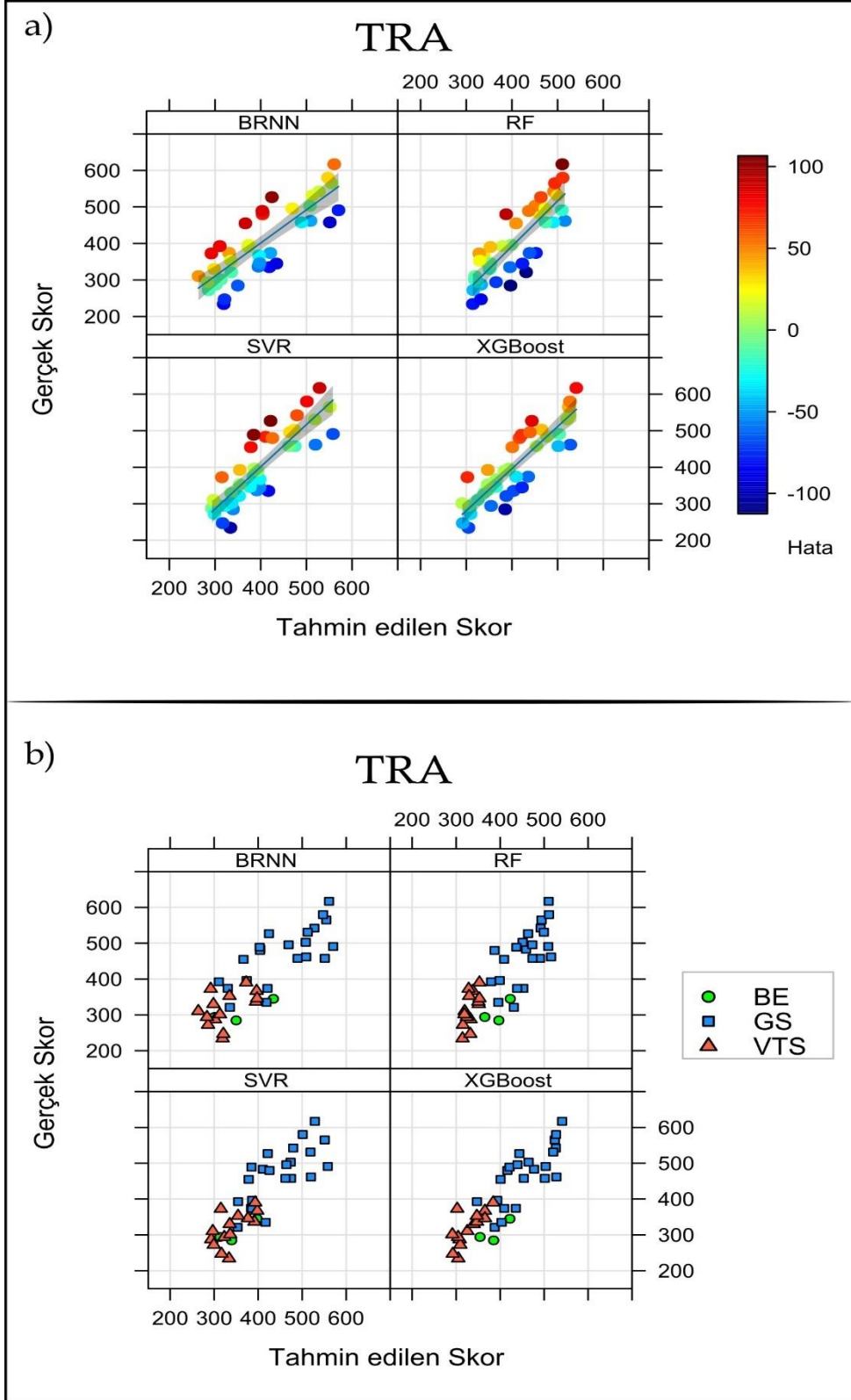
Şekil 50, Tablo 33'te elde edilen performans metriklerine ait çapraz doğrulama diyagramlarını tanımlamaktadır. Şekil 50a'da tüm modellerin hata dağılımlarının regresyon çizgisi boyunca ve yakınlarında dağıldığı görülmektedir. Bu durum tahmin hatalarının dağılımının normal yapıda olduğunu ve hata ortalamasının da sıfıra yakın olduğunu göstermektedir. Bu sonuçlar Tablo 33'te elde edilen bulgular ile uyumludur. Bunlarla birlikte XGBoost modelinin yüksek tahmin performansı sergilediği, Şekil 50a'daki çapraz doğrulama diyagramının hata dağılımlarının regresyon çizgisi etrafında sıkı şekilde toplanmasından dolayı daha açık bir şekilde görülmektedir.

Şekil 50b, aynı test verileri için modellerin farklı okul türlerine (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim) ait tahmin edilen ve gerçek skorlarının karşılaştırmasını göstermektedir. Burada özellikle tanımlayıcı regresyon çizgisinden sapan ve modellerin tahmin etme performansının düşmesine neden olan okul grubunun Mesleki ve Teknik Ortaöğretim (VTS) okul türü olduğu dikkat çekmektedir. Bununla birlikte XGBoost, her tüm okul türlerinde de tahmin edilen skorların gerçek skorlarla en uyumlu olduğu model olarak öne çıkmaktadır. Ayrıca TRA bölgesindeki Fen başarı skorları için genel ortaokul (GS) başarı düzeyinin VTS okul türüne göre belirgin şekilde artış göstermiş olması dikkat çekicidir. Özellikle okul türünün modellerin oluşturulmasında bir değişken olarak kullanılmamış olmasına rağmen, makine öğrenme algoritmasının bu başarı farklılıklarını diğer değişkenlerin yardımı ile tespit edebilmesi ve tahminlerini buna göre gerçekleştirmesi önemli bir başarı olduğu düşünülmektedir. Bu bulgular, XGBoost modelinin eğitim verileri üzerindeki güçlü performansını ve genelleştirme kabiliyetini destekler niteliktedir. XGBoost'un, farklı okul türlerindeki öğrencilerin performansını doğru bir şekilde tahmin edebilmesi, eğitim politikaları ve müdahalelerinin planlanmasında doğru tahmin skorları önerebileceğini ve bu konuda değerli bir araç olarak kullanılabileceğini göstermektedir. Özellikle Mesleki ve Teknik Ortaöğretim (VTS) öğrencileri için tahminlerdeki sapmalara rağmen, XGBoost'un genel olarak diğer modellere göre daha yüksek doğrulukla sonuçlar vermesi, bu modelin TRA bölgesi için geniş bir yelpazede güvenilir tahminler oluşturabileceğini göstermektedir.

Şekil 51, farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost) tahmin ettiği başarı skorları ile gerçek başarı skorlarının karşılaştırmalı kutu grafiğini sunmaktadır. Şekil 50 de sunulan çapraz doğrulama bulgularına göre oluşturulmuş bu grafikler, modellerin tahmin performansını ve gerçek verilerle uyumunu değerlendirmek için önemli

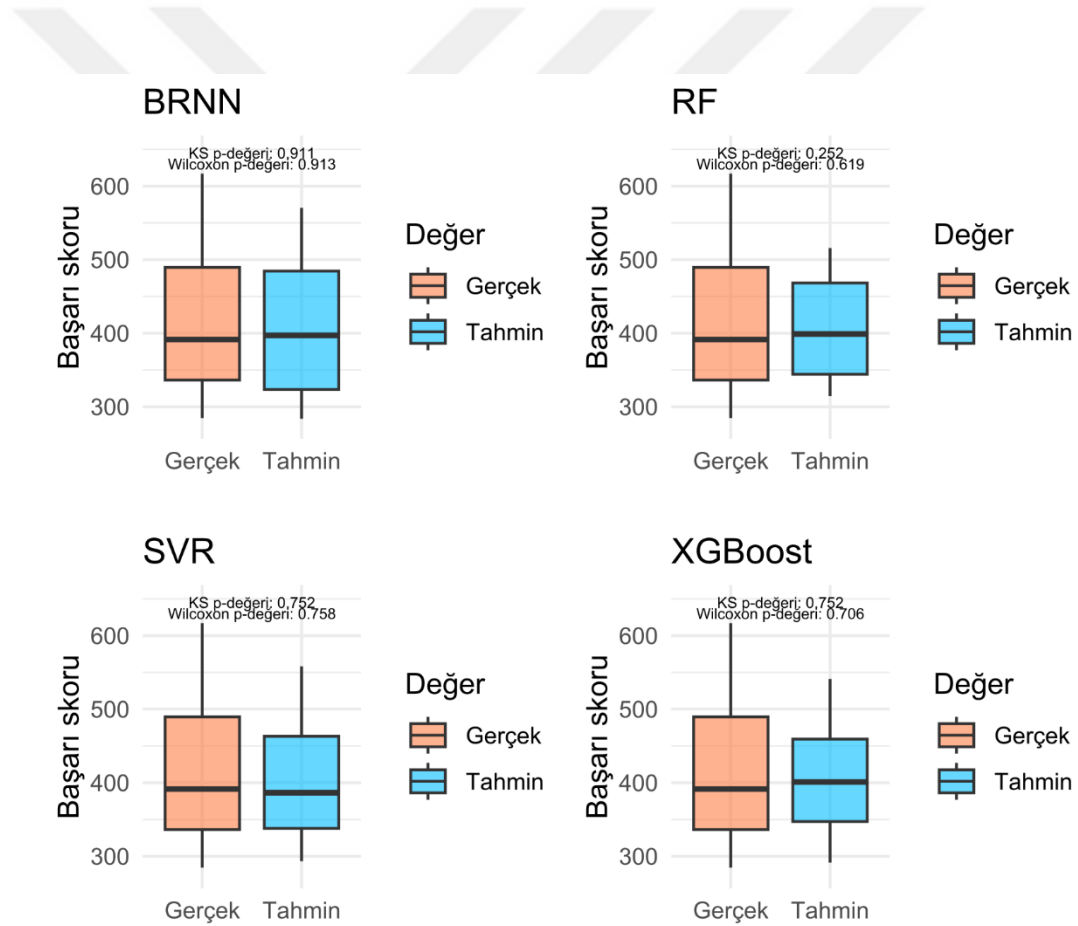
bir yakařım saęlamaktadır. Buna g re RF hari  t m modellerin tahmin skorlarının hem Kolmogorov -Simirnov testi (KS) hem de Wilcoxon sıralama toplamı testine g re ger ek skorların genel daęılımlarından ve merkezi eęilimlerinden anlamlı  l  de bir farklılıęı ($p > 0.05$) olmadığı tespit edilmiřtir. Bu durum makine  ęrenmesi modellerinin tahmin edilen ve ger ek bařarı skorları arasındaki uyumun y ksek olduęunu ortaya koymaktadır.





Şekil 50: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)

Şekil 51, en yüksek performans seviyesi gösteren XGBoost algoritması açısından değerlendirildiğinde, Kutu grafiklerinin gerçek ve tahmin skorları için temsil ettiği ortalama ve medyan değerlerinin neredeyse örtüştüğü ve dağılımların (kutu dışındaki uç örnekler için yeterli seviyelerde olmasa da) Q1-Q3 aralığı için oldukça yakın olduğunu görülmektedir. Şekil 51’deki bulgulara göre Tahmin edilen ve gerçek skorların genel dağılımlarının (KS p-değeri 0.752) ve merkezi eğilimlerinin (Wilcoxon p-değeri 0.706), istatistiksel olarak birbirlerinden farklı olmadığı belirlenmiştir. Bu durum, XGBoost algoritmasının TRA verileri üzerinde en yüksek doğruluk ve güvenilirliği sunmakla birlikte hem bireysel gerçek skorları hem de gerçek skorlara ait dağılım yapısını doğru şekilde tahmin edebilme yeteneğini ortaya koymaktadır.



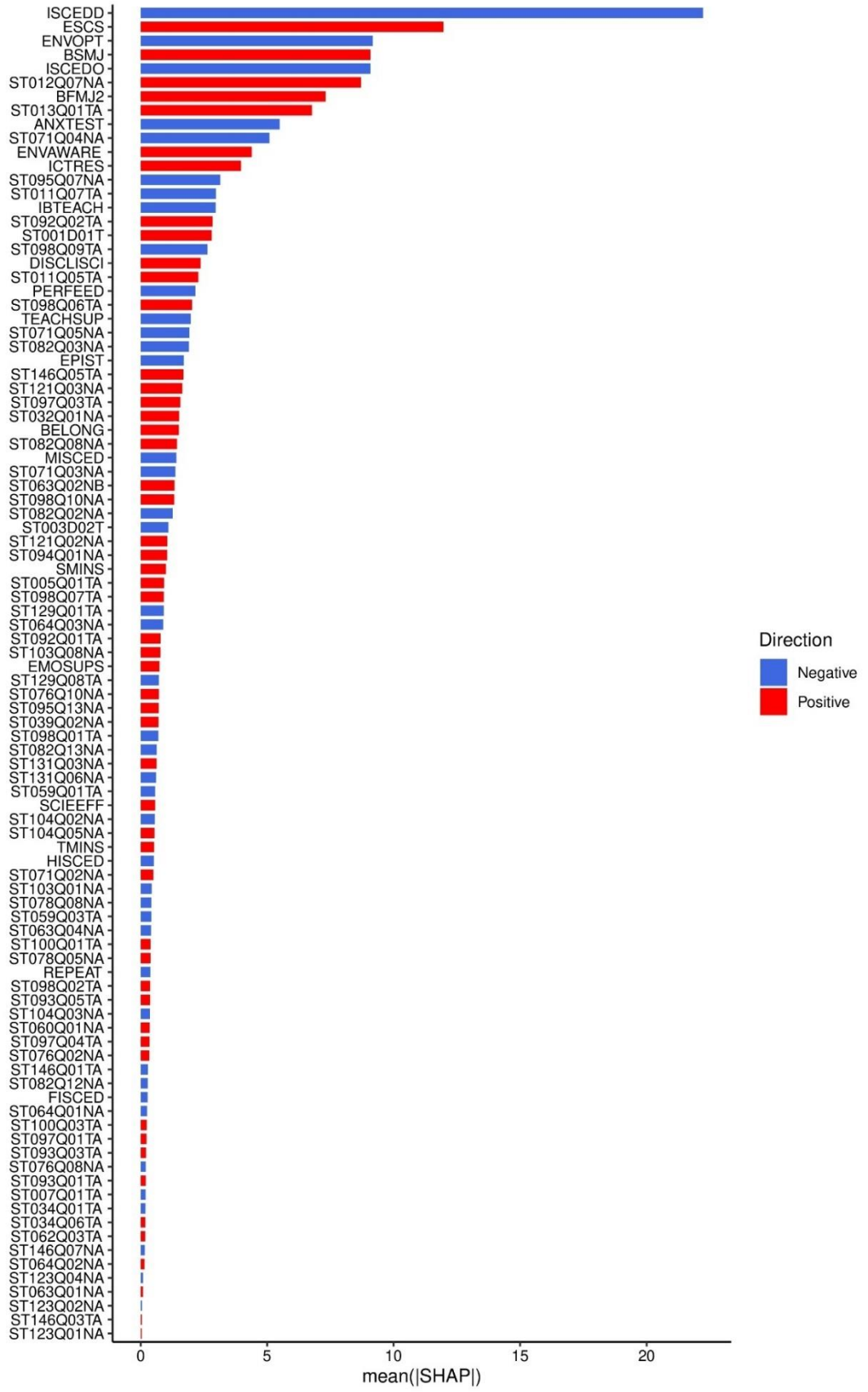
Şekil 51: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.

4.10.4 TRA Bölgesi İçin Değişkenlerin Önem Seviyeleri

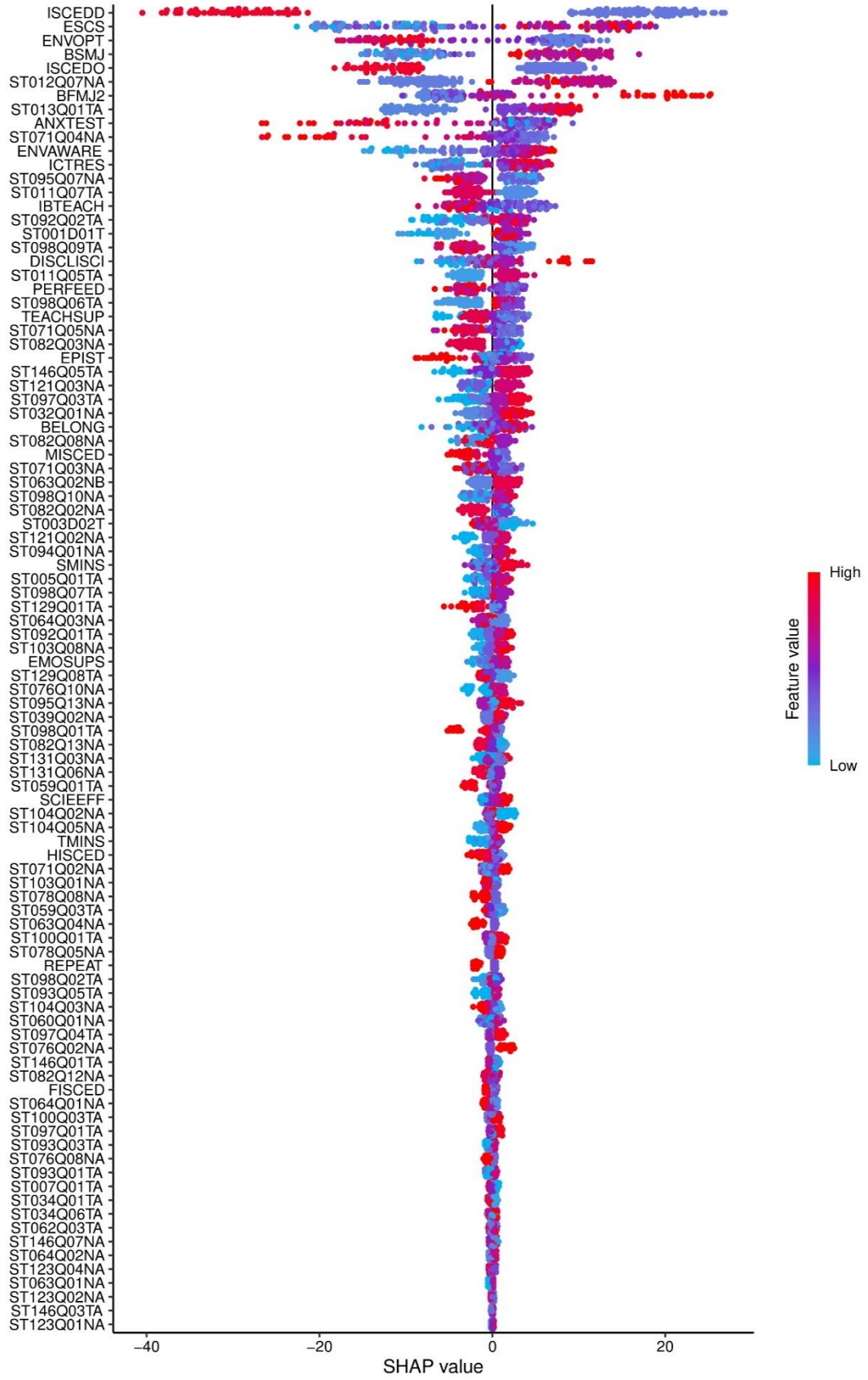
PISA 2015 Türkiye örnekleme TRA bölgesi Fen okuryazarlık alanı için en iyi tahmin performansı gösteren XGBoost algoritmasına göre SHAP (SHapley Additive exPlanations) değerleri kullanılarak değişkenlerin önem parametreleri belirlenmiştir. Sonuçlar her bir değişkenin ortalama SHAP değerlerinin mutlak değerine göre azalan sırada Tablo 34’te, sıralanmıştır. Bu sıralamaya göre diğer bölgelerdeki değişkenlerin karşılaştırması ise Kısım 4.13’te Jaccard indeksine göre yapılacaktır. Burada model için önemli olan sıralı değişkenler üç önem grubuna ayrılmış ve bu işlem model için kullanılan toplam değişkenleri 3 eşit parçaya ayırarak gerçekleştirilmiştir. Burada yüksek derecede önemli değişkenler 1. Grup (ilk 33 tanesi), orta derecede önemli değişkenler 2. Grup (34 - 66) ve düşük derecede önemli değişkenler 3. Grup (67 - 99) olarak etiketlenmiştir. Ayrıca Şekil 52 ile Tablo 34 de özetlenen değişkenlerin ortalama SHAP değerleri görselleştirilmiştir. Bu şekil her bir tahmin edici değişkenin model tahminlerine olan ortalama etkisini ve yönünü (pozitif ve negatif katkıları) vurgulamaktadır. Şekil 53 her bir değişkenin test veri setindeki her bir gözlemdeki etkisini göstermektedir. Bu şekil içerisinde her bir nokta, bir veri noktasının bir özelliği için SHAP değerini temsil eder. Bu şeklin sağ tarafında pozitif SHAP değerleri, sol tarafında ise negatif SHAP değerleri bulunmaktadır. Tahmin edici değişkenlerin model çıktısına olan etkisi, bu SHAP değerlerinin büyüklüğüne göre belirlenir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Tablo34, Şekil 52 ve Şekil 53’teki bulgulara göre, TRA bölgesi için model tahmininde önem seviyesi yüksek ilk 10 değişken, parantez içleri SHAP değerlerini ve model katkı yönlerini göstermek üzere, şu şekilde sıralanmıştır: ISCEDD (22.229, negatif), ESCS (11.973, pozitif), ENVOPT (9.175, negatif), BSMJ (9.088, pozitif), ISCEDO (9.085, negatif), ST012Q07NA (8.710, pozitif), BFMJ2 (7.316, pozitif), ST013Q01TA (6.773, pozitif), ANXTEST (5.497, negatif), ST071Q04NA (5.089, negatif). Şekil 4 de Kırmızı ve mavi renkler, sırasıyla pozitif ve negatif yönlü etkileri temsil etmektedir. Bu, modelin tahminlerinin hangi özellikler tarafından ne kadar ve hangi yönde etkilendiğini daha net bir şekilde anlaşılmasına yardımcı olmaktadır. Buna göre negatif özellikteki değişkenlerin (ANXTEST, ENVOPT, ISCEDD, ST064Q03NA, ST071Q04NA) değerlerindeki artış öğrencilerin Fen başarı skorlarını düşürmekte buna karşı pozitif özellikteki değişkenlerin değerlerindeki artış ise öğrencilerin Fen başarı skorlarını arttırıcı yönde etki etmektedir.

Tablo 34: TRA (Kuzeydoğu Anadolu) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.

ID	Değişken	SHAP değeri	Önem Grubu	ID	Değişken	SHAP değeri	Önem Grubu
1	ISCEDD	22.229	Grup 1	55	ST131Q03NA	0.630	Grup 2
2	ESCS	11.973	Grup 1	56	ST131Q06NA	0.611	Grup 2
3	ENVOPT	9.175	Grup 1	57	ST059Q01TA	0.569	Grup 2
4	BSMJ	9.088	Grup 1	58	SCIEEFF	0.569	Grup 2
5	ISCEDO	9.085	Grup 1	59	ST104Q02NA	0.562	Grup 2
6	ST012Q07NA	8.710	Grup 1	60	ST104Q05NA	0.547	Grup 2
7	BFMJ2	7.316	Grup 1	61	TMINS	0.530	Grup 2
8	ST013Q01TA	6.773	Grup 1	62	HISCED	0.523	Grup 2
9	ANXTEST	5.497	Grup 1	63	ST071Q02NA	0.499	Grup 2
10	ST071Q04NA	5.089	Grup 1	64	ST103Q01NA	0.437	Grup 2
11	ENVAWARE	4.392	Grup 1	65	ST078Q08NA	0.422	Grup 2
12	ICTRES	3.965	Grup 1	66	ST059Q03TA	0.421	Grup 2
13	ST095Q07NA	3.151	Grup 1	67	ST063Q04NA	0.410	Grup 3
14	ST011Q07TA	2.973	Grup 1	68	ST100Q01TA	0.393	Grup 3
15	IBTEACH	2.964	Grup 1	69	ST078Q05NA	0.390	Grup 3
16	ST092Q02TA	2.838	Grup 1	70	REPEAT	0.383	Grup 3
17	ST001D01T	2.808	Grup 1	71	ST098Q02TA	0.374	Grup 3
18	ST098Q09TA	2.637	Grup 1	72	ST093Q05TA	0.372	Grup 3
19	DISCLSCI	2.367	Grup 1	73	ST104Q03NA	0.364	Grup 3
20	ST011Q05TA	2.281	Grup 1	74	ST060Q01NA	0.357	Grup 3
21	PERFEED	2.168	Grup 1	75	ST097Q04TA	0.353	Grup 3
22	ST098Q06TA	2.035	Grup 1	76	ST076Q02NA	0.343	Grup 3
23	TEACHSUP	1.978	Grup 1	77	ST146Q01TA	0.281	Grup 3
24	ST071Q05NA	1.926	Grup 1	78	ST082Q12NA	0.278	Grup 3
25	ST082Q03NA	1.909	Grup 1	79	FISCED	0.277	Grup 3
26	EPIST	1.702	Grup 1	80	ST064Q01NA	0.253	Grup 3
27	ST146Q05TA	1.688	Grup 1	81	ST100Q03TA	0.240	Grup 3
28	ST121Q03NA	1.644	Grup 1	82	ST097Q01TA	0.233	Grup 3
29	ST097Q03TA	1.568	Grup 1	83	ST093Q03TA	0.212	Grup 3
30	ST032Q01NA	1.518	Grup 1	84	ST076Q08NA	0.199	Grup 3
31	BELONG	1.512	Grup 1	85	ST093Q01TA	0.199	Grup 3
32	ST082Q08NA	1.436	Grup 1	86	ST007Q01TA	0.196	Grup 3
33	MISCED	1.412	Grup 1	87	ST034Q01TA	0.193	Grup 3
34	ST071Q03NA	1.376	Grup 2	88	ST034Q06TA	0.189	Grup 3
35	ST063Q02NB	1.338	Grup 2	89	ST062Q03TA	0.182	Grup 3
36	ST098Q10NA	1.328	Grup 2	90	ST146Q07NA	0.164	Grup 3
37	ST082Q02NA	1.266	Grup 2	91	ST064Q02NA	0.151	Grup 3
38	ST003D02T	1.096	Grup 2	92	ST123Q04NA	0.096	Grup 3
39	ST121Q02NA	1.054	Grup 2	93	ST063Q01NA	0.092	Grup 3
40	ST094Q01NA	1.049	Grup 2	94	ST123Q02NA	0.055	Grup 3
41	SMINS	0.996	Grup 2	95	ST146Q03TA	0.042	Grup 3
42	ST005Q01TA	0.927	Grup 2	96	ST123Q01NA	0.040	Grup 3
43	ST098Q07TA	0.913	Grup 2	97	GRADE	0.000	Grup 3
44	ST129Q01TA	0.910	Grup 2	98	ST063Q06NA	0.000	Grup 3
45	ST064Q03NA	0.892	Grup 2	99	ST061Q01NA	0.000	Grup 3
46	ST092Q01TA	0.781	Grup 2				
47	ST103Q08NA	0.779	Grup 2				
48	EMOSUPS	0.742	Grup 2				
49	ST129Q08TA	0.724	Grup 2				
50	ST076Q10NA	0.723	Grup 2				
51	ST095Q13NA	0.714	Grup 2				
52	ST039Q02NA	0.701	Grup 2				
53	ST098Q01TA	0.697	Grup 2				
54	ST082Q13NA	0.639	Grup 2				



Şekil 52: TRA bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.



Şekil 53: TRA bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı

4.11 TRB Bölgesi PISA-2015 Bulguları

4.11.1 TRB Bölgesi Hakkında Temel Bilgiler

2015 PISA uygulamasının Türkiye örnekleminin on birinci bölgesi TRB olarak adlandırılmıştır. TRB bölgesi 2015 PISA uygulamasında ayrılan bölgelerden Ortadoğu Anadolu'yu temsil etmektedir. Ortadoğu Anadolu olarak ayrılan bölge Malatya, Tunceli, Elâzığ, Bingöl, Muş, Bitlis, Van, Hakkâri illerini kapsamaktadır. TRB bölgesinden (Ortadoğu Anadolu) 7. Sınıf ve üzeri düzeyde eğitim gören 276 öğrencinin 2015 PISA sınavına katıldığı görülmektedir. 2015 PISA sınavına katılan tüm Türkiye örnekleminin %4.68'inin Ortadoğu Anadolu bölgesinden katıldığını görmekteyiz.

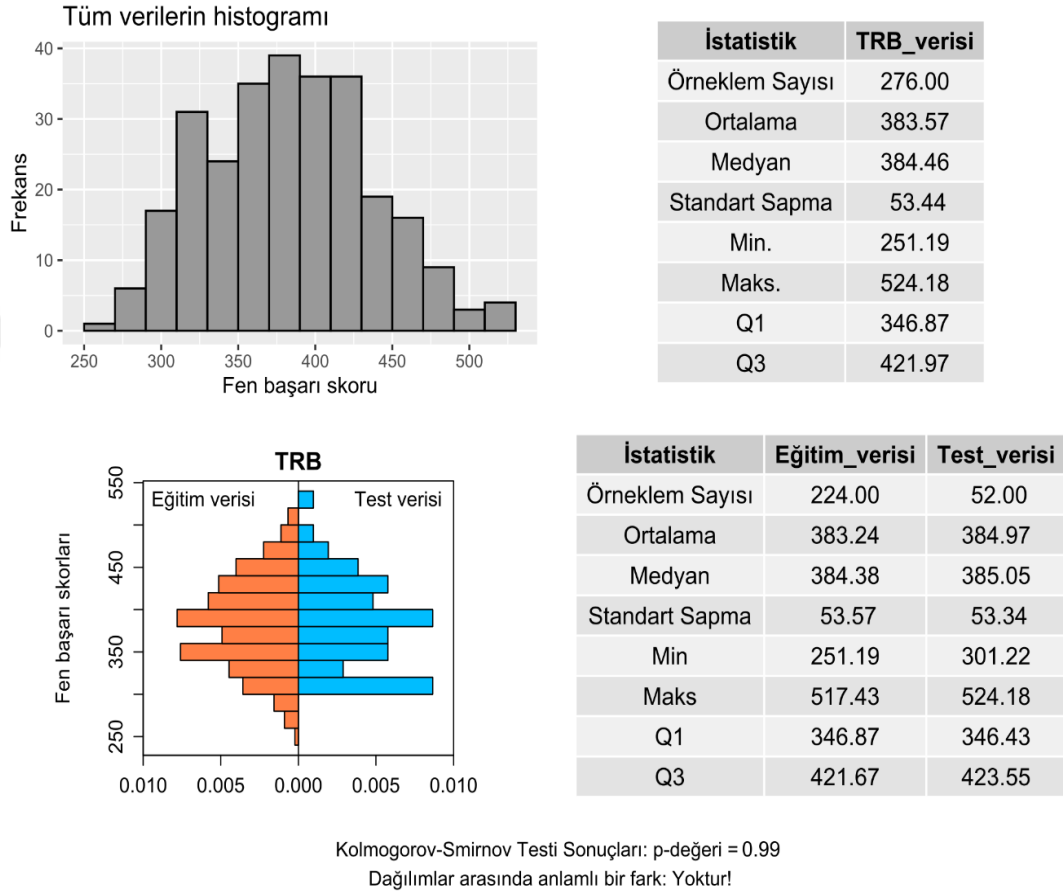
4.11.2 TRB Bölgesinin Veri Yapısı

Şekil 54'te, Türkiye'nin TRB bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı skorlarına ait istatistiksel özeti sunulmaktadır. Şekil 54'ün üst panelinde, tüm veri setini temsil eden bir histogram ve istatistiksel özet bilgileri gösterilmektedir. Aynı şeklin alt panelinde ise, makine öğrenmesi modelinin oluşturulması ve modelin performansının değerlendirilmesi süreçlerinde kullanılacak olan eğitim ve test alt veri setlerine ait istatistiksel bilgiler bulunmaktadır. Bu alt veri setleri ana veri setinden rastgele bir şekilde ve literatüre uygun olarak %80 (Eğitim verisi) / %20 (Test verisi) oranında örneklenerek oluşturulmuştur (Gholamy ve diğerleri, 2018).

Şekil 54'ün üst panelindeki histograma göre, fen puanlarının ortalama puan etrafında merkezi bir eğilimle dağıldığı görülmektedir. Bu TRB bölgesindeki öğrencilerin (276 kişi) Fen başarı skorlarının normal dağılıma uyduğunu belirtmektedir. Ayrıca bu bölge için tüm veri setine ait ortalama puan 383.57, medyan 384.46, standart sapma 53.44 olarak hesaplanmıştır. Bunlara ek olarak minimum ve maksimum puanlar sırasıyla 251.19 ve 524.18 olup, birinci çeyrek (Q1) 346.87 ve üçüncü çeyrek (Q3) 421.97 olarak belirlenmiştir.

Şekil 54'ün alt panelinde, Eğitim veri setinin (224 kişi) ortalama puanı 383.24, medyanı 384.38 ve standart sapması 53.57 olarak tespit edilmiştir. Eğitim verisi için minimum ve maksimum puanlar sırasıyla 251.19 ve 517.43 olup, birinci çeyrek (Q1) 346.87 ve üçüncü çeyrek (Q3) 421.67 olarak belirlenmiştir. Bunların yanında Test veri setinin (52 kişi), ortalama puanı 384.97, medyanı 385.05 ve standart sapması 53.34 olarak tespit edilmiştir. Test verisi için minimum ve maksimum puanlar sırasıyla 301.22 ve 524.18 olup, birinci çeyrek (Q1) 346.43 ve üçüncü çeyrek (Q3) 423.55 olarak belirlenmiştir. Bu veri seti

için rasgele bir şekilde oluşturulan eğitim ve test alt veri setlerine ait dağılımlarının birbirlerinden anlamlı bir şekilde farklı olmadığı Kolmogorov-Smirnov testi sonucuna göre ($p > 0.05$) tespit edilmiştir.



Şekil 54: Türkiye'nin TRB bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)

4.11.3 TRB Bölgesinde Makine Öğrenme Modeli

Tablo 35, TR (Ortadoğu Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı ve 10 tekrarlı çapraz doğrulama performans metriklerini ve optimize edilmiş hiperparametre değerlerini göstermektedir. Bu tabloda elde edilen bulgular, XGBoost, RF, SVR ve BRNN algoritmaları ile Kısım 3.3'te oluşturulan makine öğrenme modellerin 10 tekrarlı ve 10 katlı çapraz doğrulama tahmin performansları elde edilmiştir. Bu tablonun ayrıntılı sonuçları ve kısaltmalar için gerekli açıklamalar EK-1 içerisinde sunulmuştur. İncelenen ve oluşturulan makine öğrenmesi modellerin başarı

performansları RMSE, R^2 ve MAE metrikleri üzerinden değerlendirilmiş ve XGBoost modelinin en düşük RMSE (39.66), en yüksek R^2 (0.48) ve düşük MAE (32.48) değerleri ile eğitim verileri için en iyi performansı gösterdiği tespit edilmiştir. Buna karşın, BRNN ve SVR, RF modelleri sırasıyla 40.79, 41.75 ve 42.93 RMSE değerleri ile XGBoost algoritmasının ardından sıralanmışlardır. Bu veriler, TRB bölgesinin eğitim verileri için XGBoost modelinin Fen okuryazarlık puanlarının tahmininde diğer modellere göre daha iyi bir öğrenme ortaya koyduğunu göstermektedir.

Tablo 35: TRB (Ortadoğu Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri

Model	RMSE	R^2	MAE	Hiperparametre değerleri
XGBoost	39.66	0.48	32.48	nrounds 300.00000000 max_depth 10.00000000 eta 0.01801067 gamma 0.60228148 colsample_bytree 0.42126312 min_child_weight 16.00000000 subsample 0.66222751
BRNN	40.79	0.45	33.59	neurons 1
SVR	41.75	0.42	34.36	sigma 0.003852673 C 1.432272807
RF	42.93	0.43	34.59	mtry "32" splitrule "extratrees" min.node.size "3"

Tablo 36'da TRB bölgesine ait verileri için oluşturulan makine öğrenme algoritmalarının test verisi performans göstergeleri detaylı bir şekilde sunulmaktadır. Bu tabloya göre, SVR modelinin RMSE değeri 40.67, R^2 değeri 0.41 ve MAE değeri 32.17 olarak belirlenmiş ve diğer modellerden daha üstün bir performans sergilediği tespit edilmiştir. Ayrıca, SVR modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.592, güven aralığının alt sınırı 0.411 ve üst sınırı 0.728 ile yüksek bir uyumu gösterdiği belirlenmiştir. Buna karşın ikinci en iyi performans gösteren BRNN modelinde, RMSE değeri 42.65, R^2 değeri 0.38 ve MAE değeri 34.07 ile SVR modeline göre daha düşük olup, uyum korelasyon katsayısı merkezi tahmini 0.597 (lwr.ci: 0.401- upr.ci: 0.741) olarak hesaplanmıştır. RF ve XGBoost modellerinde ise sırasıyla RMSE değerleri 46.80 ve 49.23, R^2 değerleri 0.22 ve 0.18, MAE değerleri 37.16 ve 40.03 olarak belirlenmiştir. RF modelinin uyum korelasyon katsayısı merkezi tahmini 0.323 (lwr.ci: 0.156- upr.ci: 0.471) iken,

XGBoost modelinin uyum korelasyon katsayısı merkezi tahmini 0.375 (lwr.ci: 0.156- upr.ci: 0.559) olarak hesaplanmıştır. Bu sonuçlar, test verisinde SVR modelinin en düşük hata oranları ve en yüksek açıklama gücü ile diğer modellerden daha üstün performans sergilediğini ortaya koymaktadır.

Tablo 36: TRB (Ortadoğu Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri

Model	RMSE	R^2	MAE	ρ_c
SVR	40.67	0.41	32.17	SVR.est 0.5920260 SVR.lwr.ci 0.4108939 SVR.upr.ci 0.7281946
BRNN	42.65	0.38	34.07	BRNN.est 0.5974683 BRNN.lwr.ci 0.4006124 BRNN.upr.ci 0.7415990
RF	46.80	0.22	37.16	RF.est 0.3228765 RF.lwr.ci 0.1563042 RF.upr.ci 0.4715908
XGBoost	49.23	0.18	40.03	XGBoost.est 0.3754609 XGBoost.lwr.ci 0.1562611 XGBoost.upr.ci 0.5594129

ρ_c = Uyum Korelasyon Katsayısı (Concordance Correlation Coefficient)

est = ρ_c için merkezi tahmin edilen değeri

lwr.ci = güven aralığının alt sınırını (lower confidence interval)

upr.ci = güven aralığının üst sınırını (upper confidence interval)

Tablo 35 ve Tablo 36 verilerine göre yapılan genel değerlendirmede, XGBoost modelinin eğitim verilerinde en iyi performansı gösterdiğini fakat test verilerinde ezber (overfitting) yaptığı görülmektedir. Test verilerinde en yüksek performansı SVR modelinin sergilediği görülmektedir. Bu durum, SVR modelinin TRB bölgesi için genelleştirme kabiliyetinin güçlü olduğunu göstermektedir. Ayrıca test verilerinin uyum korelasyon katsayısının (ρ_c) yüksek olması, modelin tahminlerinin gerçek verilerle yüksek uyum gösterdiğini belirtmektedir. Bu, SVR modelinin hem eğitim hem de test verilerinde tutarlı ve güvenilir bir performans sergilediğinin kanıtıdır.

Diğer yandan, BRNN, RF ve XGBoost modelleri ise eğitim ve test verileri arasında benzer hata oranlarına sahip olsalar da performansları SVR modellerine kıyasla daha düşük seviyelerde kalmıştır. Bu durum BRNN, RF ve XGBoost modellerinin genelleştirme kabiliyetinin sınırlı olduğunu ve yeni veriler üzerinde daha düşük performans sergilediğini

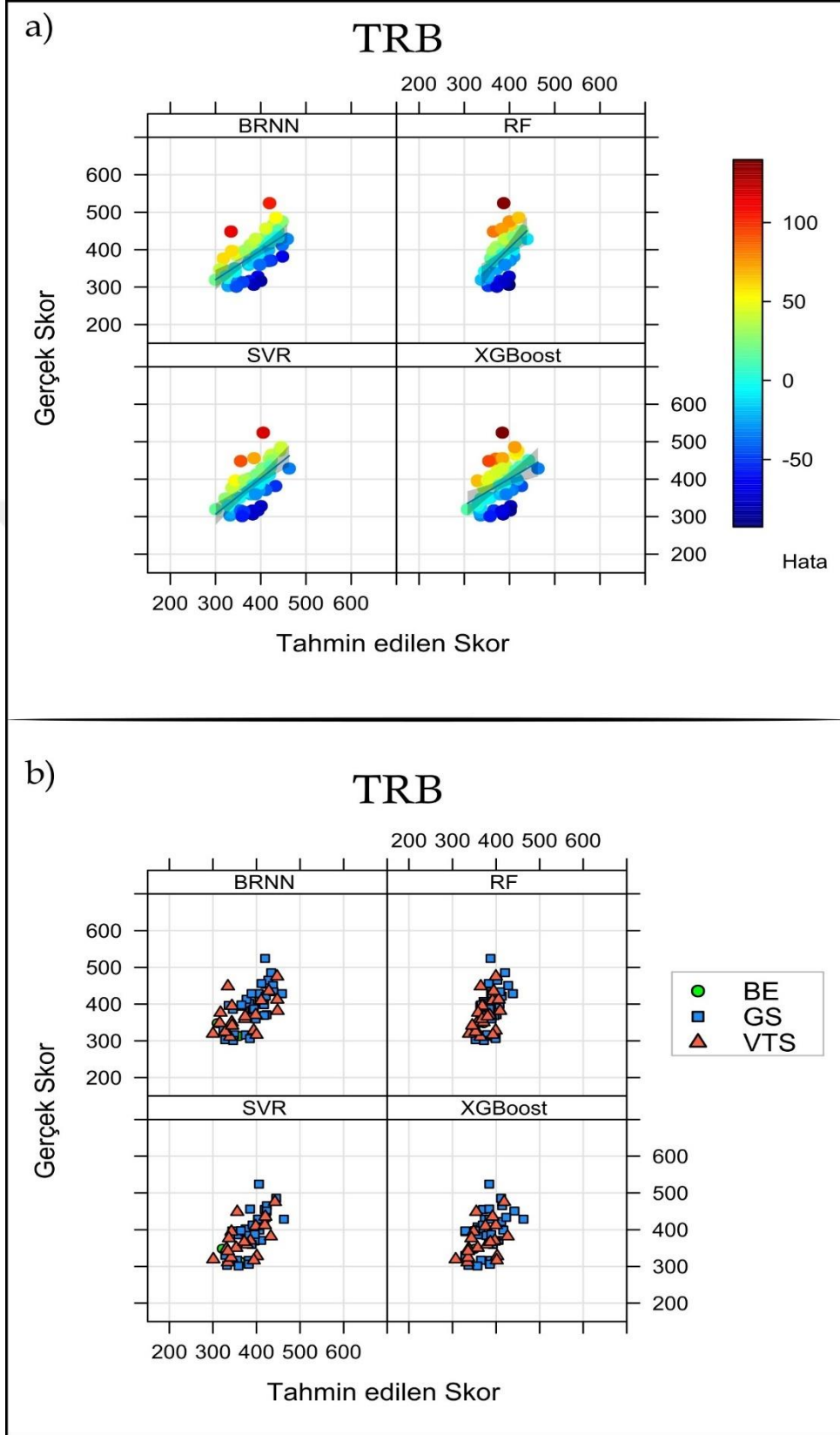
göstermektedir. Her üç modelin de ρ_c değerleri, tahminlerinin gerçek verilerle makul bir uyum gösterdiğini belirtmektedir.

Şekil 55, Tablo 36’da elde edilen performans metriklerine ait çapraz doğrulama diyagramlarını tanımlamaktadır. Şekil 55a’da tüm modellerin hata dağılımlarının regresyon çizgisi boyunca ve yakınlarında dağıldığı görülmektedir. Bu durum tahmin hatalarının dağılımının normal yapıda olduğunu ve hata ortalamasının da sıfıra yakın olduğunu göstermektedir. Bu sonuçlar Tablo 36 da elde edilen bulgular ile uyumludur. Bunlarla birlikte XGBoost modelinin yüksek tahmin performansı sergilediği, Şekil 55a’daki çapraz doğrulama diyagramının hata dağılımlarının regresyon çizgisi etrafında sıkı şekilde toplanmasından dolayı daha açık bir şekilde görülmektedir.

Şekil 55b, aynı test verileri için modellerin farklı okul türlerine (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim) ait tahmin edilen ve gerçek skorlarının karşılaştırmasını göstermektedir. Burada özellikle tanımlayıcı regresyon çizgisinden sapan ve modellerin tahmin etme performansının düşmesine neden olan okul grubunun Mesleki ve Teknik Ortaöğretim (VTS) okul türü olduğu dikkat çekmektedir. Bununla birlikte XGBoost, her tüm okul türlerinde de tahmin edilen skorların gerçek skorlarla en uyumlu olduğu model olarak öne çıkmaktadır. Ayrıca TRB bölgesindeki Fen başarı skorları için genel ortaokul (GS) başarı düzeyinin VTS okul türüne göre belirgin şekilde artış göstermiş olması dikkat çekicidir. Özellikle okul türünün modellerin oluşturulmasında bir değişken olarak kullanılmamış olmasına rağmen, makine öğrenme algoritmasının bu başarı farklılıklarını diğer değişkenlerin yardımı ile tespit edebilmesi ve tahminlerini buna göre gerçekleştirmesi önemli bir başarı olduğu düşünülmektedir. Bu bulgular, XGBoost modelinin eğitim verileri üzerindeki güçlü performansını ve genelleştirme kabiliyetini destekler niteliktedir. XGBoost’un, farklı okul türlerindeki öğrencilerin performansını doğru bir şekilde tahmin edebilmesi, eğitim politikaları ve müdahalelerinin planlanmasında doğru tahmin skorları önerebileceğini ve bu konuda değerli bir araç olarak kullanılabileceğini göstermektedir. Özellikle Mesleki ve Teknik Ortaöğretim (VTS) öğrencileri için tahminlerdeki sapmalara rağmen, XGBoost’un genel olarak diğer modellere göre daha yüksek doğrulukla sonuçlar vermesi, bu modelin TRB bölgesi için geniş bir yelpazede güvenilir tahminler oluşturabileceğini göstermektedir.

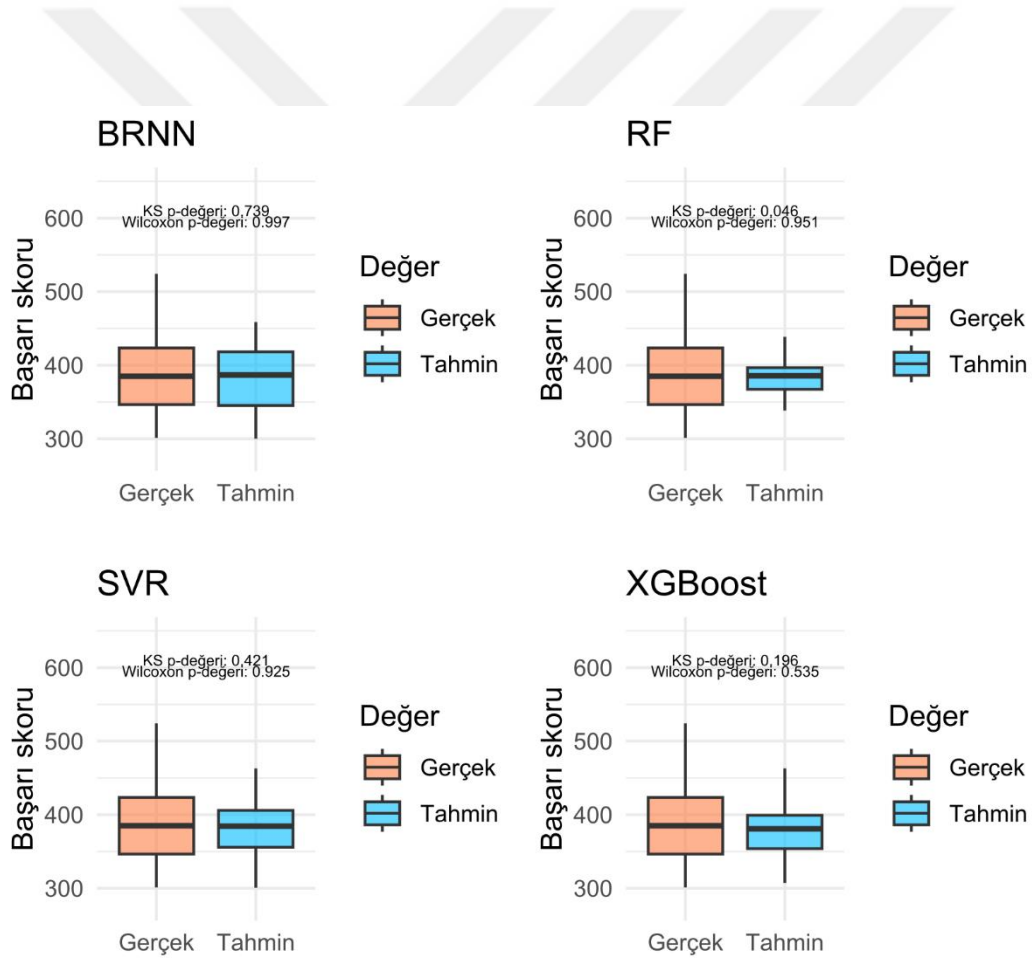
Şekil 56, farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost) tahmin ettiği başarı skorları ile gerçek başarı skorlarının karşılaştırmalı kutu grafiğini sunmaktadır. Şekil 55 de sunulan çapraz doğrulama bulgularına göre oluşturulmuş bu grafikler, modellerin tahmin performansını ve gerçek verilerle uyumunu değerlendirmek için önemli bir yaklaşım sağlamaktadır. Buna göre RF hariç tüm modellerin tahmin skorlarının hem Kolmogorov -Simirnov testi (KS) hem de Wilcoxon sıralama toplamı testine göre gerçek skorların genel dağılımlarından ve merkezi eğilimlerinden anlamlı ölçüde bir farklılığı ($p > 0.05$) olmadığı tespit edilmiştir. Bu durum makine öğrenmesi modellerinin tahmin edilen ve gerçek başarı skorları arasındaki uyumun yüksek olduğunu ortaya koymaktadır.





Şekil 55: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)

Şekil 56, en yüksek performans seviyesi gösteren SVR algoritması açısından değerlendirildiğinde, Kutu grafiklerinin gerçek ve tahmin skorları için temsil ettiği ortalama ve medyan değerlerinin neredeyse örtüştüğü ve dağılımların (kutu dışındaki uç örnekler için yeterli seviyelerde olmasa da) Q1-Q3 aralığı için oldukça yakın olduğunu görülmektedir. Şekil 56'daki bulgulara göre tahmin edilen ve gerçek skorların genel dağılımlarının (KS p-değeri 0.421) ve merkezi eğilimlerinin (Wilcoxon p-değeri 0.925), istatistiksel olarak birbirlerinden farklı olmadığı belirlenmiştir. Bu durum, SVR algoritmasının TRB verileri üzerinde en yüksek doğruluk ve güvenilirliği sunmakla birlikte hem bireysel gerçek skorları hem de gerçek skorlara ait dağılım yapısını doğru şekilde tahmin edebilme yeteneğini ortaya koymaktadır.



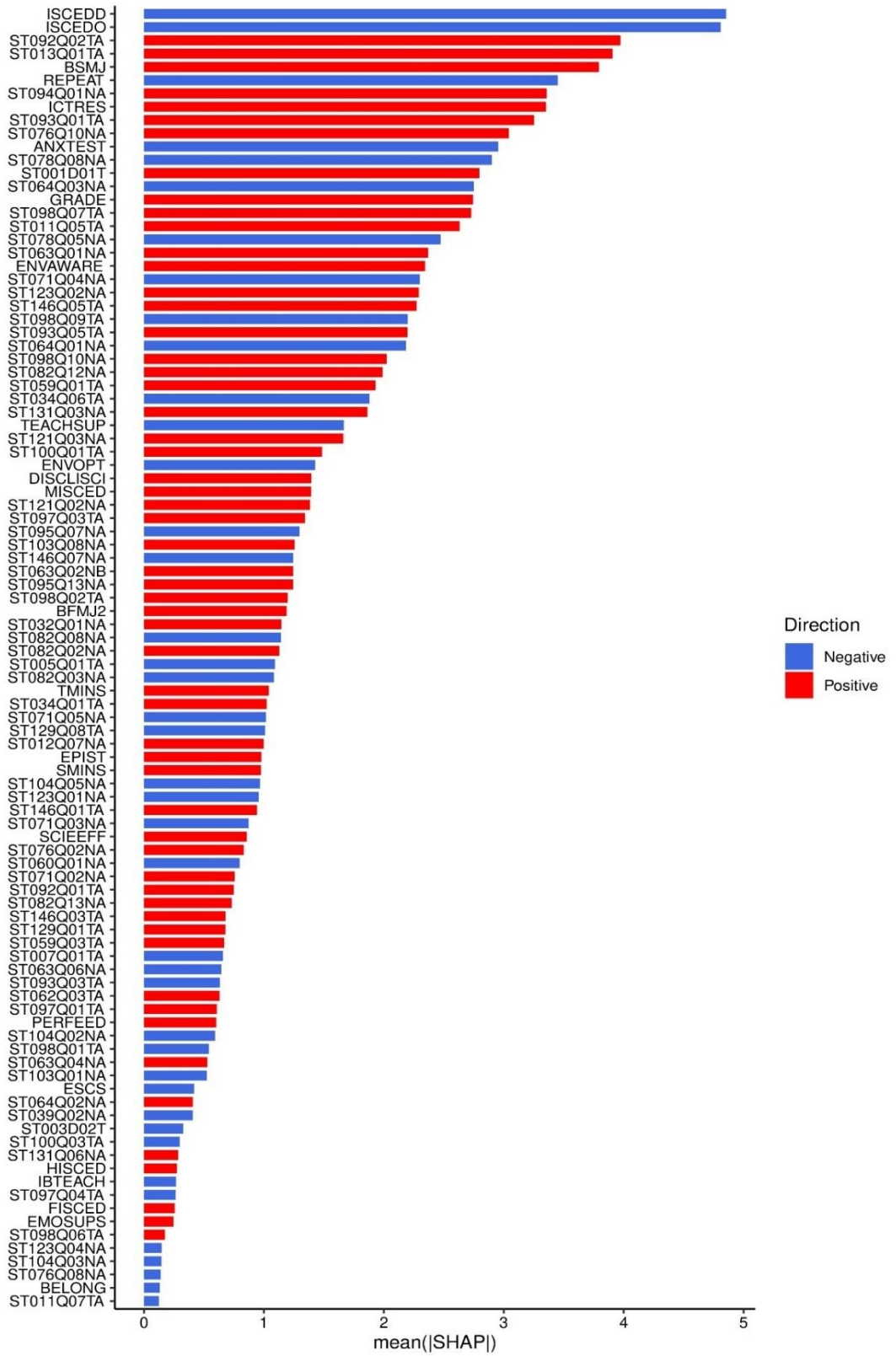
Şekil 56: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.

4.11.4 TRB Bölgesi İçin Değişkenlerin Önem Seviyeleri

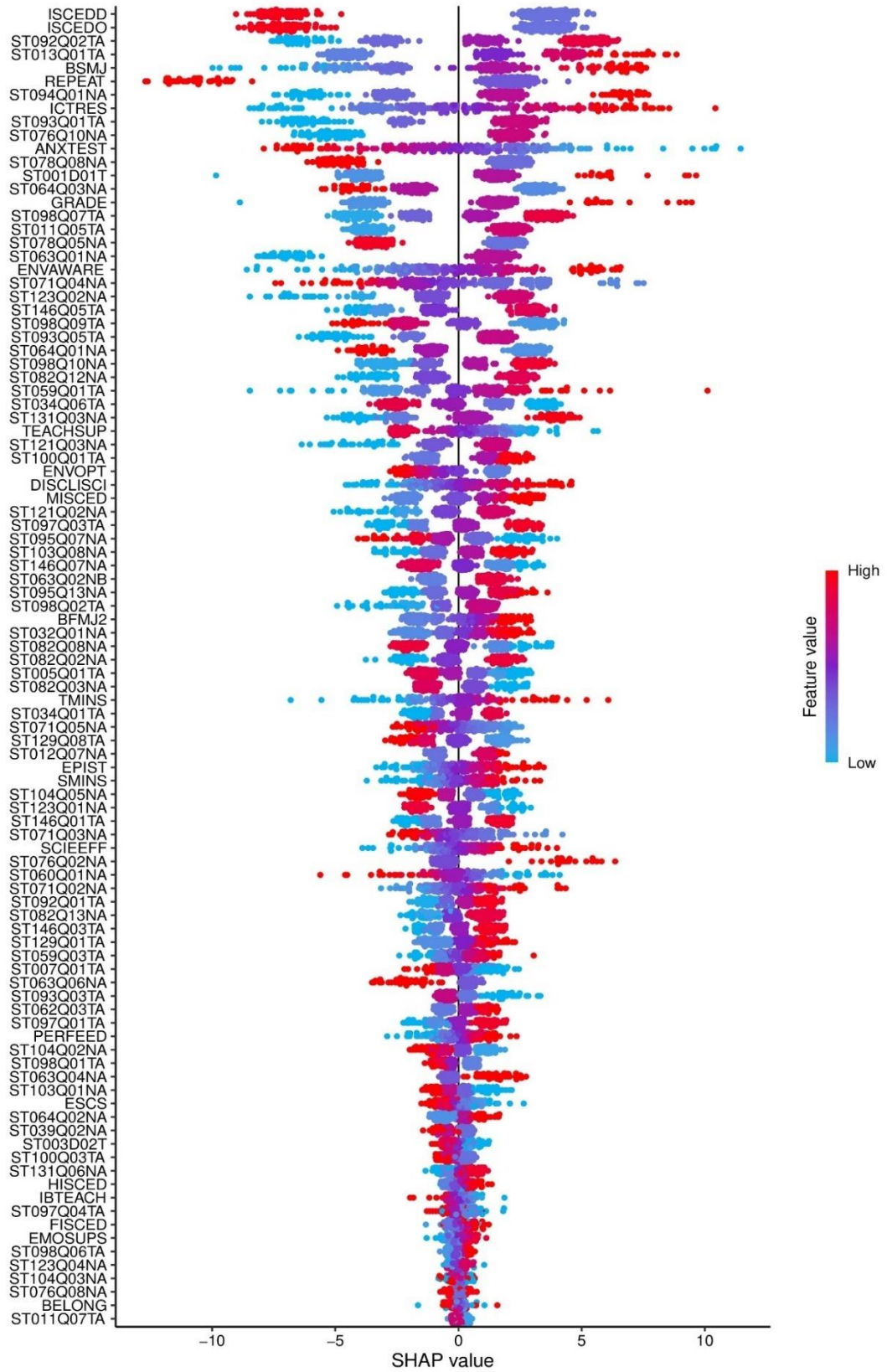
PISA 2015 Türkiye örnekleme TRB bölgesi Fen okuryazarlık alanı için iyi tahmin performans gösteren SVR algoritmasına göre SHAP (SHapley Additive exPlanations) değerleri kullanılarak) değişkenlerin önem parametreleri belirlenmiştir. Sonuçlar her bir değişkenin ortalama SHAP değerlerinin mutlak değerine göre azalan sırada Tablo 37’de, sıralanmıştır. Bu sıralamaya göre diğer bölgelerdeki değişkenlerin karşılaştırması ise Kısım 4.13’te Jaccard indeksine göre yapılacaktır. Burada model için önemli olan sıralı değişkenler üç önem grubuna ayrılmış ve bu işlem model için kullanılan toplam değişkenleri 3 eşit parçaya ayırarak gerçekleştirilmiştir. Burada yüksek derecede önemli değişkenler 1. Grup (ilk 33 tanesi), orta derecede önemli değişkenler 2. Grup (34- 66) ve düşük derecede önemli değişkenler 3. Grup (67- 99) olarak etiketlenmiştir. Ayrıca Şekil 57 ile Tablo 37 de özetlenen değişkenlerin ortalama SHAP değerleri görselleştirilmiştir. Bu şekil her bir tahmin edici değişkenin model tahminlerine olan ortalama etkisini ve yönünü (pozitif ve negatif katkıları) vurgulamaktadır. Şekil 58 her bir değişkenin test veri setindeki her bir gözlemdeki etkisini göstermektedir. Bu şekil içerisinde her bir nokta, bir veri noktasının bir özelliği için SHAP değerini temsil eder. Bu şeklin sağ tarafında pozitif SHAP değerleri, sol tarafında ise negatif SHAP değerleri bulunmaktadır. Tahmin edici değişkenlerin model çıktısına olan etkisi, bu SHAP değerlerinin büyüklüğüne göre belirlenir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Tablo37, Şekil 57 ve 58’deki bulgulara göre, TRB bölgesi için model tahmininde önem seviyesi yüksek ilk 10 değişken, parantez içleri SHAP değerlerini ve model katkı yönlerini göstermek üzere, şu şekilde sıralanmıştır: ISCEDD (4.853, negatif), ISCEDO (4.807, negatif), ST092Q02TA (3.972, pozitif), ST013Q01TA (3.907, pozitif), BSMJ (3.794, pozitif), REPEAT (3.450, negatif), ST094Q01NA (3.357, pozitif), ICTRES (3.352, pozitif), ST093Q01TA (3.253, pozitif), ST076Q10NA (3.042, pozitif). Şekil 4 de Kırmızı ve mavi renkler, sırasıyla pozitif ve negatif yönlü etkileri temsil etmektedir. Bu, modelin tahminlerinin hangi özellikler tarafından ne kadar ve hangi yönde etkilendiğini daha net bir şekilde anlaşılmasına yardımcı olmaktadır. Buna göre negatif özellikteki değişkenlerin (ISCEDD, ISCEDO, REPEAT) değerlerindeki artış öğrencilerin Fen başarı skorlarını düşürmekte buna karşı pozitif özellikteki değişkenlerin değerlerindeki artış ise öğrencilerin Fen başarı skorlarını artırıcı yönde etki etmektedir.

Tablo 37: TRB (Ortadoğu Anadolu) verileri için oluşturulan En yüksek performanslı algoritmaya (SVR) göre değişkenlerin önem seviyelerinin sıralanması.

ID	Değişken	SHAP değeri	Önem Grubu	ID	Değişken	SHAP değeri	Önem Grubu
1	ISCEDD	4.853	Grup 1	55	ST129Q08TA	1.010	Grup 2
2	ISCEDO	4.807	Grup 1	56	ST012Q07NA	0.998	Grup 2
3	ST092Q02TA	3.972	Grup 1	57	EPIST	0.978	Grup 2
4	ST013Q01TA	3.907	Grup 1	58	SMINS	0.976	Grup 2
5	BSMJ	3.794	Grup 1	59	ST104Q05NA	0.968	Grup 2
6	REPEAT	3.450	Grup 1	60	ST123Q01NA	0.957	Grup 2
7	ST094Q01NA	3.357	Grup 1	61	ST146Q01TA	0.942	Grup 2
8	ICTRES	3.352	Grup 1	62	ST071Q03NA	0.872	Grup 2
9	ST093Q01TA	3.253	Grup 1	63	SCIEEFF	0.856	Grup 2
10	ST076Q10NA	3.042	Grup 1	64	ST076Q02NA	0.833	Grup 2
11	ANXTEST	2.954	Grup 1	65	ST060Q01NA	0.798	Grup 2
12	ST078Q08NA	2.900	Grup 1	66	ST071Q02NA	0.756	Grup 2
13	ST001D01T	2.796	Grup 1	67	ST092Q01TA	0.748	Grup 3
14	ST064Q03NA	2.751	Grup 1	68	ST082Q13NA	0.733	Grup 3
15	GRADE	2.743	Grup 1	69	ST146Q03TA	0.681	Grup 3
16	ST098Q07TA	2.728	Grup 1	70	ST129Q01TA	0.680	Grup 3
17	ST011Q05TA	2.632	Grup 1	71	ST059Q03TA	0.669	Grup 3
18	ST078Q05NA	2.473	Grup 1	72	ST007Q01TA	0.659	Grup 3
19	ST063Q01NA	2.369	Grup 1	73	ST063Q06NA	0.645	Grup 3
20	ENVAWARE	2.344	Grup 1	74	ST093Q03TA	0.634	Grup 3
21	ST071Q04NA	2.300	Grup 1	75	ST062Q03TA	0.631	Grup 3
22	ST123Q02NA	2.291	Grup 1	76	ST097Q01TA	0.608	Grup 3
23	ST146Q05TA	2.274	Grup 1	77	PERFEED	0.602	Grup 3
24	ST098Q09TA	2.199	Grup 1	78	ST104Q02NA	0.592	Grup 3
25	ST093Q05TA	2.196	Grup 1	79	ST098Q01TA	0.541	Grup 3
26	ST064Q01NA	2.185	Grup 1	80	ST063Q04NA	0.528	Grup 3
27	ST098Q10NA	2.023	Grup 1	81	ST103Q01NA	0.524	Grup 3
28	ST082Q12NA	1.989	Grup 1	82	ESCS	0.419	Grup 3
29	ST059Q01TA	1.931	Grup 1	83	ST064Q02NA	0.407	Grup 3
30	ST034Q06TA	1.880	Grup 1	84	ST039Q02NA	0.407	Grup 3
31	ST131Q03NA	1.862	Grup 1	85	ST003D02T	0.328	Grup 3
32	TEACHSUP	1.666	Grup 1	86	ST100Q03TA	0.297	Grup 3
33	ST121Q03NA	1.662	Grup 1	87	ST131Q06NA	0.285	Grup 3
34	ST100Q01TA	1.485	Grup 2	88	HISCED	0.274	Grup 3
35	ENVOPT	1.427	Grup 2	89	IBTEACH	0.267	Grup 3
36	DISCLISCI	1.395	Grup 2	90	ST097Q04TA	0.265	Grup 3
37	MISCED	1.393	Grup 2	91	FISCED	0.257	Grup 3
38	ST121Q02NA	1.385	Grup 2	92	EMOSUPS	0.246	Grup 3
39	ST097Q03TA	1.341	Grup 2	93	ST098Q06TA	0.175	Grup 3
40	ST095Q07NA	1.296	Grup 2	94	ST123Q04NA	0.148	Grup 3
41	ST103Q08NA	1.256	Grup 2	95	ST104Q03NA	0.146	Grup 3
42	ST146Q07NA	1.244	Grup 2	96	ST076Q08NA	0.141	Grup 3
43	ST063Q02NB	1.243	Grup 2	97	BELONG	0.133	Grup 3
44	ST095Q13NA	1.243	Grup 2	98	ST011Q07TA	0.124	Grup 3
45	ST098Q02TA	1.198	Grup 2	99	ST061Q01NA	0.000	Grup 3
46	BFMJ2	1.190	Grup 2				
47	ST032Q01NA	1.146	Grup 2				
48	ST082Q08NA	1.142	Grup 2				
49	ST082Q02NA	1.130	Grup 2				
50	ST005Q01TA	1.093	Grup 2				
51	ST082Q03NA	1.085	Grup 2				
52	TMINS	1.041	Grup 2				
53	ST034Q01TA	1.023	Grup 2				
54	ST071Q05NA	1.017	Grup 2				



Şekil 57: TRB bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.



Şekil 58: TRB bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.

4.12 TRC Bölgesi PISA-2015 Bulguları

4.12.1 TRC Bölgesi Hakkında Temel Bilgiler

2015 PISA uygulamasında Türkiye örnekleminin on ikinci bölgesi TRC olarak adlandırılmıştır. TRC bölgesi 2015 PISA uygulamasında ayrılan bölgelerden biri olan Güneydoğu Anadolu bölgesidir. Güneydoğu Anadolu Bölgesi Kilis, Gaziantep, Adıyaman, Şanlıurfa, Diyarbakır, Mardin, Siirt ve Şırnak illerini kapsamaktadır. Güneydoğu Anadolu bölgesinden 7.sınıf ve üzeri düzeyde eğitim gören 687 öğrencinin 2015 PISA sınavına katıldığı görülmektedir. 2015 PISA sınavına katılan tüm Türkiye örnekleminin %11.65'inin Güneydoğu Anadolu bölgesinden katıldığını görmekteyiz.

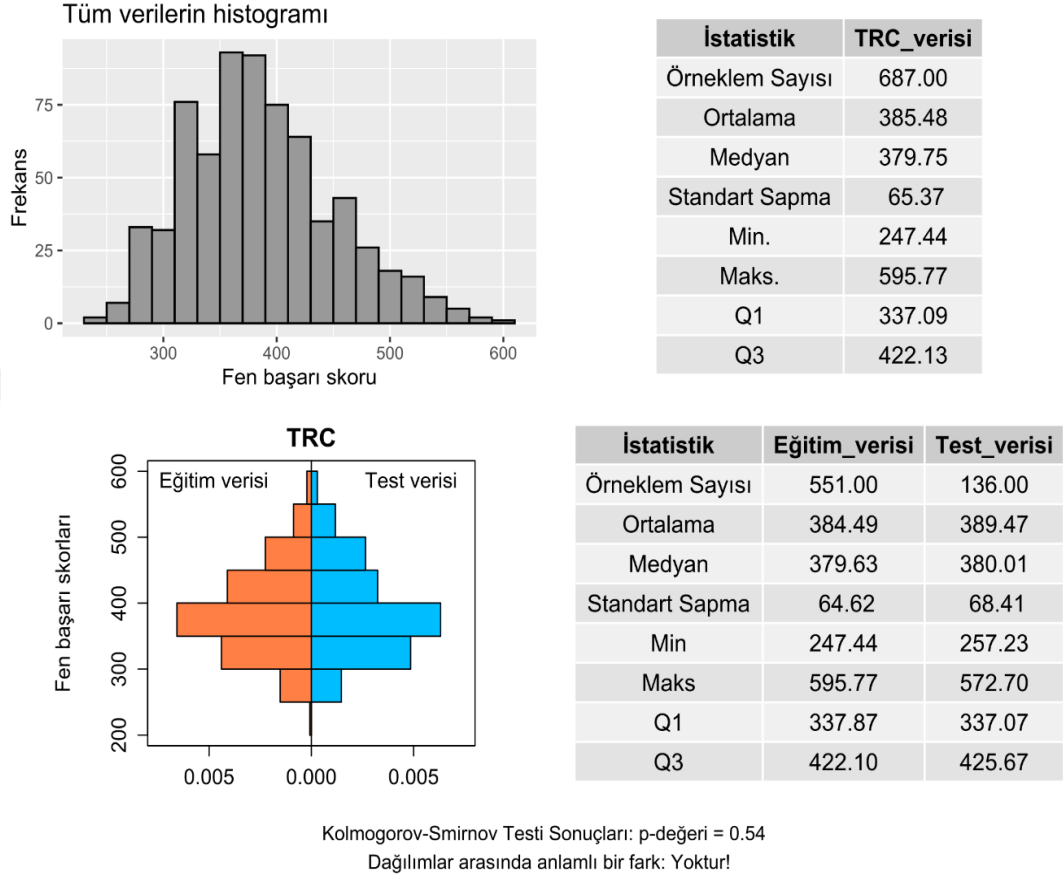
4.12.2 TRC Bölgesinin Veri Yapısı

Şekil 59'da, Türkiye'nin TRC bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özeti sunulmaktadır. Şekil 59'un üst panelinde, tüm veri setini temsil eden bir histogram ve istatistiksel özet bilgileri gösterilmektedir. Aynı şeklin alt panelinde ise, makine öğrenmesi modelinin oluşturulması ve modelin performansının değerlendirilmesi süreçlerinde kullanılacak olan eğitim ve test alt veri setlerine ait istatistiksel bilgiler bulunmaktadır. Bu alt veri setleri ana veri setinden rastgele bir şekilde ve literatüre uygun olarak %80 (Eğitim verisi) / %20 (Test verisi) oranında örneklenerek oluşturulmuştur (Gholamy ve diğerleri, 2018).

Şekil 59'un üst panelindeki histograma göre, fen puanlarının ortalama puan etrafında merkezi bir eğilimle dağıldığı görülmektedir. Bu TRB bölgesindeki öğrencilerin (687 kişi) Fen başarı puanlarının normal dağılıma uyduğunu belirtmektedir. Ayrıca bu bölge için tüm veri setine ait ortalama puan 385.48, medyan 379.75, standart sapma 65.37 olarak hesaplanmıştır. Bunlara ek olarak minimum ve maksimum puanlar sırasıyla 247.44 ve 595.77 olup, birinci çeyrek (Q1) 337.09 ve üçüncü çeyrek (Q3) 422.13 olarak belirlenmiştir.

Şekil 59'un alt panelinde, eğitim veri setinin (551 kişi) ortalama puanı 384.49, medyanı 379.63 ve standart sapması 64.62 olarak tespit edilmiştir. Eğitim verisi için minimum ve maksimum puanlar sırasıyla 247.44 ve 595.77 olup, birinci çeyrek (Q1) 337.87 ve üçüncü çeyrek (Q3) 422.10 olarak belirlenmiştir. Bunların yanında Test veri setinin (136 kişi), ortalama puanı 389.47, medyanı 380.01 ve standart sapması 68.41 olarak tespit edilmiştir. Test verisi için minimum ve maksimum puanlar sırasıyla 257.23 ve 572.70 olup, birinci çeyrek (Q1) 337.07 ve üçüncü çeyrek (Q3) 425.67 olarak belirlenmiştir. Bu veri seti

için rasgele bir şekilde oluşturulan eğitim ve test alt veri setlerine ait dağılımlarının birbirlerinden anlamlı bir şekilde farklı olmadığı Kolmogorov-Smirnov testi sonucuna göre ($p > 0.05$) tespit edilmiştir.



Şekil 59: Türkiye'nin TRC bölgesinden 2015 PISA sınavına katılan öğrencilerin fen başarı puanlarına ait istatistiksel özet bilgileri ve dağılım histogramları (Alt paneldeki histogramların x eksenı olasılık yoğunluklarını göstermektedir)

4.12.3 TRC Bölgesinde Makine Öğrenme Modeli

Tablo 38, TRC (Güneydoğu Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı ve 10 tekrarlı çapraz doğrulama performans metriklerini ve optimize edilmiş Hiperparametre değerlerini göstermektedir. Bu tabloda elde edilen bulgular, XGBoost, RF, SVR ve BRNN algoritmaları ile Kısım 3.3'te oluşturulan makine öğrenme modellerin 10 tekrarlı ve 10 katlı çapraz doğrulama tahmin performansları elde edilmiştir. Bu tablonun ayrıntılı sonuçları ve kısaltmalar için gerekli açıklamalar EK-1 içerisinde sunulmuştur. İncelenen ve oluşturulan makine öğrenmesi modellerin başarı performansları RMSE, R^2 ve MAE metrikleri üzerinden değerlendirilmiş ve XGBoost

modelinin en düşük RMSE (42.62), en yüksek R^2 (0.57) ve düşük MAE (33.41) değerleri ile eğitim verileri için en iyi performansı gösterdiği tespit edilmiştir.

Buna karşın SVR, RF ve BRNN modelleri sırasıyla 43.34, 44.61 ve 52.50 RMSE değerleri ile XGBoost algoritmasının ardından sıralanmışlardır. Bu veriler, TRC bölgesinin eğitim verileri için XGBoost modelinin Fen okuryazarlık puanlarının tahmininde diğer modellere göre daha iyi bir öğrenme ortaya koyduğunu göstermektedir.

Tablo 38: TRC (Güneydoğu Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Eğitim verilerinin 10 katlı çapraz doğrulama performans metrikleri ve optimize edilmiş Hiperparametre değerleri

Model	RMSE	R^2	MAE	Hiperparametre değerleri
XGBoost	42.62	0.57	33.41	nrounds 493.00000000 max_depth 3.00000000 eta 0.04146221 gamma 2.62185513 colsample_bytree 0.63833080 min_child_weight 0.00000000 subsample 0.55761709
SVR	43.34	0.55	33.27	sigma 0.004302388 C 1.225900717
RF	44.61	0.55	34.84	mtry "97" splitrule "extratrees" min.node.size "13"
BRNN	52.50	0.42	41.31	neurons 3

Tablo 39'da, TRC bölgesine ait verileri için oluşturulan makine öğrenme algoritmalarının test verisi performans göstergeleri detaylı bir şekilde sunulmaktadır. Bu tabloya göre, XGBoost modelinin RMSE değeri 42.27, R^2 değeri 0.61 ve MAE değeri 34.73 olarak belirlenmiş ve diğer modellerden daha üstün bir performans sergilediği tespit edilmiştir. Ayrıca, XGBoost modelinin uyum korelasyon katsayısı (ρ_c) merkezi tahmin edilen değeri 0.759, güven aralığının alt sınırı 0.685 ve üst sınırı 0.818 ile yüksek bir uyumu gösterdiği belirlenmiştir. Buna karşın ikinci en iyi performans gösteren SVR modelinde, RMSE değeri 43.17, R^2 değeri 0.60 ve MAE değeri 35.87 ile XGBoost'a göre daha düşük olup, uyum korelasyon katsayısı merkezi tahmini 0.740 (lwr.ci: 0.664 - upr.ci: 0.801) olarak hesaplanmıştır. RF ve BRNN modellerinde ise sırasıyla RMSE değerleri 45.07 ve 55.22, R^2 değerleri 0.59 ve 0.40, MAE değerleri 36.88 ve 44.40 olarak belirlenmiştir. RF modelinin uyum korelasyon katsayısı merkezi tahmini 0.677 (lwr.ci: 0.601- upr.ci: 0.740) iken, RF

modelinin uyum korelasyon katsayısı merkezi tahmini 0.627 (lwr.ci: 0.515- upr.ci: 0.718) olarak hesaplanmıştır. Bu sonuçlar, XGBoost modelinin en düşük hata oranları ve en yüksek açıklama gücü ile diğer modellerden daha üstün performans sergilediğini ortaya koymaktadır.

Tablo 39: TRC (Güneydoğu Anadolu) verileri için oluşturulan makine öğrenme algoritmalarının Test verisi Performans göstergeleri

Model	RMSE	R ²	MAE	ρ_c
XGBoost	42.27	0.61	34.73	XGBoost.est 0.7595747 XGBoost.lwr.ci 0.6852416 XGBoost.upr.ci 0.8182467
SVR	43.17	0.60	35.87	SVR.est 0.7401485 SVR.lwr.ci 0.6636476 SVR.upr.ci 0.8013276
RF	45.07	0.59	36.88	RF.est 0.6767312 RF.lwr.ci 0.6007903 RF.upr.ci 0.7405671
BRNN	55.22	0.40	44.40	BRNN.est 0.6267599 BRNN.lwr.ci 0.5149313 BRNN.upr.ci 0.7176131

ρ_c = Uyum Korelasyon Katsayısı (Concordance Correlation Coefficient)

est = ρ_c için merkezi tahmin edilen değeri

lwr.ci = güven aralığının alt sınırını (lower confidence interval)

upr.ci = güven aralığının üst sınırını (upper confidence interval)

Tablo 38 ve Tablo 39 verilerine göre yapılan genel değerlendirmede, XGBoost modelinin hem eğitim hem de test verilerinde en yüksek performansı sergilediği görülmektedir. Bu durum, modelin genelleştirme kabiliyetinin güçlü olduğunu göstermektedir. Ayrıca test verilerinin uyum korelasyon katsayısının (ρ_c) yüksek olması, modelin tahminlerinin gerçek verilerle yüksek uyum gösterdiğini belirtmektedir. Bu, XGBoost modelinin hem eğitim hem de test verilerinde tutarlı ve güvenilir bir performans sergilediğinin kanıtıdır.

Diğer yandan SVR, RF ve BRNN modelleri ise eğitim ve test verileri arasında benzer hata oranlarına sahip olsalar da performansları XGBoost modellerine kıyasla daha düşük seviyelerde kalmıştır. Bu durum SVR, RF ve BRNN modellerinin genelleştirme kabiliyetinin sınırlı olduğunu ve yeni veriler üzerinde daha düşük performans sergilediğini göstermektedir. Her üç modelin de ρ_c değerleri, tahminlerinin gerçek verilerle makul bir uyum gösterdiğini belirtmektedir.

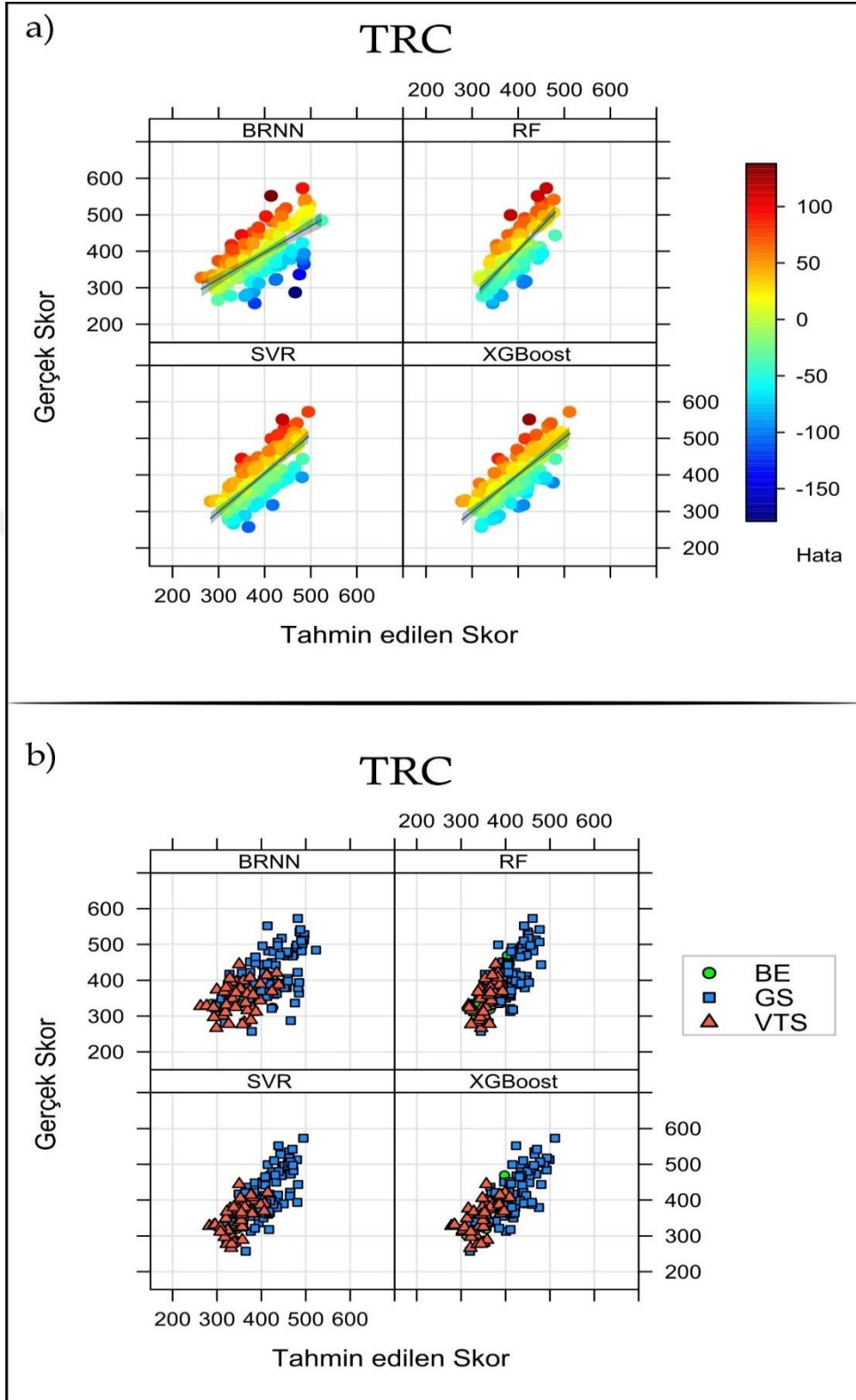
Şekil 60, Tablo 39 da elde edilen performans metriklerine ait çapraz doğrulama diyagramlarını tanımlamaktadır. Şekil 60a'da tüm modellerin hata dağılımlarının regresyon çizgisi boyunca ve yakınlarında dağıldığı görülmektedir. Bu durum tahmin hatalarının dağılımının normal yapıda olduğunu ve hata ortalamasının da sıfıra yakın olduğunu göstermektedir. Bu sonuçlar Tablo 39 da elde edilen bulgular ile uyumludur. Bunlarla birlikte XGBoost modelinin yüksek tahmin performansı sergilediği, Şekil 60a' daki çapraz doğrulama diyagramının hata dağılımlarının regresyon çizgisi etrafında sıkı şekilde toplanmasından dolayı daha açık bir şekilde görülmektedir.

Şekil 60b, aynı test verileri için modellerin farklı okul türlerine (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim) ait tahmin edilen ve gerçek skorlarının karşılaştırmasını göstermektedir. Burada özellikle tanımlayıcı regresyon çizgisinden sapan ve modellerin tahmin etme performansının düşmesine neden olan okul grubunun Mesleki ve Teknik Ortaöğretim (VTS) okul türü olduğu dikkat çekmektedir. Bununla birlikte XGBoost, her tüm okul türlerinde de tahmin edilen skorların gerçek skorlarla en uyumlu olduğu model olarak öne çıkmaktadır. Ayrıca TRC bölgesindeki Fen başarı skorları için genel ortaokul (GS) başarı düzeyinin VTS okul türüne göre belirgin şekilde artış göstermiş olması dikkat çekicidir. Özellikle okul türünün modellerin oluşturulmasında bir değişken olarak kullanılmamış olmasına rağmen, makine öğrenme algoritmasının bu başarı farklılıklarını diğer değişkenlerin yardımı ile tespit edebilmesi ve tahminlerini buna göre gerçekleştirmesi önemli bir başarı olduğu düşünülmektedir. Bu bulgular, XGBoost modelinin eğitim verileri üzerindeki güçlü performansını ve genelleştirme kabiliyetini destekler niteliktedir. XGBoost'un, farklı okul türlerindeki öğrencilerin performansını doğru bir şekilde tahmin edebilmesi, eğitim politikaları ve müdahalelerinin planlanmasında doğru tahmin skorları önerebileceğini ve bu konuda değerli bir araç olarak kullanılabileceğini göstermektedir. Özellikle Mesleki ve Teknik Ortaöğretim (VTS) öğrencileri için tahminlerdeki sapmalara rağmen, XGBoost'un genel olarak diğer modellere göre daha yüksek doğrulukla sonuçlar vermesi, bu modelin TRC bölgesi için geniş bir yelpazede güvenilir tahminler oluşturabileceğini göstermektedir.

Şekil 61, farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost) tahmin ettiği başarı skorları ile gerçek başarı skorlarının karşılaştırmalı kutu grafiğini sunmaktadır.

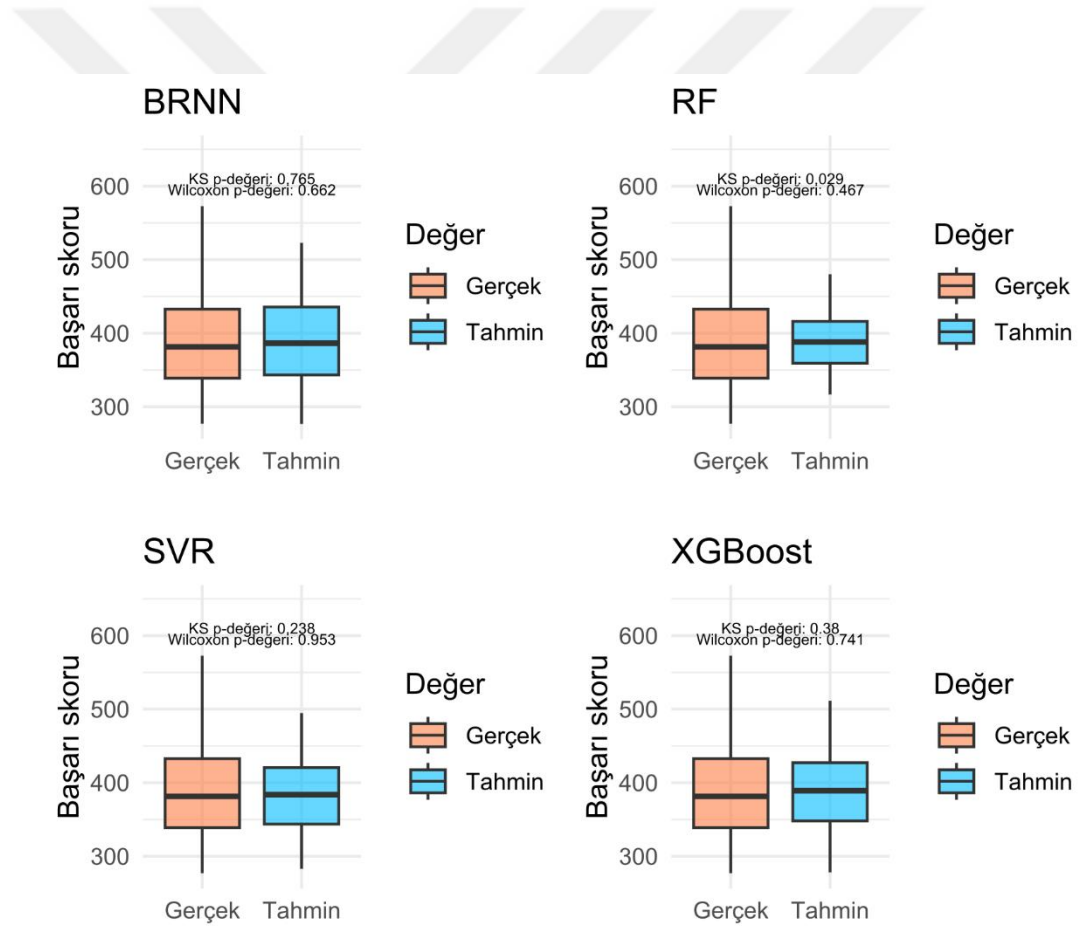
Şekil 60 de sunulan çapraz doğrulama bulgularına göre oluşturulmuş bu grafikler, modellerin tahmin performansını ve gerçek verilerle uyumunu değerlendirmek için önemli bir yakaşım sağlamaktadır. Buna göre tüm modellerin tahmin skorlarının hem Kolmogorov-Simirnov testi (KS) hem de Wilcoxon sıralama toplamı testine göre gerçek skorların genel dağılımlarından ve merkezi eğilimlerinden anlamlı ölçüde bir farklılığı ($p > 0.05$) olmadığı tespit edilmiştir. Bu durum makine öğrenmesi modellerinin tahmin edilen ve gerçek başarı skorları arasındaki uyumun yüksek olduğunu ortaya koymaktadır.





Şekil 60: a) Test verileri için her bir makine öğrenme modelinin çapraz doğrulama diyagramları, b) test verilerinin okul türlerine göre çapraz doğrulama diyagramları (BE: Temel Eğitim, GS: Ortaokul, VTS: Mesleki ve Teknik Ortaöğretim)

Şekil 61, en yüksek performans seviyesi gösteren XGBoost algoritması açısından değerlendirildiğinde, Kutu grafiklerinin gerçek ve tahmin skorları için temsil ettiği ortalama ve medyan değerlerinin neredeyse örtüştüğü ve dağılımların (kutu dışındaki uç örnekler için yeterli seviyelerde olmasa da) Q1-Q3 aralığı için oldukça yakın olduğunu görülmektedir. Şekil 61’deki bulgulara göre Tahmin edilen ve gerçek skorların genel dağılımlarının (KS p-değeri 0.038) ve merkezi eğilimlerinin (Wilcoxon p-değeri 0.741), istatistiksel olarak birbirlerinden farklı olmadığı belirlenmiştir. Bu durum, XGBoost algoritmasının TRC verileri üzerinde en yüksek doğruluk ve güvenilirliği sunmakla birlikte hem bireysel gerçek skorları hem de gerçek skorlara ait dağılım yapısını doğru şekilde tahmin edebilme yeteneğini ortaya koymaktadır.



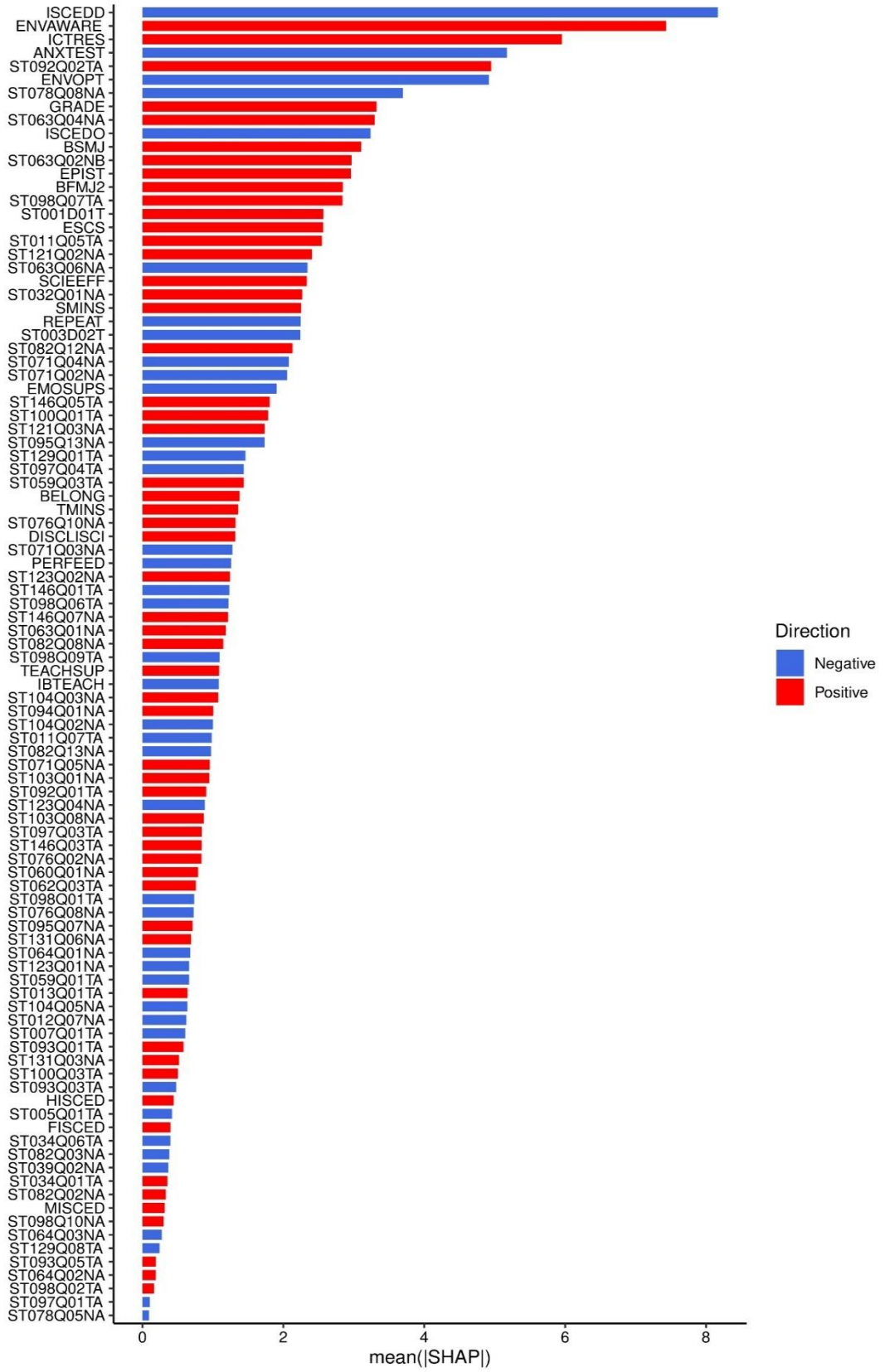
Şekil 61: Farklı makine öğrenme modellerinin (BRNN, RF, SVR, XGBoost), test alt veri seti için tahmin edilen ve gerçek başarı skorlarının karşılaştırmalı kutu grafikleri ve istatistiksel sınamaları.

4.12.4 TRC Bölgesi İçin Değişkenlerin Önem Seviyeleri

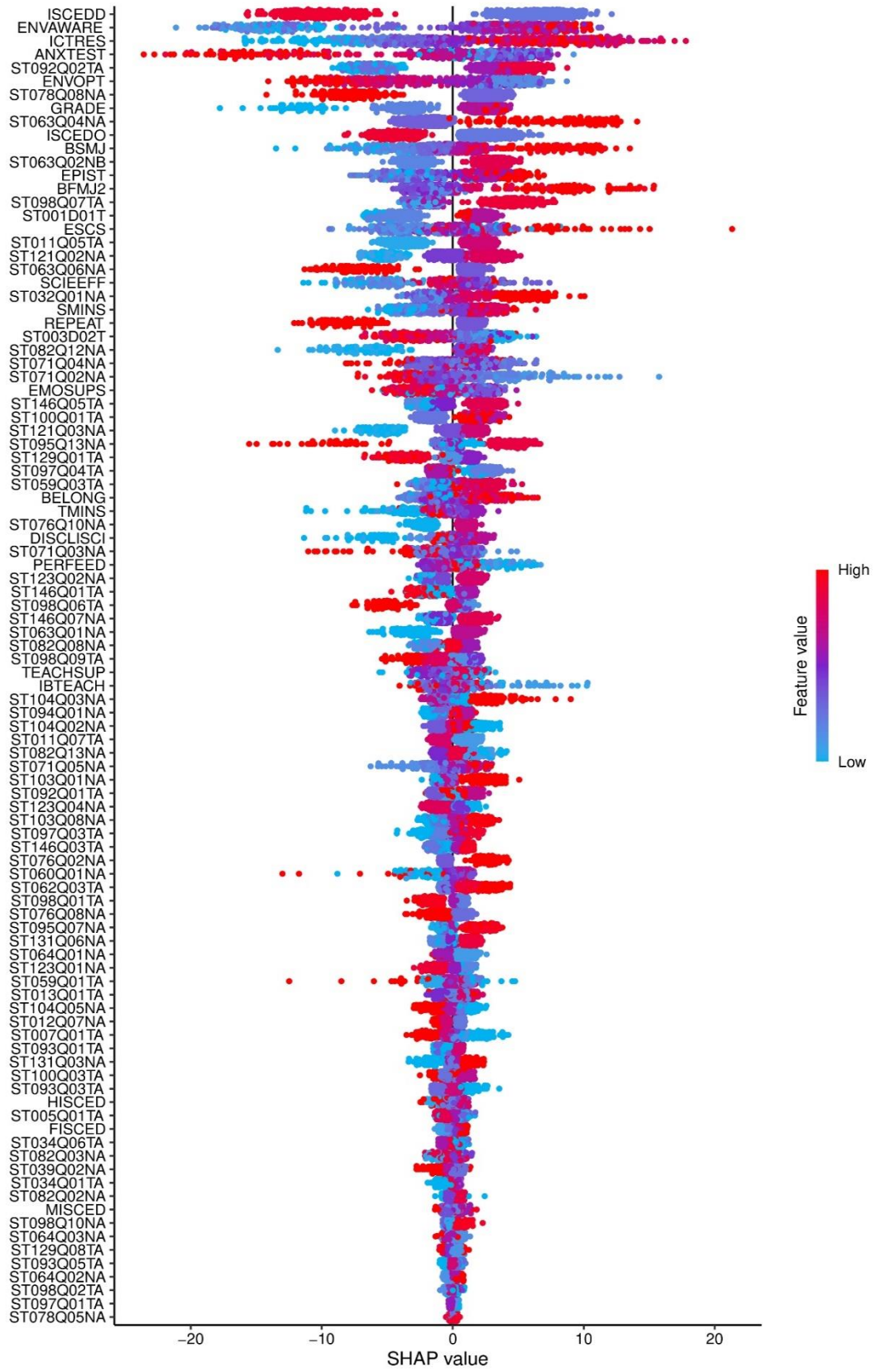
PISA 2015 Türkiye örnekleme TRC bölgesi Fen okuryazarlık alanı için en iyi tahmin performans gösteren XGBoost algoritmasına göre SHAP (SHapley Additive exPlanations) değerleri kullanılarak değişkenlerin önem parametreleri belirlenmiştir. Sonuçlar her bir değişkenin ortalama SHAP değerlerinin mutlak değerine göre azalan sırada Tablo 40'ta, sıralanmıştır. Bu sıralamaya göre diğer bölgelerdeki değişkenlerin karşılaştırması ise Kısım 4.13'te Jaccard indeksine göre yapılacaktır. Burada model için önemli olan sıralı değişkenler üç önem grubuna ayrılmış ve bu işlem model için kullanılan toplam değişkenleri 3 eşit parçaya ayırarak gerçekleştirilmiştir. Burada yüksek derecede önemli değişkenler 1. Grup (ilk 33 tanesi), orta derecede önemli değişkenler 2. Grup (34- 66) ve düşük derecede önemli değişkenler 3. Grup (67- 99) olarak etiketlenmiştir. Ayrıca Şekil 62 ile Tablo 40'ta özetlenen değişkenlerin ortalama SHAP değerleri görselleştirilmiştir. Bu şekil her bir tahmin edici değişkenin model tahminlerine olan ortalama etkisini ve yönünü (pozitif ve negatif katkıları) vurgulamaktadır. Şekil 63 her bir değişkenin test veri setindeki her bir gözlemdeki etkisini göstermektedir. Bu şekil içerisinde her bir nokta, bir veri noktasının bir özelliği için SHAP değerini temsil eder. Bu şeklin sağ tarafında pozitif SHAP değerleri, sol tarafında ise negatif SHAP değerleri bulunmaktadır. Tahmin edici değişkenlerin model çıktısına olan etkisi, bu SHAP değerlerinin büyüklüğüne göre belirlenir. Kırmızı noktalar, yüksek (pozitif) değerli özellikleri, mavi noktalar (negatif) ise düşük değerli özellikleri temsil etmektedir. Tablo 40, Şekil 62 ve 63'teki bulgulara göre, TRC bölgesi için model tahmininde önem seviyesi yüksek ilk 10 değişken, parantez içleri SHAP değerlerini ve model katkı yönlerini göstermek üzere, şu şekilde sıralanmıştır: ISCEDD (8.167, negatif), ENVWARE (7.433, pozitif), ICTRES (5.954, pozitif), ANXTEST (5.175, negatif), ST092Q02TA (4.950, pozitif), ENVOPT (4.918, negatif), ST078Q08NA (3.696, negatif), GRADE (3.324, pozitif), ST063Q04NA (3.297, pozitif), ISCEDO (3.238, negatif). Şekil 4 de Kırmızı ve mavi renkler, sırasıyla pozitif ve negatif yönlü etkileri temsil etmektedir. Bu, modelin tahminlerinin hangi özellikler tarafından ne kadar ve hangi yönde etkilendiğini daha net bir şekilde anlaşılmasına yardımcı olmaktadır. Buna göre negatif özellikteki değişkenlerin (ISCEDD, ANXTEST, ENVOPT, ST078Q08NA, ISCEDO) değerlerindeki artış öğrencilerin Fen başarı skorlarını düşürmekte buna karşı pozitif özellikteki değişkenlerin değerlerindeki artış ise öğrencilerin Fen başarı skorlarını artırıcı yönde etki etmektedir.

Tablo 40: TRC (Güneydoğu Anadolu) verileri için oluşturulan En yüksek performanslı algoritmaya (XGBoost) göre değişkenlerin önem seviyelerinin sıralanması.

ID	Değişken	SHAP değeri	Önem Grubu	ID	Değişken	SHAP değeri	Önem Grubu
1	ISCEDD	8.167	Grup 1	55	ST011Q07TA	0.984	Grup 2
2	ENVAWARE	7.433	Grup 1	56	ST082Q13NA	0.975	Grup 2
3	ICTRES	5.954	Grup 1	57	ST071Q05NA	0.956	Grup 2
4	ANXTEST	5.175	Grup 1	58	ST103Q01NA	0.951	Grup 2
5	ST092Q02TA	4.950	Grup 1	59	ST092Q01TA	0.906	Grup 2
6	ENVOPT	4.918	Grup 1	60	ST123Q04NA	0.887	Grup 2
7	ST078Q08NA	3.696	Grup 1	61	ST103Q08NA	0.872	Grup 2
8	GRADE	3.324	Grup 1	62	ST097Q03TA	0.843	Grup 2
9	ST063Q04NA	3.297	Grup 1	63	ST146Q03TA	0.842	Grup 2
10	ISCEDO	3.238	Grup 1	64	ST076Q02NA	0.839	Grup 2
11	BSMJ	3.105	Grup 1	65	ST060Q01NA	0.790	Grup 2
12	ST063Q02NB	2.967	Grup 1	66	ST062Q03TA	0.759	Grup 2
13	EPIST	2.962	Grup 1	67	ST098Q01TA	0.735	Grup 3
14	BFMJ2	2.844	Grup 1	68	ST076Q08NA	0.727	Grup 3
15	ST098Q07TA	2.840	Grup 1	69	ST095Q07NA	0.712	Grup 3
16	ST001D01T	2.570	Grup 1	70	ST131Q06NA	0.691	Grup 3
17	ESCS	2.566	Grup 1	71	ST064Q01NA	0.679	Grup 3
18	ST011Q05TA	2.548	Grup 1	72	ST123Q01NA	0.663	Grup 3
19	ST121Q02NA	2.407	Grup 1	73	ST059Q01TA	0.663	Grup 3
20	ST063Q06NA	2.344	Grup 1	74	ST013Q01TA	0.641	Grup 3
21	SCIEEFF	2.333	Grup 1	75	ST104Q05NA	0.640	Grup 3
22	ST032Q01NA	2.268	Grup 1	76	ST012Q07NA	0.623	Grup 3
23	SMINS	2.253	Grup 1	77	ST007Q01TA	0.608	Grup 3
24	REPEAT	2.247	Grup 1	78	ST093Q01TA	0.585	Grup 3
25	ST003D02T	2.241	Grup 1	79	ST131Q03NA	0.519	Grup 3
26	ST082Q12NA	2.131	Grup 1	80	ST100Q03TA	0.507	Grup 3
27	ST071Q04NA	2.079	Grup 1	81	ST093Q03TA	0.480	Grup 3
28	ST071Q02NA	2.054	Grup 1	82	HISCED	0.444	Grup 3
29	EMOSUPS	1.905	Grup 1	83	ST005Q01TA	0.420	Grup 3
30	ST146Q05TA	1.808	Grup 1	84	FISCED	0.400	Grup 3
31	ST100Q01TA	1.784	Grup 1	85	ST034Q06TA	0.398	Grup 3
32	ST121Q03NA	1.736	Grup 1	86	ST082Q03NA	0.383	Grup 3
33	ST095Q13NA	1.736	Grup 1	87	ST039Q02NA	0.368	Grup 3
34	ST129Q01TA	1.464	Grup 2	88	ST034Q01TA	0.355	Grup 3
35	ST097Q04TA	1.438	Grup 2	89	ST082Q02NA	0.332	Grup 3
36	ST059Q03TA	1.438	Grup 2	90	MISCED	0.316	Grup 3
37	BELONG	1.380	Grup 2	91	ST098Q10NA	0.301	Grup 3
38	TMINS	1.359	Grup 2	92	ST064Q03NA	0.276	Grup 3
39	ST076Q10NA	1.322	Grup 2	93	ST129Q08TA	0.244	Grup 3
40	DISCLSCI	1.317	Grup 2	94	ST093Q05TA	0.192	Grup 3
41	ST071Q03NA	1.279	Grup 2	95	ST064Q02NA	0.190	Grup 3
42	PERFEED	1.259	Grup 2	96	ST098Q02TA	0.165	Grup 3
43	ST123Q02NA	1.243	Grup 2	97	ST097Q01TA	0.104	Grup 3
44	ST146Q01TA	1.235	Grup 2	98	ST078Q05NA	0.094	Grup 3
45	ST098Q06TA	1.223	Grup 2	99	ST061Q01NA	0	Grup 3
46	ST146Q07NA	1.217	Grup 2				
47	ST063Q01NA	1.184	Grup 2				
48	ST082Q08NA	1.148	Grup 2				
49	ST098Q09TA	1.098	Grup 2				
50	TEACHSUP	1.089	Grup 2				
51	IBTEACH	1.084	Grup 2				
52	ST104Q03NA	1.074	Grup 2				
53	ST094Q01NA	1.006	Grup 2				
54	ST104Q02NA	1.001	Grup 2				



Şekil 62: TRC bölgesi için SHAP değerlerinin ortalamasını gösteren özellik önem sıralaması.



Şekil 63: TRC bölgesi test verileri için SHAP değerlerini gösteren özellik önem diyagramı.

4.13 Önem Parametresi Yüksek Değişkenlerin Karşılaştırılması

Bu çalışma sonunda elde edilen bulgulara göre, SHAP (Shapley Additive Explanations) değerleri kullanılarak belirlenen en önemli değişkenler, her bir bölge için analiz edilmiş ve değişkenlerin önem sıralamasına bağlı olarak bölgeler arası benzerlikleri Jaccard benzerlik indeksi kullanılarak değerlendirilmiştir.

Jaccard benzerlik indeksine göre önemi yüksek olan değişkenlerin benzerlikleri bölgeler temelinde “Çok Düşük”, “Düşük”, “Orta”, “Yüksek” ve “Çok Yüksek” olmak üzere beş ana sınıfta değerlendirilmiştir. Bu grafik, Jaccard indeksi ile benzer eleman sayısı arasındaki ilişkiyi temsil etmektedir ve bu ilişki, benzerlik seviyelerine göre kategorize edilmiştir.

Yatay eksen, "Benzer eleman sayısı"nı göstermekte olup, iki set arasındaki ortak eleman sayısını ifade eder. Dikey eksen ise "Jaccard indeksi"ni göstermektedir. Jaccard indeksi, iki kümenin kesişim elemanlarının birleşim elemanlarına oranı olarak hesaplanır ve benzerlik seviyesini ölçmek için sıklıkla kullanılan bir metriktir. Bu indeksi 0 ile 1 arasında değişir; 0, tamamen farklı kümeleri, 1 ise tamamen aynı kümeleri temsil eder.

Grafikteki verilere göre, artan benzer eleman sayısı ile Jaccard indeksinin pozitif bir doğrusal ilişki gösterdiği anlaşılmaktadır. Veriler, farklı renklerle kodlanmış beş ayrı benzerlik seviyesi ile gruplandırılmıştır:

Çok Düşük (Koyu mavi): Jaccard indeksinin düşük olduğu ve kümeler arasındaki benzerliğin minimum seviyede olduğu durumları ifade eder.

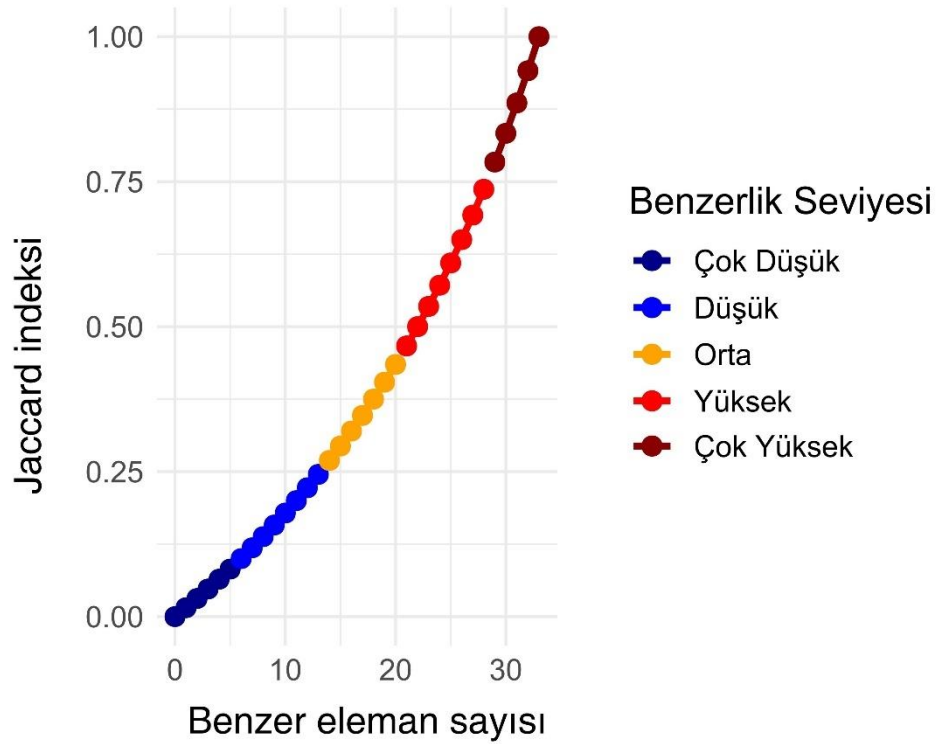
Düşük (Açık mavi): Jaccard indeksinin bir miktar arttığı, ancak benzerlik seviyesinin hala düşük olduğu durumlar.

Orta (Sarı): Orta seviyedeki benzerliği ifade eder ve kümeler arasındaki ortak eleman sayısının önemli ölçüde arttığını gösterir.

Yüksek (Kırmızı): Kümeler arasındaki benzerliğin belirgin şekilde yüksek olduğu durumlar.

Çok Yüksek (Koyu kırmızı): Kümelerin neredeyse tamamen aynı olduğunu veya çok yüksek bir kesişim yüzdesine sahip olduğunu ifade eder.

Grafik, iki küme arasındaki benzerliğin hem sayısal hem de görsel olarak değerlendirilmesine olanak tanımaktadır. Benzer eleman sayısı arttıkça, Jaccard indeksinin logaritmik bir şekilde artış göstermesi dikkat çekmektedir. Ayrıca, renk kategorileri, belirli bir indeks aralığını tanımlayarak, kümeler arasındaki benzerliğin seviyelerinin daha kolay yorumlanmasını sağlar. Bu sınıflar Şekil 64’teki dağılım eğrisine göre tespit edilmiştir. Böylece Jaccard indeksi ile bölgeler arasındaki ortak değişkenlerin oranı belirlenmiş ve elde edilen sayısal veriler ışığında bölgelerin eğitim performansına etki eden faktörlerin benzerliği veya farklılığı detaylı bir şekilde analiz edilmiştir. Çalışmada araştırılan toplam 99 değişken en iyi performans gösteren modellere göre SHAP değerlerine göre sıralanmış, bulgular her istatistiki bölge için oluşturulan kısımların sonunda tablo olarak verilmiş ve bu tablolarda önem seviyelerine göre sıralanmış değişkenler 3 gruba ayrılmıştır. Bu 99 değişkenin Türkçe karşılıkları EK2’de gösterilmiştir.



Şekil 64: Jaccard indeksi değerlerine karşılık gelen benzer değişken sayılarının gruplandırılması

Şekil 65’te elde edilen bulgulara göre görsel olarak oluşturulan Jaccard benzerlik matrisleri ve sınıflandırma haritalarını göstermektedir. Bu sınıflandırma haritasında köşegendeki değerler Türkiye istatistiki bölgelerini, köşegenin üstündeki sayılar ortak

değişken sayılarını ve köşegenin altındaki ondalıklı değerler bölgelerin benzerlik oranlarını ifade etmektedir. Sağdaki jaccard indeksi renk etiketi bu benzerlik oranlarının değerlerini bir renk skalasında tanımlamakla birlikte, soldaki benzerlik sınıfları ise köşegenin üstündeki ortak değişken sayılarına ait bilgiyi belirtmektedir. Buna göre elde edilen bulgular incelendiğinde, Şekil 65 ve her bölge için belirlenen önem seviyesi sıralaması tablolarına göre (Yukarıdaki kısımlarda her bölge için gerçekleştirilen analizlerin sonunda bu tablolar verilmiştir), aşağıda açıklanan sonuçlara ulaşılmıştır:

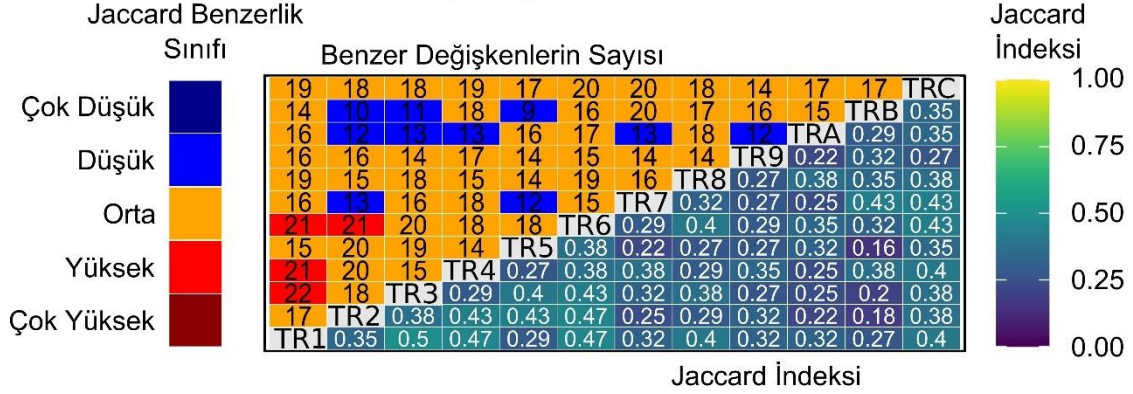
Şekil 65'te verilen diyagramlardan 1. Grup benzerlik matrisi incelendiğinde TR1 ile TR3 bölgesi (22 ortak değişken), TR1 ile TR4 (21 ortak değişken). TR1 ile TR6 (21 ortak değişken) ve TR2 ile TR6 (21 ortak değişken) bölgeleri arasında yüksek seviyede benzerlik düzeyleri tespit edilmiştir. 1. Grup için yüksek benzerlik gösteren değişkenlerin anlamları EK-3, EK-4, EK-5 ve EK-6'da verilmiştir. Şekil 65'te verilen diyagramlardan 1. Grup benzerlik matrisi incelendiğinde genel olarak orta seviyede benzerlik düzeyleri tespit edilmiştir. Bu duruma uymayan ve düşük benzerlik sınıfında olan birkaç bölge karşılaştırması da bulunmaktadır. Bunlar; TR1 bölgesi ile TR5 (15 ortak değişken) ve TR9 (15 ortak değişken) bölgeleri, TR2 bölgesi ile TR8 (15 ortak değişken), TR9 (14 ortak değişken), TRA (12 ortak değişken) ve TRB (14 ortak değişken) bölgeleri, TR3 bölgesi ile TRA (13 ortak değişken) bölgesi, TR4 bölgesi ile TRB (14 ortak değişken) bölgesi, TR5 bölgesi ile TR8 (14 ortak değişken), TR9 (13 ortak değişken) ve TRB (13 ortak değişken) bölgeleri ve TR9 bölgesi ile TRA (15 ortak değişken) bölgesidir. Orta sınıf grubunda ortak değişken sayısı ise 16 ile 20 arasında değişmektedir. Önem seviyesi yüksek olan 1. Grup değişkenler arasında bölgeler arasında çok düşük ve çok yüksek benzerlik düzeyleri tespit edilmemiştir.

Şekil 65'te verilen diyagramlardan 2. Grup benzerlik matrisi incelendiğinde bölgeler arasında çok düşük benzerlik düzeyleri baskın olup orta düzeyde benzerliklerin oldukça sınırlı olduğu görülmektedir. Orta düzey benzerlik gösteren bölgeler arasında sadece on dokuz karşılaştırma bulunmaktadır. Bunlardan yüksek seviyeye en yakın olan TR1 bölgesi ile TR4 (17 ortak değişken) ve TR7 bölgesi ile TR8 (17 ortak değişken) bölgeleridir. Yüksek seviyeye en yakın olan bölgeler için benzer değişkenler ve anlamları EK-7 ve EK-8 de verilmiştir. Bu bölgeler harici tüm karşılaştırmalarda benzerlik seviyelerileri çok düşük benzerlik sınıfı aralığında kalmıştır.

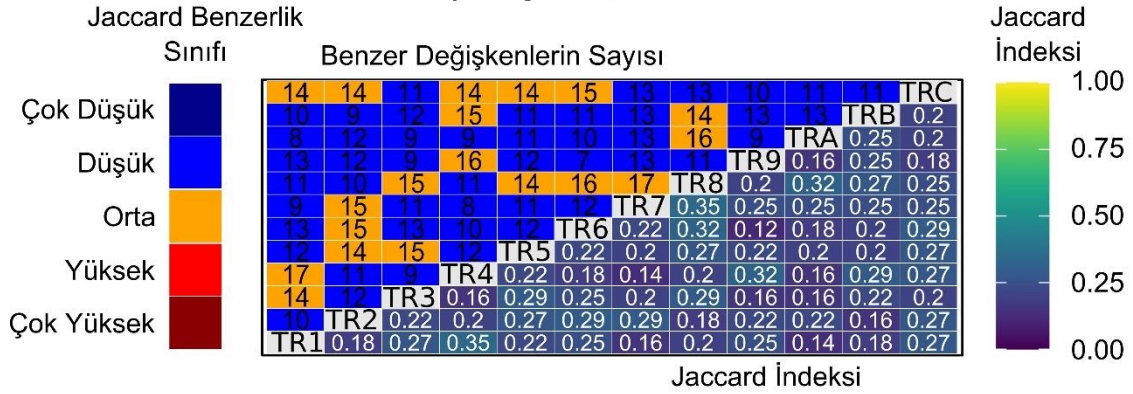
Şekil 65'te verilen diyagramlardan 3. Grup benzerlik matrisi incelendiğinde TR6 bölgesi ile TR8 bölgesi (23 ortak değişken) ve TR1bölgesi ile TR8 (21 ortak değişken) bölgeleri arasında yüksek düzeyde benzerlik olduğu görülmektedir. Yüksek benzerlik gösteren ortak değişkenlerin anlamı EK-9 ve EK-10'da verilmiştir. Şekil 65'te verilen diyagramlardan 3. Grup benzerlik matrisi incelendiğinde genel olarak orta seviyede benzerlik gösterdiği görülmektedir. Bunların haricinde bazı bölgelerdeki benzerlik düzeylerinin düşük seviyelerde olduğu dikkat çekmektedir. Bunlar; TR1 bölgesi ile TR2 (14 ortak değişken) ve TRA (15 ortak değişken) bölgeleri, TR7 bölgesi ile TR4 (15 değişken) ve TR5 (14 değişken) bölgeleri, TR9 bölgesi ile TR5 (13 değişken) TR7 (15 ortak değişken) bölgesi, TRB bölgesi ile TR5 (13 ortak değişken) ve TR2 (14 ortak değişken) bölgesi, TRC bölgesi ile TR2 (13 ortak değişken), TR3 (15 ortak değişken), TR5 (15 ortak değişken) ve TR7 (15 ortak değişken) bölgeleridir.

Bulgular genel olarak değerlendirildiğinde, bölgeler arası eğitim performansını etkileyen faktörlerin yalnızca belirli bir kısmının örtüştüğü, ancak tamamen aynı olmadığı görülmektedir. Özellikle 1. ve 3. grup değişkenler için baskın olan orta seviyedeki benzerlikler, 2. grup değişkenler için tespit edilen düşük seviyedeki benzerlik düzeyleriyle birlikte bu durumu açıkça ortaya koymaktadır. Değişkenlerin önem sıralarına ait benzerlik oranlarının düşük ve orta seviyelerde kalması, bölgeler arasındaki fen başarısını etkileyen faktörlerin çeşitlilik gösterdiğini ve bu faktörlerin bölgesel koşullara bağlı olarak farklı ağırlıklarda öne çıktığını göstermektedir. Dolayısıyla, bölgeler arası eğitim dinamiklerindeki bu farklılıklar, her bölgeye özel eğitimsel yöntemlerin geliştirilmesini zorunlu kılmaktadır. Elde edilen bulgular ışığında, Türkiye'nin farklı bölgelerindeki eğitim performansını etkileyen faktörlerin detaylı bir şekilde analiz edilmesi, veriye dayalı eğitim politikalarının oluşturulması için bir temel teşkil etmektedir. Özellikle orta seviyede benzerlik gösteren bölgeler için mevcut eğitim stratejilerinin harmonize edilmesi kaynakların daha etkin kullanımına olanak tanırken, düşük benzerlik düzeylerine sahip bölgelerde özgün ve yerel dinamiklere uygun politikaların hayata geçirilmesi fen başarısının artırılması için önemli bir adım olacağı düşünülmektedir. Makine öğrenmesi teknikleri, SHAP metodu ve Jaccard indeksiyle yapılan bu analizlerin PISA 2015 verilerine dayalı sonuçları, eğitim sistemindeki bölgesel eşitsizliklerin giderilmesine yönelik stratejik kararlar alınmasını desteklemekle birlikte, her bölgenin potansiyelinin en üst düzeye çıkarılmasına katkı sağlayacak somut veriler sunmaktadır.

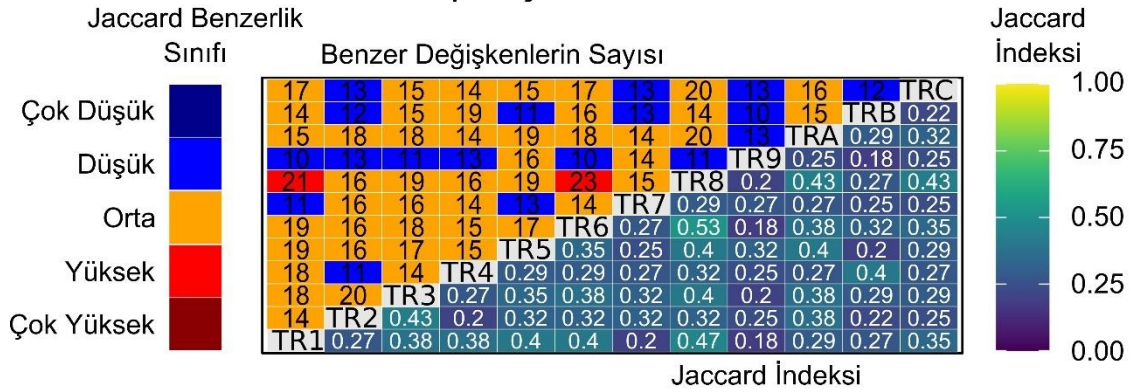
Grup 1 için Benzerlik Matrisi



Grup 2 için Benzerlik Matrisi



Grup 3 için Benzerlik Matrisi



Şekil 65: Bölgeler arasındaki jaccard benzerlik matrisi

BEŞİNCİ BÖLÜM

SONUÇ VE ÖNERİLER

Bu araştırmada, PISA 2015 Türkiye verileri kullanılarak Türkiye'nin istatistiki bölgelerine göre fen bilimleri başarı skorlarının tahmin edilmesinde makine öğrenmesi algoritmalarının performansları incelenmiştir. Bulgular, bölgeler arasındaki eğitim performansı farklılıklarını detaylı bir şekilde ortaya koymuş ve her bölgenin kendine özgü değişkenlerinin başarıya etkisini değerlendirmiştir. Çalışmadan elde edilen sonuçları üç ana başlık altında aşağıdaki gibi özetleyebiliriz.

5.1 Makine Öğrenmesi Algoritmalarının Performansı

Bu çalışmada, Türkiye'nin istatistiki bölgelerindeki öğrencilerin PISA 2015 fen bilimleri başarı skorlarını tahmin etmek amacıyla dört farklı makine öğrenmesi algoritması değerlendirilmiştir. Kullanılan algoritmalar arasında Support Vector Regression (SVR), Random Forest (RF), Bayesian Regularized Neural Networks (BRNN) ve Extreme Gradient Boosting (XGBoost) yer almaktadır. Bu algoritmaların performansları hem eğitim hem de test verileri üzerinde karşılaştırmalı olarak analiz edilmiştir.

Bulgular, makine öğrenmesi algoritmalarının başarı skorlarını tahmin etme konusunda önemli farklılıklar gösterdiğini ve bu algoritmaların bölgesel düzeyde değişken performans sergilediğini ortaya koymuştur. Özellikle XGBoost ve Support Vector Regression (SVR) modelleri, test verileri üzerinde üstün performans göstererek diğer algoritmalara kıyasla daha güvenilir tahminler sağlamıştır. Bu başarı, XGBoost ve SVR modelinin yüksek boyutlu veri setlerinde etkili bir tahmin aracı olmasıyla ilişkilendirilebilir. Ancak, bazı bölgelerin analizinde XGBoost modeli, eğitim verilerinde oldukça yüksek performans elde etmesine rağmen test verilerinde aynı başarıyı sürdürememiştir. Bu durum, modelin aşırı öğrenme (overfitting) sorunu yaşadığını ve verileri ezberleme eğiliminde olduğunu göstermektedir. Random Forest (RF) ve Bayesian Regularized Neural Networks (BRNN) modelleri ise orta düzeyde performans sergilemiştir. RF algoritması, doğrusal olmayan değişken ilişkilerini tespit etmekte başarılı olsa da, bazı bölgelerde tahmin doğruluğu beklenen seviyenin altında kalmıştır. Bu durum, RF modelinin karmaşık değişken etkileşimlerini analiz etme kapasitesinin sınırlı olduğuna ve fen bilimleri başarı skorlarındaki karmaşık ilişkileri yeterince öğrenemediğine işaret etmektedir. Öte yandan, BRNN modeli, küçük veri setleri

üzerinde başarılı sonuçlar vermesiyle bilinse de, bu çalışmada kullanılan geniş kapsamlı veri setinde düşük tahmin doğruluğu sergilemiştir.

Elde edilen bulgular aynı zamanda, makine öğrenmesi algoritmalarının eğitim verilerini analiz ederken geleneksel yöntemlere kıyasla daha hızlı ve daha doğru sonuçlar ürettiğini ortaya koymaktadır. Bu bağlamda, PISA gibi geniş ölçekli veri setlerinin analizinde makine öğrenmesi uygulamaları, öğrenci başarılarının tahmin edilmesi, bölgesel farklılıkların tespit edilmesi ve eğitim politikalarının başarı performanslarının karşılaştırılması açısından büyük bir potansiyel taşıdığı sonucuna ulaşılmıştır. Ancak algoritmaların performanslarının modelleme sürecindeki parametre ayarlarına, kullanılan veri setinin yapısına ve değişkenlerin özelliklerine bağlı olarak değişebileceği göz önünde bulundurulmalıdır.

5.2 Önem Seviyesi Yüksek Değişkenler ve Bölgesel Etkiler

PISA 2015 Fen Bilimleri başarı skorlarını etkileyen değişkenlerin Türkiye'nin istatistiki bölgelerine göre önem sıralaması analiz edilmiş, SHAP (SHapley Additive Explanations) yöntemi kullanılarak bu değişkenlerin katkı payları belirlenmiş ve böylece bölgeler arasında başarı üzerinde hangi faktörlerin ne ölçüde etkili olduğu detaylı olarak incelenmiştir.

Araştırma bulguları, Fen Bilimleri başarı skorlarını etkileyen değişkenlerin tüm bölgeler için belirli ortak özellikler taşıdığını, ancak her bir bölgenin dinamiklerine bağlı olarak bu değişkenlerin etkisinin farklılık gösterdiğini ortaya koymuştur. Bu bağlamda, ANXTEST (Kaygı Testi Skoru) değişkeninin, genel olarak fen bilimleri başarı skorları üzerinde olumsuz bir etkiye sahip olduğu görülmüştür. Öğrencilerde kaygı düzeyinin artması, Fen başarısını doğrudan düşüren bir faktör olarak tespit edilmiştir. Özellikle TR1 (İstanbul), TR3 (Ege), TR6 (Akdeniz), TR7 (Orta Anadolu), TRC (Güneydoğu Anadolu) ve TRA (Kuzeydoğu Anadolu) bölgelerinde Fen başarısının tahmininde öğrencilerin kaygı düzeyinin, diğer bölgelere kıyasla daha yüksek etkide olduğu tespit edilmiştir. Bu durum, bölgesel eğitim ortamlarının öğrenciler üzerinde oluşturduğu psikolojik baskı ve sınav kaygısı gibi faktörlerle açıklanabilir. Bununla birlikte, ST013Q01TA (evde bulunan kitap sayısı) ve ISCEDO (ailedeki ebeveynlerinden birinin sahip olduğu en yüksek eğitim seviyesi) değişkenleri, bazı bölgelerde fen bilimleri başarı skorlarını en güçlü şekilde etkileyen değişkenlerden ikisi olarak belirlenmiştir. Bu değişkenler, öğrencinin akademik

başarısının sadece okul içi öğrenme ile sınırlı olmadığını, ailelerin eğitim kültürü ve öğrenmeyi teşvik eden bir ortam oluşturmalarının da başarı üzerinde kritik bir role sahip olduğunu göstermektedir. ST013Q01TA değişkeni için TR2 (Batı Marmara), TR4 (Doğu Marmara), TR6 (Akdeniz), TR8 (Batı Karadeniz), TRA (Kuzeydoğu Anadolu) ve TRB (Ortadoğu Anadolu) bölgelerinde kitap sayısı ile başarı skorları arasındaki pozitif korelasyon daha belirgin iken, TR9 (Doğu Karadeniz) ve TRC (Güneydoğu Anadolu) bölgelerinde bu etkinin daha sınırlı kaldığı gözlemlenmiştir. Benzer şekilde ISCEDO değişkeni için TR1 (İstanbul), TR7 (Orta Anadolu), TR8 (Batı Karadeniz), TR9 (Doğu Karadeniz) bölgeri hariç diğerlerinde yüksek pozitif korelasyon ve yüksek önem düzeyi etkisi gösterdiği belirlenmiştir. Bu farklılık, sosyoekonomik koşulların etkisiyle eğitim kültürünün bölgesel düzeyde farklılaştığını göstermektedir.

Fen başarısının tahmin edilmesinde önemli değişkenlerden ikisinde BSMJ (Bilime yönelik meslek hedefi) ve ST092Q02TA (Evdeki eğitim materyalleri) değişkenleri olarak belirlenmiştir. BSMJ değişkeni, öğrencilerin gelecekte bilime dayalı meslek hedefleri belirlemelerinin fen başarısına olumlu katkı sağladığını göstermektedir. BSMJ değişkeni TR2 (Batı Marmara) hariç geri kalan bütün bölgelerde önem parametresi yüksek değişkenlerden biri olarak belirlenmiş ve bilimsel meslek hedefleri olan öğrencilerin fen bilimleri skorlarının anlamlı düzeyde yüksek olduğu gözlemlenmiştir. Ayrıca, Bu çalışma bulgularına göre ST092Q02TA değişkenin, yani evde eğitim materyallerinin (örneğin kitaplar, öğrenme araçları) varlığının fen başarısı üzerine başarı üzerindeki yüksek önem etkisini tespit edilmiştir. Bu değişkenin özellikle TRC (Güneydoğu Anadolu), TRB (Ortadoğu Anadolu), TR8 (Batı Karadeniz), TR5 (Batı Anadolu) ve TR4 (Doğu Marmara) bölgelerinde yüksek önem seviyesine sahip olduğu, bu bölgelerde eğitim materyallerinin yetersizliğinin öğrencilerin akademik performansını olumsuz yönde etkileyebileceği tespit edilmiştir. Bu bulgular, öğrencilerin bilimsel hedefler oluşturmalarının ve eğitim materyallerine erişiminin bölgesel başarı farklılıklarını azaltmada kritik bir rol oynadığını göstermektedir.

Araştırma kapsamında, dikkat çeken bir başka bulgu öğretmenlerin tutum ve davranışlarının öğrencilerin akademik başarıları üzerine her bölgede yüksek düzeyde etkisi olmadığının tespit edilmesi olmuştur. ST095Q13NA (Öğretmen adilliği ve motivasyonu) değişkeni, TR5 bölgesinde başarı skorlarının tahmin edilmesinde önemli değişkenler arasında olmasına rağmen diğer bölgelerde bu değişkenin etkisi sınırlı düzeyde kalmıştır. Bu

bulgular, öğretmenlerin öğrencilere rehberlik etme süreçlerinin bölgesel dinamiklere göre farklılaştığını ortaya koymaktadır. Benzer başka bir bulgu ST063Q04NA (öğrencinin okula aidiyet duygusu) değişkeni için belirlenmiştir. TR4 ve TRC bölgeleri için öğrencilerin okula aidiyet duygusunun, fen bilimleri başarı skorlarını olumlu yönde etkilediği tespit edilmiştir. Buna karşın, geri kalan diğer bölgelerde okula aidiyet duygusunun fen başarı skorlarının tahmin edilmesinde daha düşük önem düzeyine sahip olduğu belirlenmiştir.

5.3 Bölgeler Arası Farklılıklar ve Değişkenlerin Etkisi

PISA 2015 Fen Bilimleri başarı skorları üzerinden Türkiye'nin istatistiksel bölgeleri arasında yapılan analizde, başarı düzeyleri ile bu başarıyı etkileyen ana değişkenler ayrıntılı bir şekilde karşılaştırılmıştır. Jaccard benzerlik indeksi kullanılarak yapılan analizde, araştırma sırasında değişkenlerin bölgeler arasındaki ortak ve farklı yönleri değerlendirilmiş, sonuçlar bölgelerin Fen başarı derecelerini etkileyen faktörlerin büyük ölçüde yerel dinamiklere bağlı olduğunu göstermiştir.

Araştırma bulguları, bölgeler arasında orta düzeyde benzerlik olduğunu göstermiştir. Ancak, bazı bölgelerde başarı skorlarına etki eden değişkenlerin farklı önem sıralarına sahip olduğu belirlenmiştir. özellikle TR2 (Batı Marmara), ve TR5 (Batı Anadolu) bölgelerinin fen bilimleri başarı skorlarının tahmin edilmesi açısından değişkenlerin önem parametrelerinin sırlanmasında belirgin farklar gösterdiğini tespit edilmiştir. Bu farklılıklar, bölgelerin sosyoekonomik ve kültürel yapısına dayalı olarak eğitim süreçlerinin farklılaştığını göstermektedir. Bu sonuçlar, eğitim politikalarının bölgesel ihtiyaçlara göre özelleştirilmesi gerektiğini açıkça ortaya koymaktadır.

5.4 Öneriler

Araştırmanın bulguları ışığında, ileride yapılacak çalışmalar ve mevcut çalışmanın nasıl geliştirilebileceğine yönelik aşağıdaki öneriler sunulmaktadır.

Farklı Veri Setleri Kullanımı: Bu çalışmada PISA 2015 verileri temel alınmıştır. Gelecekteki çalışmalarda PISA'nın sonraki yıllara ait verileri (2018, 2022 vb.) veya farklı uluslararası değerlendirme programları (TIMSS, PIRLS gibi) kullanılarak benzer analizler yapılabilir. Böylece zaman içinde bölgeler arasındaki başarı değişimleri incelenerek dinamik bir değerlendirme sağlanabilir.

Makine Öğrenmesi Algoritmalarının Çeşitlendirilmesi: Bu çalışmada Random Forest, XGBoost, SVR ve BRNN algoritmaları kullanılmıştır. Gelecek çalışmalarda derin öğrenme (Deep Learning) teknikleri, hiyerarşik modeller, ensemble yöntemleri ve stacked modeller gibi daha karmaşık algoritmaların performansı da değerlendirilerek model doğruluğu artırılabilir.

Farklı Değişkenlerin Dahil Edilmesi: Çalışmada analiz edilen değişkenlerin kapsamı genişletilebilir. Özellikle çevresel faktörler, bölgesel kalkınma verileri, okul fiziksel koşulları, okul odaklı diğer değişkenler gibi ek değişkenler modele dahil edilerek fen bilimleri başarısını etkileyen faktörlerin daha kapsamlı bir analizi yapılabilir.

Boylamsal Analizlerin Yapılması: Gelecek çalışmalarda, öğrencilerin akademik başarılarının zaman içindeki gelişimini incelemek amacıyla boylamsal analizler gerçekleştirilebilir. Bu tür analizler, belirli bir bölgedeki eğitim politikalarının uzun vadeli etkilerini değerlendirme açısından da önemli bilgiler sunabilir.

Bölgesel Eşitsizliklerin Mekansal Analizi: Çalışmanın bölgesel verilerle sınırlı kalmaması için coğrafi bilgi sistemleri (GIS) kullanılarak fen bilimleri başarı skorlarının mekansal dağılımları analiz edilebilir. Böylece, başarı skorlarındaki mekansal eğilimler ve bölgesel eşitsizlikler haritalandırılarak görsel bir analiz sunulabilir.

Okul ve Öğrenci Düzeyinde Çok Düzeyli Modeller: Gelecek çalışmalarda hiyerarşik doğrusal modelleme (HLM) veya çok düzeyli modeller kullanılarak, öğrenci, okul ve bölge düzeyindeki faktörlerin etkileşimleri daha detaylı incelenebilir. Bu modeller, başarıyı etkileyen değişkenlerin farklı seviyelerdeki etkilerini analiz etme imkanı sağlar.

Diğer Alanlarının Analizi: Bu çalışmada fen bilimleri başarı skorları analiz edilmiştir. Gelecek çalışmalarda matematik ve okuma becerileri gibi diğer PISA alanlarına odaklanılarak bölgeler arasındaki farklılıkların bu alanlardaki yansımaları da incelenebilir. Böylece daha bütüncül bir değerlendirme yapılabilir.

Sosyoekonomik Faktörlerin Derinlemesine Analizi: Fen bilimleri başarı skorlarını etkileyen sosyoekonomik faktörlerin detaylı bir şekilde incelenebilmesi için nedensel analizler yapılabilir. Bu kapsamda, aile gelir düzeyi, ebeveyn eğitimi ve okul kaynakları gibi

değişkenlerin fen başarısı üzerindeki doğrudan ve dolaylı etkileri farklı veri setleri ile incelenebilir ve sonuçlar karşılaştırılabilir.

Uluslararası Karşılaştırmalar: Gelecek çalışmalarda Türkiye'nin PISA verileri, fen bilimleri başarısında öne çıkan ülkelerle karşılaştırılabilir. Bu tür karşılaştırmalar, Türkiye'nin eğitim sistemindeki güçlü ve zayıf yönlerin belirlenmesine katkı sağlayacaktır.

Fen Eğitimcilerine Öneriler: PISA 2015 verileri üzerine yapılan bu detaylı analiz, fen bilimleri başarısının sadece sınıf içi öğrenmeyle sınırlı olmadığını, aynı zamanda öğrencilerin kaygı düzeylerinden ailelerinin eğitim seviyesine, evlerindeki kaynaklardan bilimsel hedeflerine kadar pek çok faktörün etkileşimiyle şekillendiğini göstermektedir. Öncelikle, öğrencilerin kaygı düzeylerini anlamaya çalışmak ve bu konuda destekleyici bir yaklaşım sergilemek büyük önem taşımaktadır. Özellikle İstanbul, Ege, Akdeniz, Orta Anadolu, Güneydoğu Anadolu ve Kuzeydoğu Anadolu bölgelerinde görev yapan öğretmenlerin bu konuda daha dikkatli olması gerektiği düşünülmektedir.. Sınıf içi öğrenme ortamını daha güvenli ve teşvik edici hale getirerek, öğrencilerin kaygılarını azaltmalarına yardımcı olunabilir.

Öğrencilerin evdeki öğrenme ortamının öneminin farkında olmalısınız. Evde bulunan kitap sayısı ve ailelerin eğitim düzeyi gibi faktörler, öğrencinin akademik başarısında önemli bir rol oynamaktadır. Batı Marmara, Doğu Marmara, Akdeniz, Batı Karadeniz, Kuzeydoğu Anadolu ve Ortadoğu Anadolu bölgelerinde görev yapan öğretmenlerin, öğrencileri ve ailelerini bu konuda bilinçlendirmesi, mümkünse okuldaki kaynakları bu öğrencilere ulaştırmaları faydalı olacaktır. Özellikle Güneydoğu ve Ortadoğu Anadolu bölgelerinde görevli öğretmenler, evdeki eğitim materyallerinin eksikliğinin öğrencilerin başarısı üzerindeki olumsuz etkisini göz önünde bulundurarak, kaynaklara erişimi kolaylaştıracak projeler geliştirebilirler. Ayrıca, öğrencilerin bilimsel meslek hedefleri belirlemelerine yardımcı olmak, öğrencilerin derse olan motivasyonunu ve başarılarını artıracakı düşünülmektedir.

Öğrenci başarısında önemli bir faktör olarak belirtilen öğretmen adilliği ve motivasyonu, her bölgede aynı öneme sahip değildir. Bu nedenle, her bölgenin dinamiklerini ve öğrencilerin ihtiyaçlarını göz önünde bulundurarak daha kişiselleştirilmiş bir yaklaşım benimsenebilir. Batı Anadolu bölgesindeki öğretmenlerin, adil ve motive edici bir sınıf ortamı yaratmaya özellikle önem vermesi önemlidir. Ayrıca, öğrencilerin okula aidiyet duygusunun desteklenmesi, Doğu Marmara ve Güneydoğu Anadolu bölgelerinde başarıya

önemli bir katkı sağlayabilir. Öğrencilerin kendilerini okul ortamına ait hissetmelerini sağlamak için sosyal etkinlikler ve etkileşimli sınıf ortamları oluşturulabilir. Her bir öğrencinin benzersiz olduğunu ve onların başarı yolculuklarında öğretmenlere büyük bir rol düştüğü unutulmamalıdır. Bölgesel farklılıkları ve öğrencilerin bireysel ihtiyaçlarını göz önünde bulundurarak, öğrencilerin fen başarısını artırılabilir.



KAYNAKLAR

- Acar, T. ve Öğretmen, T. (2012). Çok Düzeyli İstatistiksel Yöntemler ile öğrencilerin 2006 PISA Fen Bilimleri Performansının İncelenmesi. *Eğitim ve Bilim*, 37(163), 178–189.
- Açıslı Çelik, S. ve Yeşilkanat, C. M. (2023). Predicting science achievement scores with machine learning algorithms: a case study of OECD PISA 2015--2018 data. *Neural Computing and Applications*, 35, 21201–21228. doi:10.1007/s00521-023-08901-6
- Akkoyun, S. (2020). Estimation of fusion reaction cross-sections by artificial neural networks. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 462, 51–54. doi:10.1016/J.NIMB.2019.11.014
- Akkoyun, S. ve Bayram, T. (2014). Estimations of fission barrier heights for Ra, Ac, Rf and Db nuclei by neural networks. *International Journal of Modern Physics E*, 23(10). doi:10.1142/S0218301314500645
- Aksu, G. ve Doğan, N. (2018). Veri Madenciliğinde Kullanılan Öğrenme Yöntemlerinin Farklı Koşullar Altında Karşılaştırılması (in Turkish). *Ankara University Journal of Faculty of Educational Sciences*, 51(3), 71–100. doi:10.30964/auebfd.464262
- Aksu, G. ve Güzeller, C. O. (2016). Classification of PISA 2012 mathematical literacy scores using Decision-Tree Method: Turkey sampling. *Eğitim ve Bilim*, 41(185), 101–122. doi:10.15390/EB.2016.4766
- Anil, D. (2009). Uluslararası Öğrenci Başarılarını Değerlendirme Programı (PISA)’ nda Türkiye ’ deki Öğrencilerin Fen Bilimleri Başarılarını Etkileyen Faktörler *Factors Effecting Science Achievement of Science Students in Programme for International Students ’ Achieve*. *Eğitim ve bilim*, 34(152).
- Aromsuk, T., Kongnakorn, S. ve Noonpakdee, W. (2023). Classification Covid-19 Data Using XGBoost and SHAP (Shapley Additive exPlanations). 2023 8th International Conference on Business and Industrial Research, ICBIR 2023 - Proceedings, (May), 901–906. doi:10.1109/ICBIR57571.2023.10147537
- Awad, M. ve Khanna, R. (2015). Efficient learning machines: theories, concepts, and applications for engineers and system designers. Springer nature. doi:10.1007/978-1-4302-5990-9
- Aydın, N. ve Yılmaz, A. (2010). Yapılandırıcı Yaklaşımın Öğrencilerin Üst Düzey Bilişsel Becerilerine Etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 39, 57–68.
- Balabin, R. M. ve Lomakina, E. I. (2011). Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. *Analyst*, 136(8), 1703–1712. doi:10.1039/C0AN00387E
- Batı, K., Yetişir, M. İ. ve Güneş, G. (2019). Data PISA 2015 Sonuçlarına Göre Duyuşsal Özelliklerin Öğrencilerin Fen Performansları Üzerine Etkisinin Ülkeler. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 34(4), 999–1014. doi:10.16986/HUJE.2019050025

- Bayraklı, B. ve Dengiz, O. (2020). An evaluation of heavy metal pollution risk in tea cultivation soils of micro-catchments using various pollution indexes under humid environmental condition. *Rendiconti Lincei*, 31(2), 393–409. doi:10.1007/s12210-020-00901-1
- Berberoğlu, G., Çalışkan, M. ve Karşı, N. (2019). Variables Predicting PISA Scientific Literacy Scores in Turkey. *Uluslararası Bilim ve Eğitim Dergisi*, 2(2), 38–49.
- Bernard, S., Adam, S. ve Heutte, L. (2012). Dynamic Random Forests. *Pattern Recognition Letters*, 33(12), 1580–1586. doi:10.1016/j.patrec.2012.04.003
- Biau, G. ve Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227. doi:10.1007/s11749-016-0481-7
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. doi:10.1023/A:1018054314350
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.
- Buldu, E. ve Olgan, R. (2018). Fen Okur-yazarlık Becerileri Arasındaki Farklılaşmanın Bazı Göstergeler Açısından İncelenmesi: PISA-Türkiye Bulguları. *International Journal of Human Sciences*, 15(3), 1453–1465. doi:10.14687/jhs.v15i3.5279
- Burden, F. R. ve Winkler, D. A. (1999). Robust QSAR Models Using Bayesian Regularized Neural Networks. *Journal of Medicinal Chemistry*. doi:10.1021/jm980697n
- Burden, F. ve Winkler, D. (2009). Bayesian Regularization of Neural Networks BT - Artificial Neural Networks: Methods and Applications. D. J. Livingstone (Ed.), (ss. 23–42). Totowa, NJ: Humana Press. doi:10.1007/978-1-60327-101-1_3
- Carslaw, D. C. ve Ropkins, K. (2012). Openair - An r package for air quality data analysis. *Environmental Modelling and Software*, 27–28, 52–61. doi:10.1016/j.envsoft.2011.09.008
- Chen, J., Zhang, Y., Wei, Y. ve Hu, J. (2021). Discrimination of the Contextual Features of Top Performers in Scientific Literacy Using a Machine Learning Approach. *Research in Science Education*, 51(Suppl 1), 129–158. doi:10.1007/s11165-019-9835-y
- Chen, M., Xin, J., Tang, Q., Hu, T., Zhou, Y. ve Zhou, J. (2024). Explainable machine learning model for load-deformation correlation in long-span suspension bridges using XGBoost-SHAP. *Developments in the Built Environment*, 20(August), 100569. doi:10.1016/j.dibe.2024.100569
- Chen, T. ve Guestrin, C. (2016). Extreme Gradient Boosting [R package xgboost version 1.5.0.2]. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-Aug, 785–794. doi:10.1145/2939672.2939785
- Chung, N. C., Miasojedow, B., Startek, M. ve Gambin, A. (2019). Jaccard / Tanimoto similarity test and estimation methods for biological presence-absence data. *BMC Bioinformatics* 2019, 20(644), 1–11. doi:10.1186/s12859-019-3118-5
- Cortes, C. ve Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(2), 273–

297. doi:10.1111/j.1747-0285.2009.00840.x

De, S. (2003). Weighting Support Vector Regression Algorithm. Computer Science. <https://api.semanticscholar.org/CorpusID:63973143> adresinden erişildi.

Demir, S. B. ve Yıldırım, Ö. (2016). Okulda ve Okul Dışında Bilgi ve İletişim Teknolojilerinin Kullanımının Öğrencilerin PISA 2012 Performansı ile İlişkisinin İncelenmesi. *Kastamonu Eğitim Dergisi*, 24(1), 251–262.

Demircioğlu, T. (2023). Sağlık veri modellenmesi: makine öğrenmesi yaklaşımları. Marmara Üniversitesi.

Demirez, G. (2018). Bazı Değişkenlerin Fen Başarı Puanına Etkisi: PISA 2015 Türkiye, Singapur ve Almanya Örneği. Antalya, Türkiye.

Dong, X. X., An, R., Lu, J. X. ve Geng, L. S. (2022). Novel Bayesian neural network based approach for nuclear charge radii. *Physical Review C*, 105(1). doi:10.1103/PHYSREVC.105.014308

Drucker, H., Surges, C. J. C., Kaufman, L., Smola, A. ve Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, 1, 155–161.

Eraslan, A. (2009). Finlandiya'nın PISA'daki Başarısının Nedenleri : Türkiye için Alınacak Dersler. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, 3(2), 238–248.

Erdoğan, S. ve Taştan, H. (2024). Predicting Student Achievement via Machine Learning: Evidence from Turkish Subset of PISA. *Yildiz Social Science Review*, 10(1), 7–27. doi:10.51803/yssr.1461030

Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189–1232.

Fryer, D., Strümke, I. ve Nguyen, H. (2020). Shapley Value Confidence Intervals for Attributing Variance Explained. *Front. Appl. Math. Stat.*, 6(December), 1–17. doi:10.3389/fams.2020.587199

Gebreyesus, Y., Dalton, D., Nixon, S., De Chiara, D. ve Chinnici, M. (2023). Machine Learning for Data Center Optimizations: Feature Selection Using Shapley Additive exPlanation (SHAP). *Future Internet*, 15(3), 1–17. doi:10.3390/fi15030088

Gholamy, A., Kreinovich, V. ve Kosheleva, O. (2018). Why 70 / 30 or 80 / 20 Relation Between Training and Testing Sets : A Pedagogical Explanation. *Journal of Intelligent Technologies and Applied Statistics*, 11(2), 105–111. doi:10.6148/IJITAS.201806_11(2).0003

Golino, H. ve Gomes, C. M. A. (2014). Predicting Academic Achievement of High-School Students Using Machine Learning. *Psychology*, 5, 1536–1549. doi:10.4236/psych.2014.518207

Greenwell, B. (2024). fastshap: Fast Approximate Shapley Values: <https://cran.r->

project.org/web/packages/fastshap/index.html.

Güllüpnar, F., Kuzu, A., Dursun, Ö. Ö., Kurt, A. A. ve Gültekin, M. (2013). Milli Eğitimde Teknoloji Kullanımı ve Sonuçları: Velilerin Bakış Açısından Fatih Projesi'nin Pilot Uygulamasının Değerlendirilmesi. *SDÜ Fen Edebiyat Fakültesi Sosyal Bilimler Dergisi*, 30, 195–216.

Gür, B. S., Çelik, Z. ve Özoglu, M. (2012). Policy options for Turkey : a critique of the interpretation and utilization of PISA results in Turkey. *Journal of Education Policy*, 27(1), 1–21. doi:10.1080/02680939.2011.595509

Gür, D. (2019). Türkiye'deki Öğrencilerin PISA 2015 Fen Okuryazarlığını Yordayan Değişkenlerin Bölgelere Göre İncelenmesi. Antalya, Türkiye.

Güre, Ö. B., Şevgin, H. ve Kayri, M. (2023). Reviewing the Factors Affecting PISA Reading Skills by Using Random Forest and MARS Methods. *International Journal of Contemporary Educational Research*, 10(1), 181–196. doi:10.33200/ijcer.1192590

Güzel, Z. (2018). Fen Bilimleri Öğretiminde Öz ve Akran Değerlendirme Uygulamalarının Yer Aldığı Probleme Dayalı Öğrenme Yaklaşımının Öğrencilerin Başarı ve Tutumlarına Etkisi. Konya, Türkiye.

Habenicht, O. K. ve Goldhammer, F. (2020). ICT Engagement : a new construct and its assessment in PISA 2015. *Large-scale Assessments in Education*, 8(6), 1–21. doi:10.1186/s40536-020-00084-z

Hacıoğlu, G. F. (2024). Makine öğrenmesi algoritmaları kullanılarak Pisa verilerine göre öğrenme düzeylerine etki eden faktörlerin belirlenmesi. Artvin Çoruh Üniversitesi / Lisansüstü Eğitim Enstitüsü / Fen ve Matematik Eğitimi Ana Bilim Dalı.

Hacıoğlu, Gözde Fatma. (2024). Makine Öğrenmesi Algoritmaları Kullanılarak PISA Verilerine Göre Öğrenme Düzeylerine Etki Eden Faktörlerin Belirlenmesi. Artvin, Türkiye.

Hastie, T., Friedman, J. ve Tibshirani, R. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (C. 2). New York: Springer.

Ho, T. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844. doi:10.1109/34.709601

Hu, J., Peng, Y. ve Ma, H. (2021). Examining the contextual factors of science effectiveness: a machine learning-based approach. *School Effectiveness and School Improvement*. doi:10.1080/09243453.2021.1929346

Hwang, J., Choi, K. M., Bae, Y. ve Shin, D. H. (2018). Do Teachers' Instructional Practices Moderate Equity in Mathematical and Scientific Literacy?: an Investigation of the PISA 2012 and 2015. *International Journal of Science and Mathematics Education*. doi:10.1007/s10763-018-9909-8

Immekus, J. C., Jeong, T. ve Yoo, J. E. (2022). Machine learning procedures for predictor variable selection for schoolwork-related anxiety: Evidence from PISA 2015 mathematics,

reading, and science assessments. *Large-scale Assessments in Education*, 10, 30. doi:10.1186/s40536-022-00150-8

Immekus, J. C., Jeong, T. ve Yoo, J. E. (2022). Machine learning procedures for predictor variable selection for schoolwork-related anxiety: Evidence from PISA 2015 mathematics, reading, and science assessments. *Large-scale Assessments in Education*, 10, 30. doi:10.1186/s40536-022-00150-8

Jaccard, P. (1901). Distribution comparée de la fore alpine dans quelques régions des Alpes occidentales et orientales. *Bull. Soc. Vaudoise Sci. Nat.*, 37, 241–272.

Kahraman, Ü. ve Çelik, K. (2017). Analysis of PISA 2012 results in terms of some variables<p>PISA 2012 sonuçlarının bazı değişkenler açısından incelenmesi. *Journal of Human Sciences*, 14(4), 4797. doi:10.14687/jhs.v14i4.5136

Karasu, U. (2019). Fen Okuryazarlığını Etkileyen Bazı Sosyal Değişkenlerin PISA Verilerine Göre İncelenmesi. Ankara, Türkiye.

Karatzoglou, A., Smola, A., Hornik, K. ve Zeileis, A. (2004). kernlab - An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, 11(9), 1–20. <http://www.jstatsoft.org/v11/i09/> adresinden erişildi.

Kasap, Y., Doğan, N. ve Koçak, C. (2021). PISA 2018’de Okuduğunu Anlama Başarısını Yordayan Değişkenlerin Veri Madenciliği İle Belirlenmesi. *Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, 19(4), 241–258. doi:10.18026/cbayarsos.959609

Kim, D. ve Cho, S. (2008). Bootstrap Based Pattern Selection for Support Vector Regression. *Pacific-Asia Conference on Knowledge Discovery and Data Mining* içinde. <https://api.semanticscholar.org/CorpusID:22095287> adresinden erişildi.

Koyuncu, İ. ve Gelbal, S. (2020). Comparison of Data Mining Classification Algorithms on Educational Data under Different Conditions *. *Journal of Measurement and Evaluation in Education and Psychology*, 11(4), 325–345. doi:10.21031/epod.696664.Received

Kuhn, M. (2019). The caret Package. 24 Mart 2023 tarihinde <https://topepo.github.io/caret/index.html> adresinden erişildi.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., ... Hunt, T. (2020). caret: Classification and Regression Training. R Packag. version 6.0-86. <https://CRAN.R-project.org/package=caret>.

Kurt, Ş., Yüksel, M. K. ve Burcu Dinçergök. (2023). *Journal of Turkish Data driven approach for weight restricted data envelopment analysis*, (1984), 1768–1779.

Kwok, J. T.-Y. ve Yeung, D. Y. (1996). Bayesian Regularization in Constructive Neural Networks. *International Conference on Artificial Neural Networks* içinde. <https://api.semanticscholar.org/CorpusID:2554466> adresinden erişildi.

Kılıç, S., Çene, E. ve Demir, İ. (2012). Comparison of Learning Strategies for Mathematics Achievement in Turkey with Eight Countries. *Educational Sciences: Theory & Practice*, 12(4), 2594–2598. <https://www.edam.com.tr/estp> adresinden erişildi.

Lau, K. ve Lam, T. Y. (2017). Instructional practices and science performance of 10 top-performing regions in PISA 2015. *International Journal of Science Education*. doi:10.1080/09500693.2017.1387947

Lee, S. (2017). Improving Jaccard Index Using Genetic Algorithms for Collaborative Filtering BT - Advances in Swarm Intelligence. Y. Tan, H. Takagi ve Y. Shi (Ed.), (ss. 378–385). Cham: Springer International Publishing.

Lezhnina, O. ve Kismihók, G. (2022). Combining statistical and machine learning methods to explore German students' attitudes towards ICT in PISA. *International Journal of Research and Method in Education*, 45(2), 180–199. doi:10.1080/1743727X.2021.1963226

Li, W., Yin, Y., Quan, X. ve Zhang, H. (2019). Gene Expression Value Prediction Based on XGBoost Algorithm. *Frontiers in Genetics*, 10(November), 1–7. doi:10.3389/fgene.2019.01077

Lin, L. I. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*, 45(1), 255–268 (URL: <https://www.jstor.org/stable/2532051>). doi:10.2307/2532051

Lin, Y., Lee, Y. ve Wahba, G. (2002). Support vector machines for classification in nonstandard situations. *Machine Learning*, 46(1–3), 191–202. doi:10.1023/A:1012406528296

MEB, (2010). Pisa 2009 Ulusal Ön raporu, 31(3), 331–351. URL: https://pisa.meb.gov.tr:443/eski_dosyalar/wp-content/uploads/2013/07/PISA-2009-Ulusal-On-Rapor.pdf

MEB, (2016). Uluslararası Öğrenci Değerlendirme Programı PISA 2015 Ulusal Raporu. URL: https://odsgm.meb.gov.tr/test/analizler/docs/PISA/PISA2015_Ulusal_Rapor.pdf

MEB, (2022). 2022 PISA Türkiye Raporu. URL: https://pisa.meb.gov.tr/meb_iys_dosyalar/2024_03/21120745_26152640_pisa2022_rapor.pdf

Masci, C., Johnes, G. ve Agasisti, T. (2018). Politecnico di Milano, School of Management, via Lambruschini 4/b, Milano, Italy, 1–54.

Medhat, W., Hassan, A. ve Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. doi:10.1016/j.asej.2014.04.011

Meng, L., Qiu, C. ve Boyd-Wilson, B. (2019). Germany : Evidence from PISA 2015 data. *British Journal of Educational Technology*, 50(6), 3233–3251. doi:10.1111/bjet.12729

Merrick, L. ve Taly, A. (2020). The Explanation Game: Explaining Machine Learning Models Using Shapley Values BT - Machine Learning and Knowledge Extraction. A. Holzinger, P. Kieseberg, A. M. Tjoa ve E. Weippl (Ed.), (ss. 17–38). Cham: Springer International Publishing.

Mitchell, R. ve Frank, E. (2017). Accelerating the XGBoost algorithm using GPU

computing. *PeerJ Computer Science*, 2017(7). doi:10.7717/peerj-cs.127

Molnar, C. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2. bs.). <https://christophm.github.io/interpretable-ml-book> adresinden erişildi.

Nizam, H. ve Akın, S. S. (2014). Sosyal Medyada Makine Öğrenmesi i le Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması. *inet-tr'14 - XIX. Türkiye'de İnternet Konferansı Bildirileri*.

OECD. (2014). Education at a Glance 2019. *Education at a Glance: OECD Indicators*. doi:10.1787/f8d7880d-en

OECD. (2015). URL: <https://www.oecd.org/pisa/data/2015database/> . Erişim tarihi: 23/12/2023.

OECD. (2022). PISA 2022 OECD Ülke Raporu: Türkiye. Ankara, Türkiye: T.C. Milli Eğitim Bakanlığı.

Oktaay Güzeller, C. ve Şeker, F. (2016). Uluslararası Öğrenci Değerlendirme Programı'nda (PISA 2009) Öğrencilerin Fen Başarılarının Çeşitli Değişkenler Açısından İncelenmesi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, 10(2), 1–1. doi:10.17522/balikesirnef.273863

Onan, A. ve Korukoğlu, S. (2016). A review of literature on the use of machine learning methods for opinion mining. *Pamukkale University Journal of Engineering Sciences*, 22(2), 111–122. doi:10.5505/pajes.2015.90018

Orhan, B. (2020). Investigation Of The Effect Of Student And School Background Variables, Teaching And Learning Variables And Non-Cognitive Outcomes On The Components Of Scientific Literacy In Programme For International Student Assessment (Pisa 2015). Doktora Tezi, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye.

Özerbaş, M. A. ve Safi, B. N. (2022). Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi The Comparative Examination of the Countries that Have Become Successful TIMSS and PISA and the Turkish Teacher Training Systems TIMSS ve PISA ' DA Başarılı Olan Ülkeler ve Türk Öğretmen Yetiştirme. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi* <http://kefad.ahievran.edu.tr>, 23, 1960–1992. doi:10.29299/kefad.898782

Öztürk, Ö. (2018). Using PISA 2015 Data to Analyze How the Scientific Literacy of Students from Different Socioeconomic Levels Can Be Predicted by Environmental Awareness and by Environmental Optimism. İhsan Doğramacı Bilkent University.

Panov, P. ve Džeroski, S. (2007). Combining Bagging and Random Subspaces to Create Better Ensembles. *Advances in Intelligent Data Analysis VII* içinde (ss. 118–129). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-74825-0_11

Prasad, A. M., Iverson, L. R. ve Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, 9(2), 181–199. doi:10.1007/s10021-005-0054-1

Putra, N. S. ve Agusnita. (2021). Scientific literacy competency of senior high school

students based on question formats. *Journal of Physics: Conference Series*, 1882. doi:10.1088/1742-6596/1882/1/012161

R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <http://www.r-project.org/> (date accessed: 18.08.2024) adresinden erişildi.

Resino, D. A., Constante-Amores, A., Castro, M. ve Navarro, E. (2024). School Effectiveness and High Reading Achievement of Spanish Students in PISA 2018: A Machine Learning Approach. *Educación XX1*, 27(2), 223–251. doi:10.5944/educxx1.38634

Rodriguez, P. P. ve Gianola, D. (2022). <https://cran.r-project.org/web/packages/brnn/index.html>. RCran.

Rodríguez-Pérez, R. ve Bajorath, J. (2020). Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*, 34(10), 1013–1026. doi:10.1007/s10822-020-00314-0

Romero, C. ve Ventura, S. (2007). Educational data mining : A survey from 1995 to 2005, 33, 135–146. doi:10.1016/j.eswa.2006.04.005

Rumelhart, D. E., Hinton, G. E. ve Williams, R. J. (1986). Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* içinde. Biometrika. doi:10.7551/mitpress/4943.003.0128

Saarela, M., Yener, B., Zaki, M. J. ve Kärkkäinen, T. (2015). This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail. *International Journal of Special Education*, 30(2), 70–84.

Sakata, R., Ohama, I. ve Taniguchi, T. (2018). An extension of gradient boosted decision tree incorporating statistical tests. *IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018-Novem, 964–969. doi:10.1109/ICDMW.2018.00139

Salvatore, S., Rand, K. D., Grytten, I., Ferkingstad, E., Domanska, D., Holden, L., ... Sandve, G. K. (2020). Beware the Jaccard : the choice of similarity measure is important and non-trivial in genomic colocalisation analysis. *Briefings in Bioinformatics*, 21(5), 1523–1530. doi:10.1093/bib/bbz083

Scornet, E., Biau, G. ve Vert, J.-P. (2015). Consistency of Random Forests. *The Annals of Statistics*. doi:10.1214/15-aos1321

Shapley, L. (1953). Contributions to the Theory of Games (AM-28), Volume II. T. A. (ed. . Kuhn H (Ed.), A Value for n-Person Games. içinde (ss. 307–318). Princeton University Press. <https://doi.org/10.1515/9781400881970-018>. doi:10.1515/9781400881970-018

She, H. (2019). Reflections on and implications of the programme for international student assessment 2015 performance of students in Taiwan: The role of epistemic beliefs about science in scientific literacy, (March). doi:10.1002/tea.21553

She, H. (2019). Reflections on and implications of the programme for international student

assessment 2015 performance of students in Taiwan: The role of epistemic beliefs about science in scientific literacy, (March). doi:10.1002/tea.21553

Shelley, M. ve Yildirim, A. (2013). Transfer of Learning in Mathematics, Science, and Reading among Students in Turkey: A Study Using 2009 PISA Data. *International Journal of Education in Mathematics, Science and Technology (IJEMST)*, 1(2), 83–95. <http://www.ijemst.com> adresinden erişildi.

Sheng, C. ve Yu, H. (2022). An optimized prediction algorithm based on XGBoost. *Proceedings - 2022 International Conference on Networking and Network Applications, NaNA 2022*, 442–447. doi:10.1109/NaNA56854.2022.00082

Soares, P. S. M. (2024). Combining Statistical and Machine Learning Methods to Identify Predictors of Brazilian Students' Proficiency in PISA 2018. *Journal of Psychoeducational Assessment*, 42(8), 927–954. doi:10.1177/07342829241262861

Soydan, N. T. Y., Gacar, B. K. ve Akay, E. C. (2020). Machine learning and economics: bibliometric analysis. *Pressacademia*, 12(1), 104–105. doi:10.17261/pressacademia.2020.1367

Štrumbelj, E. ve Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions, 647–665. doi:10.1007/s10115-013-0679-x

Su, Q., Chen, L. ve Qian, L. (2024). Optimization of big data analysis resources supported by XGBoost algorithm: Comprehensive analysis of industry 5.0 and ESG performance. *Measurement: Sensors*, 36(July), 101310. doi:10.1016/j.measen.2024.101310

Suganyadevi, M. V. ve Babulal, C. K. (2014). Support Vector Regression Model for the prediction of Loadability Margin of a Power System. *Applied Soft Computing Journal*, 24, 304–315. doi:10.1016/j.asoc.2014.07.015

Şhribanoğlu, S. (2024). PISA 2018 Türkiye Örnekleminde Okuma Becerisini Etkileyen Değişkenlerin Boruta Algoritması ile Belirlenmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 57, 655–701. doi:10.30964/auebfd.1254457

Ulukan, G. (2019). PISA 2015 Verileri ve Ülkelerin Gelişmişlik Düzeyleri ile Öğrencilerin Fen Okuryazarlığının İncelenmesi. Aksaray, Türkiye.

Ulukan, G. (2019). PISA 2015 Verileri ve Ülkelerin Gelişmişlik Düzeyleri ile Öğrencilerin Fen Okuryazarlığının İncelenmesi. Aksaray, Türkiye.

Uluğ, S. (2019). PISA 2015 Türkiye Uygulamasında Bazı Öğrenci Değişkenlerinin Fen Okuryazarlığı ve Okuma Becerileri Başarısına Etkisinin İncelenmesi. Antalya, Türkiye.

Uğuz, E., Şahin, S. ve Yılmaz, R. (2021). PISA 2018 Fen Bilimleri Puanlarının Değerlendirilmesinde Eğitsel Veri Madenciliğinin Kullanımı The Use of Educational Data Mining in the Evaluation of PISA 2018 Scores of Science, 3(X), 212–227.

Üstün, U., Özdemir, E., Cansız, M. ve Cansız, N. (2019). Türkiye'deki Öğrencilerin Fen Okuryazarlığını Etkileyen Faktörler Nelerdir? PISA 2015 Verisine Dayalı Bir Hiyerarşik Doğrusal Modelleme Çalışması. *Eğitim Fakültesi Dergisi*. doi:10.16986/HUJE.2019050786

- Van Buuren, S. ve Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1–67. doi:10.18637/jss.v045.i03
- Van Den Broeck, G., Lykov, A., Suciu, D. ve Schleich, M. (2022). On the Tractability of SHAP Explanations. *Journal of Artificial Intelligence Research*, 74, 851–886.
- Verikas, A., Gelzinis, A. ve Bacauskiene, M. (2011). Mining data with random forests: A survey and results of new tests. *Pattern Recognition*, 44(2), 330–349. doi:10.1016/j.patcog.2010.08.011
- Verma, V. ve Aggarwal, R. (2020). A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective. *Social Network Analysis and Mining*, 10(1), 1–16. doi:10.1007/s13278-020-00660-9
- Wang, M., Fu, W., He, X., Hao, S. ve Wu, X. (2020). A Survey on Large-scale Machine Learning. *IEEE Transactions on Knowledge and Data Engineering*. <https://arxiv.org/abs/2008.03911> adresinden erişildi.
- Wen, Z., He, B., Kotagiri, R., Lu, S. ve Shi, J. (2018). Efficient gradient boosted decision tree training on GPUs. *Proceedings - 2018 IEEE 32nd International Parallel and Distributed Processing Symposium (IPDPS)*, 234–243. doi:10.1109/IPDPS.2018.00033
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag. <https://ggplot2.tidyverse.org> adresinden erişildi.
- Wickham, H., François, R., Henry, L., Müller, K. ve Vaughan, D. (2023). dplyr: A Grammar of Data Manipulation.
- Woodhams, J., Grant, T. I. M. D. ve Price, A. R. G. (2007). From Marine Ecology to Crime Analysis: Improving the Detection of Serial Sexual Offences Using a Taxonomic Similarity Measure. *J. Investig. Psych. Offender Profil*, 4(June), 17–27. doi:10.1002/jip.55
- Wright, M. N. ve Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77(1 SE-Articles), 1–17. doi:10.18637/jss.v077.i01
- Wu, C. L., Chau, K. W. ve Li, Y. S. (2008). River stage prediction based on a distributed support vector regression. *Journal of Hydrology*, 358(1–2), 96–111. doi:10.1016/j.jhydrol.2008.05.028
- Wu, J., Liu, H., Wei, G., Song, T., Zhang, C. ve Zhou, H. (2019). Flash flood forecasting using support vector regression model in a small mountainous catchment. *Water (Switzerland)*, 11(7). doi:10.3390/w11071327
- Yesilkanat, C. M. (2020). Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm. *Chaos, Solitons and Fractals*, 140, 14–39. doi:10.1016/j.chaos.2020.110210
- Yesilkanat, C. M., Kobya, Y., Taskin, H. ve Çevik, U. (2017). Spatial interpolation and radiological mapping of ambient gamma dose rate by using artificial neural networks and

fuzzy logic methods. *Journal of Environmental Radioactivity*, 175–176, 78–93. doi:10.1016/j.jenvrad.2017.04.015

Yeşilkanat, C. M. ve Akkoyun, S. (2023). Estimation of fission barrier heights for even-even superheavy nuclei using machine learning approaches. *Journal of Physics G: Nuclear and Particle Physics*, 50(5), 55101. doi:10.1088/1361-6471/acbaaf

Yitik, E. (2019). Uluslararası Öğrenci Başarılarını Değerlendirme Programı (PISA) 2015 Sınavına Türkiye’de Katılan Öğrencilerin Bireysel, Ailesel ve Okula Ait Değişkenlerin Fen Başarısını Yordama Durumunun İncelenmesi. Burdur, Türkiye.

Yoo, J. E. ve Rho, M. (2020). Exploration of Predictors for Korean Teacher Job Satisfaction via a Machine Learning Technique, Group Mnet. *Frontiers in Psychology*, 11, 441. doi:10.3389/fpsyg.2020.00441

You, H., Park, S. ve Delgado, C. (2021). A closer look at US schools: What characteristics are associated with scientific literacy?: A multivariate multilevel analysis using PISA 2015 A closer look at US schools: What characteristics are associated with scientific literacy?: A multivariate mul. *Science Education*, 105(2), 406–437. doi:10.1002/sce.21609

Yu, P. S., Chen, S. T. ve Chang, I. F. (2006). Support vector regression for real-time flood stage forecasting. *Journal of Hydrology*, 328(3–4), 704–716. doi:10.1016/j.jhydrol.2006.01.021

Yüksel, M. (2019). PISA 2015 Türkiye ve Finlandiya Verilerine Göre Okul Özellikleri ile Öğrencilerin Okuma Becerileri Arasındaki İlişkinin İncelenmesi. Edirne, Türkiye.

Yıldırım, A. ve Karaman, Z. (2018). Ailenin Yoksulluğu/Zenginliği İle Eğitim/Akademik Başarı Arasındaki İlişki: PISA 2015 Verileri İle Ampirik Bir Çalışma. *International Journal of Academic Value Studies (Javstudies)*, 4(19), 343–350. <http://www.javstudies.com> adresinden erişildi.

Zaman, N., Doğaner, G. D., Yücel, D., Zaman, I., Karataş, E. ve Karaş, C. (2022). Fen Okuryazarlığının PISA Verilerine Göre İncelenmesi. *International Journal of Disciplines Economics & Administrative Sciences Studies*, 8(48), 895–907. doi:10.29228/ideas.66224

Zeybekoğlu, Ş. (2019). PISA 2015 Türkiye Örnekleme Fen Okuryazarlığını Açıklayan Değişkenlerin CHAID Analizi ile İncelenmesi. Antalya, Türkiye.

Zhang, P., Jia, Y. ve Shang, Y. (2022). Research and application of XGBoost in imbalanced data. *International Journal of Distributed Sensor Networks*, 18(6). doi:10.1177/15501329221106935

EKLER

EK-1: Algoritmalarda hiperparametre ayarları yapılan özelliklerin anlamları

Algoritma	Özellikler ve anlamları
Random Forest	<u>min.node.size:</u> Bir düğümün terminal düğüm olması için gereken minimum örnek sayısı. <u>mtry:</u> Her bölünme noktasında dikkate alınacak değişken sayısı <u>splitrule:</u> Ağaç dallarını ayırmak için kullanılacak kural
XGBoost	<u>nrounds:</u> Modelin eğitim için kaç iterasyon alacağını belirtir. <u>max_depth:</u> Karar ağaçlarının maksimum derinliğini belirler. eta: Öğrenme oranı <u>sub_sample:</u> Her yinelemede kullanılacak eğitim veri kümesinin yüzdesini belirtir <u>colsample_bytree:</u> Her ağacın oluşturulmasında kullanılacak özelliklerin yüzdesini belirler <u>min_child_weight:</u> Bir düğümdeki minimum ağırlıklı örnek sayısını belirtir. <u>gamma</u> Bir düğümü bölmek için gereken minimum kayıp azaltımını belirler.
SVR	<u>sigma.</u> Radyal Taban Fonksiyonu için çekirdeğin genişliğini belirler. <u>C:</u> Düzenleştirme parametresi
BRNN	Neurons: Gizli katmandaki nöron sayısı

EK-2: Algoritmelerde kullanılan 99 değişkenin anlamları

ID	Değişken	Değişkenin Anlamı	ID	Değişken	Değişkenin Anlamı
1	ST059Q03TA	Öğrencilerin derse yönelik tutumları	31	ST098Q07TA	Öğrencilerin fen bilimleri dersinde deney yapma sıklığı
2	ANXTEST	Sınav kaygısı düzeyleri	32	ST007Q01TA	Öğrencilerin babalarının en yüksek eğitim seviyeleri
3	ST092Q01TA	Öğrencilerin çevresel farkındalık düzeyleri	33	ST013Q01TA	Öğrencilerin evlerinde bulunan kitap sayısı
4	ENVOPT	Öğrencilerin çevresel iyimserlik düzeyleri	34	ST071Q05N A	Öğrencilerin bir haftalık okul programlarında diğer derslere ayırdıkları haftalık saat sayısı
5	BSMJ	Bilime yönelik meslek hedefi	35	ST076Q10N A	Öğrencilerin bir haftalık okul programlarında diğer derslere ayırdıkları haftalık saat sayısı
6	SMINS	Fen öğrenmeye ayırdıkları süre	36	ST098Q02TA	Fen derslerinde öğretmenlerinin deney yapmalarını ne sıklıkla sağladığı
7	ISCEDD	Ebeveynlerinin eğitim düzeyi	37	MISCED	Öğrencilerin annelerinin en yüksek eğitim seviyesi
8	ST098Q10NA	Okulda karşılaştıkları zorbalık deneyimleri	38	ICTRES	Öğrencilerin evlerinde bulunan bilgi ve iletişim teknolojileri kaynaklarının sayısı ve çeşitliliği
9	EPIST	Epistemolojik inançları	39	TEACHSUP	Öğrencilerin öğretmenlerinden aldıkları destek
10	ST064Q03NA	Ders çalışmak için harcadıkları süre	40	ST129Q08TA	Öğrencilerin asit yağmurlarının oluşumuna dair bilgileri
11	ST092Q02TA	Genetik olarak değiştirilmiş organizmaların kullanımı konusundaki bilgi ve anlayışı	41	ST093Q01TA	Öğrencilerin hava kirliliği konusundaki farkındalık düzeyleri
12	GRADE	Öğrencilerin eğitim seviyeleri	42	ST146Q01TA	Öğrencilerin fen ile ilgili televizyon programlarını izleme sıklıkları
13	DISCLISCI	Fen derslerindeki sınıf disiplini ortamı	43	ST123Q04N A	Öğrencilerin evlerinde internet bağlantısı olup olmadığı
14	ST095Q13NA	Fen bilimleri dersinde deney yapma sıklığı	44	ENVAWARE	Öğrencilerin çevresel sorunlar ve bunların çözümü konusundaki farkındalıkları
15	ST032Q01NA	Fiziksel aktivite yapma sıklığı	45	FISCED	Öğrencilerin babalarının en yüksek eğitim seviyesi
16	ST063Q02NB	Kimya dersine katılım durumları	46	ST063Q01N A	Öğrencilerin bir haftalık okul programlarında diğer derslere ayırdıkları haftalık saat sayısı
17	IBTEACH	Sorgulama temeli öğretim yöntemlerinin kullanım sıklığı	47	ST121Q02N A	Öğrencilerin bir haftalık okul programlarında diğer derslere ayırdıkları haftalık saat sayısı
18	ST076Q02NA	Fen bilimleri derslerinde deney yapma sıklığı	48	ST098Q06TA	Öğretmenlerin fen derslerinde bilimsel fikirlerin farklı fenomenlere nasıl uygulanabileceğini açıklama sıklığı
19	ST071Q03NA	Türkçe ders çalışma süresi	49	EMOSUPS	Öğrencilerin ailelerinden aldıkları duygusal destek
20	ISCEDO	Uluslararası standart eğitim sınıflaması	50	ST146Q05TA	Öğrencilerin bir bilim kulübüne katılma sıklığı
21	SCIEEFF	Fen bilimleri konularında kendilerine olan güvenleri	51	ST103Q01N A	Öğrencilerin fen derslerinde öğretmenlerinin öğrenci performansını nasıl değerlendirdiği
22	ST100Q03TA	Öğretmenlerinin ders sırasında öğrencilere yardım etme sıklığı	52	ST011Q07TA	Öğrencilerin evlerinde klasik edebiyat eserlerine erişim durumu
23	BFMJ2	Öğrencilerin annelerinin eğitim düzeyleri	53	ST093Q05TA	Öğrencilerin su kirliliği konusundaki çevresel iyimserlikleri
24	ST121Q03NA	Öğrencilerin motivasyon düzeyleri	54	ST071Q04N A	Öğrencilerin bir haftalık okul programlarında diğer derslere ayırdıkları haftalık saat
25	ST131Q06NA	Bir önceki eğitim yılında fen bilimleri dersinde deney yapma sıklığı	55	ST082Q03N A	Öğrencilerin Sınıf arkadaşlarının başarılı olmasını görme zevki
26	ST060Q01NA	Bir haftalık okul programlarında toplamda kaç saat ders aldıkları	56	ST071Q02N A	Öğrencilerin bir haftalık okul programlarında Matematik dersine ayırdıkları haftalık saat
27	REPEAT	Öğrencilerin sınıf tekrarı yapıp yapmadığı	57	ST146Q07N A	Öğrencilerin bir bilim kulübüne katılma sıklığı

28	PERFEED	Öğretmenlerin öğrencileri okuma performansları hakkında geri bildirim verme sıklığı	58	ST005Q01TA	Öğrencilerin annelerinin yüksek eğitim seviyesi
29	ST001D01T	Öğrencilerin bir önceki eğitim yılında hangi sınıfta olduklarını belirten soru	59	ST003D02T	Öğrencilerin doğum ayları
30	ST011Q05TA	Öğrencilerin eğitim yazılımlarına erişimi	60	BELONG	Öğrencilerin okula aidiyet hissi

EK-2: Devam ediyor

ID	Değişken	Değişkenin Anlamı	ID	Değişken	Değişkenin Anlamı
61	ESCS	Öğrencilerin sosyoekonomik statüsü	81	ST123Q01NA	Öğrencilerin ailelerinin okuldaki faaliyetlere katılım düzeyi
62	ST076Q08NA	Öğrencilerin okula gitmeden önce ebeveynleri ile konuşma sıklığı	82	ST064Q01NA	Öğrencilerin öğretmenlerinin bilimsel kavramların günlük yaşantımıza olan önemini açıkça açıklayıp açıklamadığı
63	ST098Q09TA	Öğretmenin geniş bilimsel kavramların günlük yaşantımıza olan önemini açıkça açıklayıp açıklamadığı	83	ST039Q02NA	Öğrencilerin öğretmenlerinin kendilerini diğer öğrencilere göre daha sıkı değerlendirme sıklığı
64	ST104Q03NA	Öğrencilerin evlerinde bilgisayar ve internet erişiminin olup olmadığı	84	ST082Q13NA	Öğrencilerin öğretmenlerinin bilimsel kavramların günlük yaşantımıza olan önemini açıkça açıklayıp açıklamadığı
65	ST062Q03TA	Öğretmenlerin bilimsel kavramların günlük yaşantımıza olan önemini açıkça açıklayıp açıklamadığı	85	ST012Q07NA	Öğrencilerin sahip oldukları tablet bilgisayarlar
66	ST034Q06TA	Öğrencilerin öğretmenlerinin bilimsel kavramların günlük yaşantımıza olan önemini açıkça açıklayıp açıklamadığı	86	ST064Q02NA	Öğrencilerin fen bilimleri derslerinde zorluk seviyesini seçme imkanları
67	TMINS	Öğrencilerin haftalık toplam öğrenme süresi	87	ST082Q12NA	Öğrencilerin öğretmenlerinin bilimsel kavramların günlük yaşantımıza olan önemini açıkça açıklayıp açıklamadığı
68	ST078Q05NA	Öğrencilerin öğretmenlerinin bilimsel kavramların günlük yaşantımıza olan önemini açıkça açıklayıp açıklamadığı	88	ST034Q01TA	Öğrencilerin öğretmenlerinin bilimsel kavramların günlük yaşantımıza olan önemini açıkça açıklayıp açıklamadığı
69	ST063Q06NA	Öğrencilerin öğretmenlerinin bilimsel kavramların günlük yaşantımıza olan önemini açıkça açıklayıp açıklamadığı	89	ST097Q03TA	Öğrencilerin öğretmenlerinin bilimsel kavramların günlük yaşantımıza olan önemini açıkça açıklayıp açıklamadığı
70	ST094Q01NA	Öğrencilerin fen bilimleri öğrenme sürecinde ne kadar keyif aldıkları	90	ST129Q01TA	Öğrencilerin sağlık konulu bir gazete raporunun altında yatan fen bilimleri kavramlarını anlama düzeyi
71	HISCED	Öğrencilerin ebeveynlerinin en yüksek eğitim seviyesi	91	ST061Q01NA	Öğrencilerin öğretmenlerinin fen bilimleri derslerinde öğrencilerin günlük yaşamla bağlantı kurmalarına teşvik edip etmedikleri
72	ST095Q07NA	Öğrencilerin fen bilimleri okuryazarlığı kapsamında hareket ve kuvvetler konusundaki bilgi ve becerileri	92	ST078Q08NA	Öğrencilerin öğretmenlerinin fen bilimleri derslerinde öğrencilerin günlük yaşamla bağlantı kurmalarına teşvik edip etmedikleri
73	ST082Q08NA	Öğrencilerin öğretmenlerinin bilimsel kavramların günlük yaşantımıza olan önemini açıkça açıklayıp açıklamadığı	93	ST082Q02NA	Öğrencilerin öğretmenlerinin fen bilimleri derslerinde öğrencilerin günlük yaşamla bağlantı kurmalarını teşvik edip etmedikleri
74	ST100Q01TA	Öğrencilerin öğretmenlerinin bilimsel kavramlarının günlük yaşantımıza olan önemini açıklama ölçütü	94	ST093Q03TA	Öğrencilerin bitki ve hayvanların nesli tükenme sorununu önümüzdeki 20 yıl içinde nasıl gelişeceğini düşündüklerini
75	ST097Q04TA	Öğrencilerin öğretmenlerinin bilimsel kavramların günlük yaşantımıza olan önemi	95	ST063Q04NA	Öğrencilerin öğretmenlerinin fen bilimleri derslerinde öğrencilerin günlük yaşamla bağlantı kurmalarını teşvik edip etmedikleri
76	ST146Q03TA	Öğrencilerin bir bilim kulübüne katılma sıklığı	96	ST098Q01TA	Öğrencilerin öğretmenlerinin fen bilimleri derslerinde öğrencilerin günlük yaşamla bağlantı kurmalarını teşvik edip etmedikleri
77	ST059Q01TA	Öğrencilerin fen bilimleri dersinde öğretmenlerin öğrenci performansını nasıl değerlendirdiği	97	ST104Q05NA	Öğrencilerin öğretmenlerinin fen bilimleri derslerinde öğrencilerin günlük yaşamla bağlantı kurmalarını teşvik edip etmedikleri
78	ST103Q08NA	Öğrencilerin öğretmenlerinin öğrencilerin sorularını tartışma sıklığı	98	ST123Q02NA	Öğrencilerin öğretmenlerinin fen bilimleri derslerinde öğrencilerin günlük yaşamla bağlantı kurmalarını teşvik edip etmedikleri
79	ST097Q01TA	Öğrencilerin öğretmenlerinin bilimsel kavramların günlük yaşantımıza olan önemi	99	ST131Q03NA	Öğrencilerin fen bilimleri alanındaki fikirlerinin zaman içinde değişip değişmediği
80	ST104Q02NA	Öğrencilerin öğretmenlerinin bilimsel kavramların günlük yaşantımıza olan önemini açıkça açıklayıp açıklamadığı			

EK-3: TR1 Bölgesi ve TR3 Bölgesi 1. Grup Benzerlik Seviyesi Yüksek Değişkenler.

TR1 ve TR3 Bölgesi Ortak Değişkenler	Değişkenin Anlamı
ST059Q03TA	Bilimsel Araştırma ve Sorgulama Tutumu
ANXTEST	Kaygı Testi
ST092Q01TA	Bilim Öğrenmeye Yönelik Özgüven
ENVOPT	Çevre Bilinci ve Duyarlılığı
BSMJ	Bilim Alanına İlgi
SMINS	Bilimsel Tutum
ISCEDD	Eğitim Düzeyi (Anne)
EPIST	Bilgiye Dayalı Epistemolojik İnançlar
ST064Q03NA	Bilimsel Okuryazarlık
ST092Q02TA	Bilim Öğrenmeye Yönelik Değer
ST095Q13NA	Bilime Yönelik Olumlu Tutum
ST063Q02NB	Bilimsel Tartışmalara Katılım (Öğrenci Görüşü)
IBTEACH	Öğretmen Desteği
ST076Q02NA	Bilimsel Tartışmalara Katılım (Öğrenci Tutumu)
ST071Q03NA	Bilim Öğrenmeye Yönelik Motivasyon ve İlgi
ISCEDO	Eğitim Düzeyi (Baba)
BFMJ2	Fen Bilimine İlgi
ST131Q06NA	Fen Öğrenmeye Yönelik Çaba/Motivasyon
ST060Q01NA	Bilimsel Bilgiye Yönelik Güven
SCIEEFF	Bilimsel Yeterlilik İnanç
ST098Q07TA	Bilim Öğrenmeye Yönelik Çaba
ST013Q01TA	Bilimsel Konulara Katılım

EK-4: TR1 Bölgesi ve TR4 Bölgesi 1. Grup Benzerlik Seviyesi Yüksek Değişkenler.

TR1 ve TR4 Bölgesi Ortak Değişkenler	Değişkenin Anlamı
ST059Q03TA	Bilimsel Araştırma ve Sorgulama Tutumu
ANXTEST	Kaygı Testi
ST092Q01TA	Bilim Öğrenmeye Yönelik Özgüven
ENVOPT	Çevre Bilinci ve Duyarlılığı
BSMJ	Bilim Alanına İlgi
SMINS	Bilimsel Tutum
ISCEDD	Eğitim Düzeyi (Anne)
ST098Q10NA	Bilim Öğrenmeye Yönelik Motivasyon
EPIST	Bilgiye Dayalı Epistemolojik İnançlar
ST092Q02TA	Bilim Öğrenmeye Yönelik Değer
ST032Q01NA	Okulda Bilimsel Aktiviteler
ST063Q02NB	Bilimsel Tartışmalara Katılım (Öğrenci Görüşü)
ST071Q03NA	Bilim Öğrenmeye Yönelik Motivasyon ve İlgi
ISCEDO	Eğitim Düzeyi (Baba)
ST121Q03NA	Bilimsel Etkinliklere Katılım
ST131Q06NA	Fen Öğrenmeye Yönelik Çaba/Motivasyon
GRADE	Sınıf Düzeyi
ST001D01T	Cinsiyet
ST013Q01TA	Bilimsel Konulara Katılım
ST011Q05TA	Bilim Dersinde Öğrenmeye Yönelik Çaba
ST098Q07TA	Bilim Öğrenmeye Yönelik Çaba
ST059Q03TA	Bilimsel Araştırma ve Sorgulama Tutumu
ANXTEST	Kaygı Testi
ST092Q01TA	Bilim Öğrenmeye Yönelik Özgüven

EK-5: TR1 Bölgesi ve TR6 Bölgesi 1. Grup Benzerlik Seviyesi Yüksek Değişkenler.

TR1 ve TR6 Bölgesi Ortak Değişkenler	Değişkenin Anlamı
ST059Q03TA	Bilimsel Araştırma ve Sorgulama Tutumu
ANXTEST	Kaygı Testi
ST092Q01TA	Bilim Öğrenmeye Yönelik Özgüven
ENVOPT	Çevre Bilinci ve Duyarlılığı
BSMJ	Bilim Alanına İlgi
SMINS	Bilimsel Tutum
ISCEDD	Eğitim Düzeyi (Anne)
ST098Q10NA	Bilim Öğrenmeye Yönelik Motivasyon
EPIST	Bilgiye Dayalı Epistemolojik İnançlar
ST064Q03NA	Bilimsel Okuryazarlık
ST092Q02TA	Bilim Öğrenmeye Yönelik Değer
ST095Q13NA	Bilime Yönelik Olumlu Tutum
IBTEACH	Öğretmen Desteği
ST076Q02NA	Bilimsel Tartışmalara Katılım (Öğrenci Tutumu)
ISCEDO	Eğitim Düzeyi (Baba)
SCIEEFF	Bilimsel Yeterlilik İnanç
BFMJ2	Fen Bilimine İlgi
ST121Q03NA	Bilimsel Etkinliklere Katılım
ST098Q07TA	Bilim Öğrenmeye Yönelik Çaba
ST013Q01TA	Bilimsel Konulara Katılım
PERFEED	Akran Geri Bildirimi

EK-6: TR2 Bölgesi ve TR6 Bölgesi 1. Grup Benzerlik Seviyesi Yüksek Değişkenler.

TR2 ve TR6 Bölgesi Ortak Değişkenler	Değişkenin Anlamı
ISCEDD	Eğitim Düzeyi (Anne)
ST092Q01TA	Bilim Öğrenmeye Yönelik Özgüven
ST013Q01TA	Bilimsel Konulara Katılım
ENVOPT	Çevre Bilinci ve Duyarlılığı
ESCS	Ekonomik Sosyal ve Kültürel Statü
ANXTEST	Kaygı Testi
EPIST	Bilgiye Dayalı Epistemolojik İnançlar
ISCEDO	Eğitim Düzeyi (Baba)
ST098Q07TA	Bilim Öğrenmeye Yönelik Çaba
ICTRES	Bilgisayar Kaynaklarına Erişim
ST076Q10NA	Bilimsel Tartışmalara Katılım (Öğretmen Görüşü)
SCIEEFF	Bilimsel Yeterlilik İnanç
ST071Q04NA	Bilim Öğrenmeye Yönelik Zorluk Algısı
ST121Q03NA	Bilimsel Etkinliklere Katılım
ST059Q03TA	Bilimsel Araştırma ve Sorgulama Tutumu
ST064Q03NA	Bilimsel Okuryazarlık
ST092Q02TA	Bilim Öğrenmeye Yönelik Değer
BFMJ2	Fen Bilimine İlgi
ST098Q02TA	Bilim Öğrenmeye Yönelik İlgisizlik
SMINS	Bilimsel Tutum
ST064Q02NA	Bilimsel Okuryazarlık

EK-7: TR1 Bölgesi ve TR6 Bölgesi 2. Grup Benzerlik Seviyesi Orta Seviyedeki Değişkenler.

TR1 ve TR4 Bölgesi Ortak Değişkenler	Değişkenin Anlamı
ST071Q05NA	Bilim Öğrenmeye Yönelik Çalışma Yöntemleri
ST146Q05TA	Bilimsel Aktivitelerde Yeterlilik Algısı
ST121Q02NA	Bilimsel Etkinliklere Katılım
ST098Q09TA	Bilim Öğrenmeye Yönelik İlgi
ST098Q06TA	Bilimsel Tartışmalarda Aktiflik
ENVAWARE	Çevre Bilinci
ST003D02T	Anne Çalışma Durumu
ICTRES	Bilgisayar Kaynaklarına Erişim
ST146Q07NA	Bilimsel Konulara İlgi ve Merak
ST146Q01TA	Bilimsel Araştırma ve Sorgulama Becerileri
TEACHSUP	Öğretmen Desteği
MISCED	Eğitim Düzeyi (Veli)
ST063Q01NA	Bilimsel Kaynaklara Erişim (Öğrenci Görüşü)
ST005Q01TA	Bilimsel Kavramları Anlama
ST129Q08TA	Bilim Öğrenmeye Yönelik Zorluklarla Başa Çıkma
ST093Q05TA	Bilim Öğrenmeye Yönelik Katılım
ST103Q01NA	Bilimsel Kaynaklara Erişim

EK-8 TR7 Bölgesi ve TR8 Bölgesi 2. Grup Benzerlik Seviyesi Orta Seviyedeki Değişkenler.

TR1 ve TR4 Bölgesi Ortak Değişkenler	Değişkenin Anlamı
ST071Q05NA	Bilim Öğrenmeye Yönelik Çalışma Yöntemleri
BELONG	Okula Aidiyet Hissi
ST003D02T	Anne Çalışma Durumu
ST103Q01NA	Bilimsel Kaynaklara Erişim
ST007Q01TA	Bilimsel Düşünme Becerileri
MISCED	Eğitim Düzeyi (Veli)
ST011Q05TA	Bilim Dersinde Öğrenmeye Yönelik Çaba
ST034Q01TA	Okulda Bilimle İlgili Tutum
ST098Q02TA	Bilim Öğrenmeye Yönelik İlgisizlik
SMINS	Bilimsel Tutum
ST064Q02NA	Bilimsel Okuryazarlık
EMOSUPS	Duygusal Destek
ST064Q03NA	Bilimsel Okuryazarlık
ST082Q13NA	Bilim Öğrenmeye Yönelik İlgisizlik ve Umutsuzluk
ST103Q08NA	Bilimsel Kaynaklara Ulaşım İmkânı
ST082Q08NA	Bilimsel Çalışmalara Yönelik Motivasyon
ST011Q07TA	Bilim Dersinde Öğrenmeye Yönelik Motivasyon

EK-9: TR6 Bölgesi ve TR8 Bölgesi 3. Grup Benzerlik Seviyesi Yüksek Seviyedeki Değişkenler.

TR6 ve TR8 Bölgesi Ortak Değişkenler	Değişkenin Anlamı
ST123Q02NA	Bilimsel Kavramları Öğrenmede Zorluk
ST146Q01TA	Bilimsel Araştırma ve Sorgulama Becerileri
ST005Q01TA	Bilimsel Kavramları Anlama
ST129Q08TA	Bilim Öğrenmeye Yönelik Zorluklarla Başa Çıkma
ST121Q02NA	Bilimsel Etkinliklere Katılım
ST131Q03NA	Fen Öğrenimine Yönelik Motivasyon ve Çaba
ST078Q05NA	Bilimsel Çalışmalardan Keyif Alma
ST100Q01TA	Bilimsel Çalışmalara Yönelik İlgi
ST097Q03TA	Bilimsel Konulara Yönelik İlgi ve Merak
ST034Q06TA	Bilimsel Konuları Anlamada Zorlanma
FISCED	Eğitim Düzeyi (Baba)
ST098Q06TA	Bilimsel Tartışmalarda Aktiflik
ST063Q06NA	Bilimsel Konuları Öğrenmeye Yatkınlık
ST146Q07NA	Bilimsel Konulara İlgi ve Merak
ST129Q01TA	Bilimsel Çalışmalara Yönelik İstek
ST060Q01NA	Bilimsel Bilgiye Yönelik Güven
ST104Q05NA	Bilimsel Bilgiye Yönelik İlgi
ST103Q01NA	Bilimsel Kaynaklara Erişim
ST093Q03TA	Bilimsel Konulara Katılım
HISCED	Ebeveynin Eğitim Düzeyi (En Yüksek)
ST076Q08NA	Bilimsel Tartışmalara Katılım
ST123Q01NA	Bilimsel Konuları Öğrenmede Kolaylık

EK-10 TR1 Bölgesi ve TR8 Bölgesi 3. Grup Benzerlik Seviyesi Yüksek Seviyedeki Değişkenler.

TR1 ve TR4 Bölgesi Ortak Değişkenler	Değişkenin Anlamı
ST078Q05NA	Belirli bir konu hakkındaki tutum, inanç veya bilgi düzeyi.
ST063Q06NA	Belirli bir durum veya davranışa yönelik algı veya değerlendirme.
ST100Q01TA	Bir konuya yatkınlık
ST059Q01TA	Öğrencinin bir konudaki bilgi düzeyi veya becerisi
ST097Q01TA	Öğrencinin belirli bir konudaki performansına dair öz değerlendirme.
ST104Q02NA	Öğrencinin öğrenme motivasyonu veya belirli bir konuya karşı hissettiği ilgi.
ST123Q01NA	Öğrencinin okul veya eğitimle ilgili genel düşünceleri, duyguları.
ST064Q01NA	Öğrencinin belirli bir konuda kendine güveni veya yetkinlik algısı.
ST039Q02NA	Öğrencinin belirli bir sosyal veya duygusal konuya yaklaşımı
ST082Q12NA	Öğrencinin belirli bir konu hakkındaki bilgi düzeyi.
ST034Q01TA	Öğrencinin bir öğrenme aktivitesine katılım düzeyi.
ST097Q03TA	Öğrencinin belirli bir konuda sahip olduğu yetkinlik algısı.
ST129Q01TA	Öğrencinin öğrenme stratejilerini kullanım sıklığı veya farkındalığı.
ST061Q01NA	Öğrencinin belirli bir konu veya beceri hakkındaki bilgi düzeyi.
ST063Q04NA	Öğrencinin belirli bir öğrenme ortamıyla ilgili algısı veya deneyimi.
ST098Q01TA	Öğrencinin bir konuyu ne kadar anladığına dair kendi değerlendirmesi.
ST104Q05NA	Öğrencinin bir öğrenme aktivitesine katılımından aldığı keyif veya ilgi.
ST123Q02NA	Öğrencinin okul ortamı veya öğretmenlerle ilgili düşünceleri.
ST131Q03NA	Öğrencinin öğrenme süreciyle ilgili farkındalığı veya öz değerlendirmesi.
ST095Q07NA	Öğrencinin bir konuda bilgi veya beceriye ne kadar sahip olduğunu değerlendirmesi

ÖZGEÇMİŞ

Adı Soyadı:	Büşra YILDIZ			
Derece	Okulun Adı	Program	Yer	Yıl
İlköğretim				
Ortaöğretim				
Lisans				
Yüksek Lisans				
İş Deneyimi				
Referans Kişiler				
E-posta				