



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

REGRESYON AĞAÇLARIYLA SÜREÇ VE SONUÇ VERİLERİNİN KARMAŞIK PROBLEM
ÇÖZME BECERİSİNİ YORDAMA DÜZEYLERİNİN İNCELENMESİ

Fatma Nur AYDIN

Doktora Tezi

Ankara, 2024



Liderlik, araştırma, inovasyon, kaliteli eğitim ve değişim ile

Daha ileriye... En İyiyeye...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

REGRESYON AĞAÇLARIYLA SÜREÇ VE SONUÇ VERİLERİNİN KARMAŞIK PROBLEM
ÇÖZME BECERİSİNİ YORDAMA DÜZEYLERİNİN İNCELENMESİ

INVESTIGATION OF THE LEVELS OF PROCESS AND RESULT DATA TO PREDICTION
COMPLEX PROBLEM SOLVING WITH REGRESSION TREES

Fatma Nur AYDIN

Doktora Tezi

Ankara, 2024

Kabul ve Onay

Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Fatma Nur AYDIN'ın hazırladığı “Regresyon Ağaçlarıyla Süreç ve Sonuç Verilerinin Karmaşık Problem Çözme Becerisini Yordama Düzeylerinin İncelenmesi” başlıklı bu çalışma jürimiz tarafından **Eğitim Bilimleri Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiştir.

Jüri Başkanı	Prof. Dr. Hülya KELEÇİOĞLU	İmza
--------------	----------------------------	------

Jüri Üyesi (Danışman)	Doç. Dr. Kübra ATALAY KABASAKAL	İmza
-----------------------	---------------------------------	------

Jüri Üyesi	Prof. Dr. Burcu ATAR	İmza
------------	----------------------	------

Jüri Üyesi	Doç. Dr. Beyza AKSU DÜNYA	İmza
------------	---------------------------	------

Jüri Üyesi	Prof. Dr. Dilara BAKAN KALAYCIOĞLU	İmza
------------	------------------------------------	------

İkinci Tez Danışmanı	Dr. İsmail DİLEK	Enstitü Yönetim Kurulunun 18/04/2023 Tarihli ve 2023- 17/03 sayılı kararı.
----------------------	------------------	--

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim, Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından / / tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca / / tarihi itibarıyla kabul edilmiştir.

Prof. Dr. İsmail Hakkı MİRİCİ

Eğitim Bilimleri Enstitüsü Müdürü

Öz

Bu çalışmada süreç ve sonuç verilerinin farklı makine öğrenme algoritmalarıyla karmaşık problem çözme becerisini yordama performansları incelenmiştir. Bu amaçla Uluslararası Öğrenci Değerlendirme Programı (PISA) 2012'ye katılmış 915 kişinin verileri kullanılmıştır. Süreç verileri, PISA 2012'deki problem çözme değerlendirmesinde bulunan klima kontrolü ünitesinin birinci sorusuna ait günlük dosyasından elde edilmiştir. Sonuç verileri olarak aynı değerlendirmedeki çeşitli bilişsel ve duyuşsal özellikler kullanılmıştır. Analizler için tekli regresyon ağacı, torbalama, rastgele orman, gradyan artırma, koşullu çıkarım ağacı, koşullu çıkarım ağaçlarına dayalı torbalama, koşullu çıkarım ağaçlarına dayalı rastgele orman ve koşullu çıkarım ağaçlarına dayalı gradyan artırma algoritmaları kullanılmıştır. Elde edilen temel sonuçlara göre (1) süreç verisi orta, sonuç verisi orta-iyi, süreç+sonuç verisi iyi düzeyde tahmin performansı sergilemiştir. (2) Süreç verisinde koşullu çıkarım ağaçlarına dayalı gradyan artırma, sonuç verisinde hem gradyan artırma hem de koşullu çıkarım ağaçlarına dayalı gradyan artırma, süreç+sonuç verisinde gradyan artırma daha iyi performans göstermiştir. Bununla birlikte tüm yöntemlerin metrik değerleri arasında önemli farklılıklar bulunmamaktadır. (3) Tüm algoritmalar için genel olarak; süreç verisinde en önemli değişkenler VOTAT stratejisi puanı ve toplam süre, sonuç verisinde matematik puanı ve okuduğunu anlama puanı, süreç+sonuç verisinde matematik puanı ve VOTAT stratejisi puanı olmuştur. Matematik puanının oldukça baskın biçimde etkili olması dikkat çekmiştir.

Anahtar sözcükler: süreç verisi, günlük dosyası, sonuç verisi, karmaşık problem çözme, makine öğrenme algoritmaları

Abstract

The present study examines the predictive capacity of process and result data in complex problem-solving skills, employing different machine learning algorithms. The data set comprises 915 participants from the Programme for International Student Assessment (PISA) 2012. The process data was obtained from the log file of the first question of the Climate Control unit in the problem-solving assessment in PISA 2012. Various cognitive and affective characteristics were used as result data. The following machine learning algorithms were employed: single regression tree, bagging, random forest, gradient boosting, conditional inference tree, bagging based on conditional inference trees, random forest based on conditional inference trees, and boosting algorithms based on conditional inference trees. The main results indicated that process data exhibited moderate prediction performance, result data exhibited moderate-good prediction performance, and process+result data exhibited good prediction performance. (2) Gradient boosting based on conditional inference trees performed better on process data, both gradient boosting and gradient boosting based on conditional inference trees performed better on result data, and gradient boosting performed better on process+result data. However, there were no important differences between the metric values of all methods. (3) In general, the most important variables in the process data were the VOTAT strategy score and total time, in the outcome/result data the mathematics literacy score and reading literacy score, and in the process+result data the mathematics literacy score and VOTAT strategy score. It is notable that the mathematics literacy score was particularly effective.

Keywords: process data, log file, result data, complex problem solving, machine learning algorithms

Teşekkür

Yollarımız kesiştiği için kendimi şanslı hissettiğim, tezimin her aşamasında bana gösterdikleri anlayış ve destek için minnettar olduğum sevgili danışmanlarım Doç. Dr. Kübra ATALAY KABASAKAL ve Dr. İsmail DİLEK'e sonsuz teşekkürlerimi sunuyorum. Yoğun programlarına rağmen her zaman ulaşılabilir olup, karşılaştığım zorluklarda yol gösterici olmaları bu çalışmayı tamamlamamda büyük bir rol oynadı. İyi ki varsınız...

İçinde bulunduğum her anında beni çok mutlu eden ölçme ve değerlendirme alanıyla tanışmama vesile olan çok kıymetli lisans danışmanım Prof. Dr. Mustafa Zafer BALBAĞ'a çok çok teşekkür ederim.

Doktora tez sürecimde ve jürimde bulunmayı kabul ederek, kıymetli fikir ve önerileriyle çalışmamın zenginleşmesine katkıda bulunan Prof. Dr. Burcu ATAR, Doç. Dr. Beyza AKSU DÜNYA, Prof. Dr. Hülya KELECİOĞLU, Prof. Dr. Dilara BAKAN KALAYCIOĞLU'na çok teşekkür ederim.

Doktora ders döneminde akademik gelişimime katkıda bulunmuş, üzerimizde emeği olan tüm hocalarımıza çok teşekkür ederim.

Bugünlere gelmemde büyük emekleri olan sevgili aileme sonsuz teşekkür ederim.

Doktora öğrenimim boyunca yurt içi doktora burs olanağı sağlayan TÜBİTAK'a teşekkür ederim.

İçindekiler

Kabul ve Onay.....	ii
Öz.....	iii
Abstract.....	iv
Teşekkür.....	v
Tablolar Dizini.....	viii
Şekiller Dizini.....	ix
Simgeler ve Kısaltmalar Dizini.....	xi
Bölüm 1 Giriş.....	1
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	15
Karmaşık Problem Çözme Becerisi.....	15
Karmaşık Problem Çözme Becerisinin ve Süreç Verisinin Kullanıldığı Araştırmalar	29
Süreç ve Sonuç Verilerinin Farklı Yapılarla ve Bağlamlarda Kullanıldığı Araştırmalar	40
Veri Madenciliği (Makine Öğrenme Algoritmaları).....	43
Sınıflandırma ve Regresyon Ağaçları.....	45
Regresyon Ağaçlarının Temel Elemanları.....	47
Regresyon Ağaçlarının Oluşturulmasındaki İstatistiksel Temeller.....	48
Regresyon Ağaçlarının Kullanıldığı Araştırmalar	55
Topluluk Öğrenme Yöntemleri	57
Torbalama (Bagging)	58
Rastgele Orman.....	59
Gradyan Artırma (Gradient Boosting).....	63
Topluluk Öğrenme Yöntemlerinin Kullanıldığı Araştırmalar.....	68
Koşullu Çıkarım Ağaçları.....	75
Koşullu Çıkarım Ağaçlarına Dayalı Torbalama ve Rastgele Orman.....	80
Koşullu Çıkarım Ağaçlarına Dayalı Gradyan Artırma	83

Koşullu Çıkarım Ağaçlarına Dayalı Algoritmaların Kullanıldığı Araştırmalar	84
Bölüm 3 Yöntem	92
Araştırmanın Türü	92
Araştırmanın Evreni ve Örneklemi	92
Verilerin Elde Edilmesi	93
Bağımlı Değişkenin Elde Edilmesi.....	93
Süreç ve Sonuç Verisi Değişkenleri	96
Verilerin Analizi	100
Bölüm 4 Bulgular, Yorumlar ve Tartışma.....	110
Bölüm 5 Sonuç ve Öneriler.....	145
Kaynaklar	151
EK-A: Değişkenlere İlişkin Betimsel İstatistikler.....	173
EK-B: Değişkenlere İlişkin Dağılım Grafikleri ve Korelasyon Değerleri	174
EK-C: Madde Parametre Kestirimlerine İlişkin Değerler	176
EK-D: Süreç Verisinden Elde Edilen Regresyon Ağaçları	177
EK-E: Sonuç Verisinden Elde Edilen Regresyon Ağaçları	178
EK-F: Süreç+Sonuç Verisinden Elde Edilen Regresyon Ağaçları	179
EK-G: Süreç Verisi-Değişken Önem Grafikleri	180
EK-H: Sonuç Verisi-Değişken Önem Grafikleri.....	184
EK-I: Süreç+Sonuç Verisi-Değişken Önem Grafikleri.....	188
EK-J: Araştırma Etik Komisyon İzin Muafiyeti Formu	192
EK-K: Etik Beyanı	193
EK-L: Doktora Tez Çalışması Orijinallik Raporu	194
EK-M: Dissertation Originality Report.....	195
EK-N: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı.....	196

Tablolar Dizini

Tablo 1 <i>Bağımlı Değişkenin Hesaplanmasında Kullanılan Maddeler</i>	93
Tablo 2 <i>Model-Veri Uyumı Sonuçları</i>	94
Tablo 3 <i>GKPM için Model-Veri Uyumı Sonuçları</i>	95
Tablo 4 <i>Yetenek Kestirimlerine İlişkin Güvenirlik Değerleri</i>	95
Tablo 5 <i>Oluşturulan Süreç Verisi Değişkenleri</i>	98
Tablo 6 <i>Azim ve Problem Çözmeye Açıklık ile İlgili Anket Soruları</i>	100
Tablo 7 <i>Kullanılan Paketler</i>	101
Tablo 8 <i>Farklı Örneklem Büyüklüğü Koşullarına Göre Elde Edilen Regresyon Ağaçlarına İlişkin Metrik Değerleri</i>	105
Tablo 9 <i>Topluluk Öğrenme Algoritmaları İçin En Uygun Hiper-parametre Değerleri</i>	107
Tablo 10 <i>Koşullu Çıkarım Ağaçlarına Dayalı Algoritmalar İçin En Uygun Hiper-parametre Değerleri</i>	108
Tablo 11 <i>Süreç Verisi-Makine Öğrenme Algoritmaları Metrik Değerleri</i>	110
Tablo 12 <i>Sonuç Verisi-Makine Öğrenme Algoritmaları Metrik Değerleri</i>	113
Tablo 13 <i>Süreç+Sonuç Verisi-Makine Öğrenme Algoritmaları Metrik Değerleri</i>	116
Tablo 14 <i>Süreç Verisi-Değişken Önem Sıralamaları</i>	124
Tablo 15 <i>Sonuç Verisi-Değişken Önem Sıralamaları</i>	129
Tablo 16 <i>Süreç+Sonuç Verisi-Değişken Önem Sıralamaları</i>	131

Şekiller Dizini

Şekil 1 <i>Klima Kontrolü Ünitesi Bilgilendirme Ekranı (OECD, 2014a, s. 37)</i>	20
Şekil 2 <i>Klima Kontrolü Ünitesi Birinci Sorusu (OECD, 2014a, s. 38)</i>	21
Şekil 3 <i>Klima Kontrolü Ünitesi İkinci Sorusu (OECD, 2014a, s. 38)</i>	21
Şekil 4 <i>Örnek Bir Ağaç (Sol: İki boyutlu örnek üzerinde özyinelemeli ikili bölünme çıktısı. Sağ: Soldaki bölünmeye karşılık gelen ağaç)</i>	47
Şekil 5 <i>Permütasyon Önemi Şeması (Sol: koşulsuz permütasyon önemi. Sağ: koşullu permütasyon önemi)</i>	82
Şekil 6 <i>Günlük Dosyasından Örnek Kesit</i>	97
Şekil 7 <i>Eksik Veri Tamamlama İşlemlerine Göre Elde Edilen Histogramlar</i>	102
Şekil 8 <i>Süreç Verisinden Elde Edilen Regresyon Ağacı</i>	111
Şekil 9 <i>Süreç Verisinden Elde Edilen Koşullu Çıkarım Ağacı</i>	112
Şekil 10 <i>Sonuç Verisinden Elde Edilen Regresyon Ağacı</i>	114
Şekil 11 <i>Sonuç Verisinden Elde Edilen Koşullu Çıkarım Ağacı</i>	115
Şekil 12 <i>Süreç+Sonuç Verisinden Elde Edilen Regresyon Ağacı</i>	116
Şekil 13 <i>Süreç+Sonuç Verisinden Elde Edilen Koşullu Çıkarım Ağacı</i>	118
Şekil 14 <i>Süreç Verisi-Kısmi Bağımlılık Grafikleri</i>	125
Şekil 15 <i>Sonuç Verisi-Kısmi Bağımlılık Grafikleri</i>	129
Şekil 16 <i>Süreç+Sonuç Verisi-Kısmi Bağımlılık Grafikleri</i>	132
Şekil 17 <i>Süreç Verisi-Rastgele Orman Değişken Önemi (Sol: Ölçeklenmiş. Sağ: Ölçeklenmemiş)</i>	134
Şekil 18 <i>Sonuç Verisi-Rastgele Orman Değişken Önemi (Sol: Ölçeklenmiş. Sağ: Ölçeklenmemiş)</i>	135
Şekil 19 <i>Süreç+Sonuç Verisi-Rastgele Orman Değişken Önemi (Sol: Ölçeklenmiş. Sağ: Ölçeklenmemiş)</i>	135
Şekil 20 <i>Süreç Verisi-Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman Değişken Önem Grafiği (Koşullu Olmayan Versiyon)</i>	137
Şekil 21 <i>Süreç Verisi-Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman Değişken Önem Grafiği (Koşullu Versiyon)</i>	138
Şekil 22 <i>Sonuç Verisi-Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman Değişken Önem Grafiği (Koşullu Olmayan Versiyon)</i>	140
Şekil 23 <i>Sonuç Verisi-Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman Değişken Önem Grafiği (Koşullu Versiyon)</i>	140
Şekil 24 <i>Süreç+Sonuç Verisi-Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman Değişken Önem Grafiği (Koşullu Olmayan Versiyon)</i>	142

Şekil 25 Süreç+Sonuç Verisi-Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman Değişken Önem Grafiği (Koşullu Versiyon).....	143
--	-----



Simgeler ve Kısaltmalar Dizini

KPÇ: Karmaşık Problem Çözme

PISA: Programme for International Student Assessment



Bölüm 1

Giriş

Bu bölümde araştırmayla ilgili problem durumu, amaç, önem, alt problemler, sayıltılar, sınırlılıklar ve tanımlara yer verilmiştir.

Problem Durumu

Ölçme ve değerlendirme uygulamalarının birincil amacı, bireylerin ölçülmesi hedeflenen örtük özelliğe ilişkin yeterliklerinin geçerli ve güvenilir bir şekilde ortaya koyulmasıdır (Zhang ve ark., 2023). Bu amaç doğrultusunda maddelerden elde edilen puanlara dayalı olarak çıkarımlar yapmak yaygındır. Bireylerin madde yanıtlarından elde edilen veriler sonuç veya ürün verileri olarak adlandırılabilir (Goldhammer ve ark., 2020). Bu bağlamda eldeki çalışmada yanıt doğruluğu, seçilen tepki kategorisi ve bunlara dayalı gerçekleştirilen yetenek kestirimleri “sonuç verisi” olarak ele alınmıştır. Kullanımları oldukça yaygın olan sonuç verileri pek çok ölçme-değerlendirme problemine geçerli ve güvenilir çözüm sunmuştur. Ancak bu veriler görevin ne kadar iyi tamamlandığına ilişkin önemli bilgiler içerirken (Zhang ve ark., 2023) performansın ortaya çıkma biçimindeki davranışsal farklılıklara dair bilgi vermekte bazı sınırlılıklara sahiptir (Goldhammer ve ark., 2017).

Teknolojik gelişmeler sayesinde bir problemin çözümünde görevin doğruluğuna ilişkin bilgilerin yanı sıra, çözüm esnasındaki davranışsal sürecin çeşitli yönleri de kaydedilebilmektedir (Engelhardt ve Goldhammer, 2019). Bu sebeple popüler hale gelen bilgisayar tabanlı değerlendirmeler, problem çözme gibi karmaşık süreçleri içeren üst düzey becerilerin ölçülmesinde sıklıkla kullanılmaya başlamıştır (Tang ve ark., 2021). Bilgisayar tabanlı değerlendirmelerde bireylerin görevleri çözüp çözemediklerine dair verilerin yanında (Liao ve ark., 2019) görevlerin yanıtlanması boyunca gerçekleştirilen işlemler de kaydedilir (Liu ve ark., 2018). Simüle edilmiş ortamlarda sunulan gerçek hayatla ilişkilendirilmiş görevlere yanıt verilmesini gerektiren bu değerlendirmeler sırasında bireyler bilgisayarla etkileşime girer (Tang ve ark., 2021). Bu etkileşim sırasında yapılan tüm eylemler otomatik

olarak yakalanabilir ve bu sayede zaman damgalarıyla birlikte eylemlerin kaydedildiği günlük dosyaları oluşur (Han ve Wilson, 2022; Jiao ve ark., 2021). Bu dosyalar sayesinde bireylerin test platformuyla etkileşime girme biçimleri izlenebilir (Reis Costa ve ark., 2021). Buna göre, bireylerin problemleri ne kadar iyi çözdüklerine ek olarak nasıl çözdüklerine ilişkin daha ayrıntılı bilgi veren ve problem çözme süreçlerini yansıtan, bilgisayar tabanlı değerlendirmelerde kaydedilen insan-insan veya insan-bilgisayar etkileşimi verilerine “süreç verisi” denir (Fu, Chen ve ark., 2023; Zhan ve Qiao, 2022). Bu veriler yanıt doğruluğuna göre oluşturulmuş verilere ek olarak yardımcı, daha ayrıntılı ve değerli bilgiler sağlama potansiyeline sahiptir (Tang ve ark., 2021; Jiao ve ark., 2021). Bireylerin problem çözme esnasındaki faaliyetlerinin dışı vurumu olan bu veriler, problem çözme stratejileri ve bireylerin yetenekleri hakkında zengin bilgiler içerir (Han ve Wilson, 2022). Süreç verileri tanımlayıcı veya biçimlendirici amaçlarla kullanılabileceği gibi, düzey belirleyici (summative) değerlendirmeler için de bilgi kaynağı olabilir (de Klerk ve ark., 2015). Daha kapsamlı bir bakış açısından bilişsel testlerde süreç verileri Goldhammer ve Zehner (2017) tarafından “ölçülen yapı/yapıların görev ürünü (yani madde puanı) üzerindeki etkisine aracılık eden bilişsel (aynı zamanda meta-bilişsel, motivasyonel ve duyuşsal) durumlar ve ilgili davranışlar hakkındaki ampirik bilgi” olarak tanımlanmıştır. Bu bağlamda süreç verilerinin operasyonel tanımı, bir test maddesi üzerindeki çalışma sürecini yansıtan ampirik veriler olarak ifade edilebilir (Goldhammer ve Zehner, 2017). Süreç verisi kaynakları olarak günlük dosyaları, eş zamanlı sesli düşünme protokolleri, video ve ses içerikli ekran kayıtları, göz izleme, yüz ifadelerinin kaydı, fizyolojik sensör (EDA, EMG, EKG, EEG, fMRI, fNIR)¹ verileri sayılabilir (Azevedo, 2015).

Bireylerin problem çözerken gerçekleştirdikleri yanıt davranışlarının altında yatan bilişsel mekanizmaların ortaya konulmasının (Greiff, Wüstenberg ve ark., 2015) bir anlamda problem çözme esnasında işletilen zihinsel sürecin haritalandırılması olarak görülebileceği düşünülmektedir. Yanıt doğruluğu gibi sonuç göstergelerine dayalı gerçekleştirilen yeterlik

¹ EDA: Electrodermal activity, EMG: Electromyogram, EKG: Electrocardiogram, EEG: Electroencephalography, fMRI: Functional magnetic resonance, fNIR: Functional Near-Infrared

kestirimlerinin geçerliğine katkı sunmak amaçlı bilgi sağlayabilecek süreç verileri aynı zamanda gerçek zamanlı geri bildirim sağlayan etkileşimli simülasyon tabanlı test sistemlerinin geliştirilmesi, eksik verilerin nedenlerinin anlaşılması, yapılabilecek ek öğretim uygulamalarının içeriğinin daha isabetli hale getirilmesi gibi konularda tamamlayıcı bilgiler sağlayacaktır (Goldhammer ve ark., 2020; Greiff, Wüstenberg ve ark., 2015; Jiao ve ark., 2021; Teig ve ark., 2020). Buna göre süreç verileri, bireyin yanıt süreçlerine ilişkin yeni bakış açıları sağlayabileceği gibi yetenek kestirimlerinin kesinliğini artırabilir (Klotzke ve Fox, 2019) ve yetkinliklerin teorik modellerini geliştirebilir (Goldhammer ve ark., 2017). Örneğin, Reis Costa ve arkadaşları (2021) çalışmasında yanıt sürelerinin modele dahil edilmesiyle yetenek tahmini kesinliğinin arttığını raporlamışlardır. Başka bir araştırmada çeşitli süreç verilerinden elde edilen bilgilerle yetenek kestirimi yapıldığı durumda, madde tepki kuramına dayalı yapılan kestirime göre daha yüksek güvenirlikte tahminler elde edilmiştir (Zhang ve ark., 2023). Benzer bir bulgu Scalise ve Clarke Midura (2018) tarafından da raporlanmıştır. Araştırmacılar çeşitli süreç verilerini dahil ederek çok boyutlu madde tepki kuramı (MTK) ve Bayes ağlarını içeren hibrit bir modelleme yaklaşımı çok boyutlu MTK-Bayes aracılığıyla yaptıkları kestirimlerde, ölçülen yapının her iki boyutu için de süreç verisi içermeyen modellere göre daha yüksek güvenirlik elde edildiğini ifade etmişlerdir. Wang ve arkadaşları (2019) tarafından yapılan bir çalışmada yanıt süresinin modele dahil edilmesiyle yetenek tahminlerinin standart hatalarının azaldığı bulunmuştur. Son olarak Tang ve arkadaşları (2020) çalışmasında süreç verisinden elde edilen özelliklerle yapılan yetenek kestiriminin, yanıt doğruluğuna göre yapılandan daha iyi performans sergilediğini bulmuşlardır. Öte yandan süreç verilerinden elde edilen bilgilerle ölçülmesi hedeflenen becerinin yorumlanmasına ilişkin çeşitli geçerlik çıkarımları yapmak da mümkündür (De Boeck ve Jeon, 2019). Örneğin bu bilgilerle, bireyin eylemlerinin uygun olmayan yanıt davranışlarıyla örtüşme düzeyi belirlenebilir ve buna göre uyarı veren bir test sisteminin geliştirilmesi söz konusu olabilir (Goldhammer ve ark., 2020). Öte yandan bu veriler, bireylerin çözüme ilişkin stratejilerinin anlaşılmasına imkân sağlar (Liu ve ark., 2018). Bu stratejilerin anlaşılmasıyla daha gelişmiş ek öğretim içerikleri oluşturulabilir. Çünkü problem

çözerken bir stratejiyi uygulayıp başarısız olan ve hiç uygulamayıp başarısız olan öğrencinin farklı desteğe ihtiyacı olacaktır (Greiff, Wüstenberg ve ark., 2015). Teig ve arkadaşları (2020) tarafından ifade edildiği üzere, bireyler bir görevde benzer puan alsalar bile süreçte aynı şekilde performans göstermemiş olabilirler. Bazı öğrenciler doğru stratejiyi uygulayıp çözüme ulaşırken bazıları deneme yanılma yoluyla da doğru sonuca ulaşmış olabilir. Ya da bazı öğrenciler hiçbir strateji uygulamadan rastgele doğru sonuca ulaşmış olabilir. Benzer şekilde, düşük puan alan öğrencilerden bir kısmı süreçte belli bir noktaya kadar doğru stratejiyi uygulamış fakat çözüme ulaşamamışken bazıları hiçbir adımda doğru stratejiyi uygulamayarak düşük puan almış olabilir. Dolayısıyla süreç verileri yöntemlerde, modellerde, teorilerde ve öğretimsel içeriklerin daha etkili hale getirilmesi konularında ilerlemelere yol açma potansiyeline sahiptir (Azevedo, 2015). Bu bağlamda çeşitli süreç verileri kullanılarak geçerlik yaklaşımı önerisi (örn. Engelhardt ve Goldhammer, 2019), tepki süreçlerini analiz etmek için yeni istatistiksel modellemeler (örn. Chen ve arkadaşları, 2019), yanıt süresi ve yanıt doğruluğunun ortak modellemesi (örn. Bolsinova ve Molenaar, 2018; Guo ve ark., 2020; Klotzke ve Fox, 2019; Kroehne ve ark., 2019; Reis Costa ve ark., 2021; Wang ve ark., 2019), süreç verilerinin hiyerarşik yapısının ortaya koyulmasına yönelik modellemeler (örn. Liu ve ark., 2018), süreç verilerinin analizinin tanısal sınıflandırma modelleriyle birlikte ele alınması (örn. Zhan ve Qiao, 2022), sınavla ilgili ön bilgisi olan bireylerin tespiti (örn. Toton ve Maynes, 2019), hızlı tahmin davranışının tespiti (örn. Lee ve ark., 2019) konularını ele alan pek çok çalışmanın olması dikkat çekicidir. Özetle; süreç verilerinin, sonuç verilerinden elde edilen bilgilere katkı sunarak bireylerin yeteneklerine ilişkin çıkarımları desteklemek için başka türlü gözlemlenemeyecek verileri elde etmemize ve yapıları daha iyi anlamlandırmamıza imkân sağlayabileceği söylenebilir.

Süreç verilerinin psikometrik çıkarımlara kattığı zenginliğin yanında bazı dezavantajları da bulunmaktadır. Günlük dosyaları doğrudan kullanıma hazır değildir. Bu dosyalar öğrenmeyle ilgili olma ihtimali olan tüm öğrenci eylemlerini yakalamak üzere tasarlandığından oldukça büyüktür ve eylem dizileri standart olmayan bir formattadır (yani,

farklı katılımcılar için farklı veri uzunlukları vardır). Bu nedenle ilgisiz olabilecek büyük miktarda veri içerirler. Çok fazla bilgi içermeleri sebebiyle tasarım sürecine dahil olmayan bir araştırmacı ham günlük dosyası verisini anlamlandırmada zorlanabilir. Bu durumlarla başa çıkabilmek için farklı veri yapılarıyla uğraşmayı içeren ön işleme ve hazırlık süreci gerekir. Dolayısıyla günlük dosyalarındaki verilere mevcut psikometrik modellerin doğrudan uygulanması mümkün değildir. Öte yandan günlük dosyalarından oluşturulabilecek anlamlı süreç verileri, sistem tasarımı sırasında kaydedilmesi planlanan durumlarla sınırlıdır (Goldhammer ve ark., 2017; Kerr ve ark., 2011; Fu, Chen ve ark., 2023; Zhang ve ark., 2023; Xu ve ark., 2018). Örneğin çalışma kapsamında ele alınan Uluslararası Öğrenci Değerlendirme Programı (PISA) 2012 (Ekonomik İş birliği ve Kalkınma Örgütü [OECD], 2014) problem çözme değerlendirmesindeki klima kontrolü maddesinde, kontrol çubuklarının durumu sadece “uygula” butonuna tıklandığında günlük dosyasına kaydedilmektedir (Chen ve ark., 2019). Bu sınırlılıklarla başa çıkabilmek için çeşitli yöntemler/yaklaşımlar ortaya çıkmıştır. Örneğin, Hao ve arkadaşları (2016) oyun/simülasyon tabanlı değerlendirmelerden elde edilen günlük dosyalarının düzenlenmesi ve önemli özelliklerin çıkarılabilmesi için Python yazılımında *glassPy* paketi geliştirmişlerdir. Bu paketteki fonksiyonlar sayesinde ham günlük dosyaları daha düzenli XML şemasına uygun hale getirilmekte ve olaylar/olay dizileri hakkında özet bilgiler elde edilebilmektedir. Öte yandan Xu ve arkadaşları (2018) çalışmasında günlük dosyasından bireylerin kümelenmesine imkân tanıyacak bir model önermişlerdir. Önerilen model günlük dosyasından doğrudan örtük sınıflar oluşturmayı sağlamaktadır. Günlük dosyalarını işlemek, analiz etmek için geliştirilen başka bir araç ise R yazılımındaki *ProcData* paketidir. Bu paketteki fonksiyonlarla süreç verilerinden özellik çıkarma, düzensiz bilgileri düzenli sayısal vektörlere dönüştürme gibi işlemler yapılabilir (Tang ve ark., 2021). Günlük dosyalarının sınırlılıklarına yönelik geliştirilen çözümlere ek olarak bu dosyaların kullanımını önemli ölçüde kolaylaştıran PISA 2012 döngüsüne ait veriler dikkat çekmektedir. Bu

döngüye ait günlük dosyalarının açık erişimde² olması araştırmacılar için önemli bir fırsattır. Dahası bu günlük dosyaları iyi yapılandırılmıştır (Han ve Wilson, 2022). Aynı zamanda araştırmacıların karmaşık değerlendirmelerde karşılaşılabilecekleri durumları temsil etmektedir (Qiao ve Jiao, 2018). Bu günlük dosyaları ilgisiz olabilecek kayıtları içermezler ve bu sayede özellik çıkarılması için elverişli bir formattadır. Buna ek olarak PISA değerlendirmeleri yüksek genellenebilirlik potansiyeline sahiptir. Dolayısıyla bu değerlendirmeye ait günlük dosyalarının kullanımı oldukça yaygındır (örn. Chen ve ark., 2019; Eichmann ve ark., 2019; Eichmann ve ark., 2020; Fu, Chen ve ark., 2023; Fu, Zhan ve ark., 2023; Greiff, Wüstenberg ve ark., 2015; Han ve ark., 2019; Liu ve ark., 2018; Pejić ve Molcer, 2021; Qiao ve Jiao, 2018; Ren ve ark., 2019; Tang ve ark., 2021; Xu ve ark., 2018; Xu ve ark., 2024; Zhan ve Qiao, 2022).

Bilgisayar tabanlı değerlendirmelerde bireylerin görev alanını keşfetmeleri ve belirlenen hedeflere göre çözüm üretmeleri gerekir (Fu, Zhan ve ark., 2023). Dolayısıyla bu tür değerlendirmeler çözüme ulaşmak birden fazla adım gerektirdiği için daha karmaşık bilişsel süreçleri ölçme imkânı sunarlar (Zhan ve Qiao, 2022). Bilgisayar tabanlı simülatif değerlendirmeler, bireylerin eylemlerinin ve görev ortamının etkileşimli olarak değişmesine izin vermesi sebebiyle insanın karar verme ve problem çözme davranışını incelemek için önemli bir fırsat sunmaktadır (Funke, 2001). Süreç verisi ile ilgili çalışmalar incelendiğinde, teknoloji açısından zengin ortamlarda öğrenci yeteneklerini ölçmek için simülasyon görevlerini kullanan ve rutin olmayan doğası nedeniyle önemli 21. yüzyıl becerilerinden biri olarak görülen karmaşık problem çözme (KPÇ) karşımıza çıkmaktadır (Greiff, Wüstenberg ve ark., 2015; Xu ve ark., 2018). KPÇ değerlendirmesi için 1980'li yıllardan itibaren bilgisayar tabanlı ölçme araçları kullanılmıştır (Vollmeyer ve ark., 1996). Bu beceriye ilişkin değerlendirmeler temelde gerçek hayat problemlerine odaklanır. Bunlar gerçek hayat problemlerindeki gibi çok adımlı kararlar vermeyi gerektiren, iyi yapılandırılmamış ve

² Uluslararası Yetişkin Yeterlikleri Değerlendirme Programı (PIAAC) 2012 (Goodman ve ark., 2013) değerlendirmesine ait günlük dosyaları da açık erişimdedir.

dinamik yapıya sahiptir (Funke, 2001). Aynı anda birden fazla hedefe ulaşmak için birden fazla değişkenin kontrol edilmesini gerektirmeleri anlamında bu problemler karmaşıktır (Vollmeyer ve ark., 1996). Dörner ve Funke'ye (2017) göre KPÇ becerisine ilişkin ölçümler yalnızca bir çözümün sonucuna indirgenmemelidir. Süreçte yapılan işlemler bireyin problem çözme yetkinliği hakkında nihai sonuçtan daha farklı bilgiler içerebilir. Bununla birlikte bir beceri ve onu ölçmeyi hedefleyen ölçme araçları ne kadar çok keşif davranışı ve eylem gerektirirse o kadar zengin günlük dosyaları oluşur (Greiff ve ark., 2016). Bu özelliklerinden dolayı KPÇ becerisini ölçen görevler zengin günlük dosyaları oluşmasını sağlar (Stadler ve ark., 2019). Dolayısıyla buradan elde edilecek bilgiler genel performans ve problem çözme esnasındaki davranışlar arasındaki ilişkiyi araştırmak için önemli bir kaynak haline gelmektedir (Greiff, Wüstenberg ve ark., 2015). Bu bağlamda KPÇ becerisinin, süreç verilerinin ele alındığı bir çalışmada zengin bir bilgi kaynağı olacağı düşünülmüştür.

KPÇ ile ilgili araştırmalar yapının daha iyi anlaşılması ve günlük dosyalarının potansiyelinin keşfedilmesinden, günlük dosyalarından özellik çıkarma yöntemlerinin geliştirilmesine kadar çeşitlilik göstermektedir. Amaçlarına göre bazı çalışmalarda sadece sonuç veya süreç verileri, bazılarının da ise her iki veri türü birlikte kullanılmıştır. Buna göre ilgili beceriye ait yapının daha ayrıntılı bir şekilde ortaya koyulmasını amaçlayan bazı çalışmalarda okul notları, madde doğruluğu, akıl yürütme testleri gibi sonuç verilerinin kullanıldığı görülmektedir (örn. Greiff, Fischer ve ark., 2013; Wüstenberg ve ark., 2012). Öte yandan pek çok çalışmada yalnızca süreç verilerinin kullanıldığı görülmüştür (örn. Chen ve ark., 2019; Greiff, Wüstenberg ve ark., 2015; Greiff ve ark., 2016; Pejic ve Molcer, 2021; Stadler ve ark., 2019; Ren ve ark., 2019; Liu ve ark., 2018; Qiao ve Jiao, 2018; Xu ve ark., 2018). Buna karşılık KPÇ bağlamında süreç ve sonuç verilerinin birlikte kullanıldığı çalışmalar oldukça sınırlıdır (örn. Scherer ve ark., 2015; Zhan ve Qiao, 2022; Wüstenberg ve ark., 2012). Bu iki veri türünün birlikte ele alınmasıyla öğrencilerinin gelişiminin temsiline ve izlenebilirliğinin daha doğru olabileceği ifade edilmektedir (Zhang ve Deane, 2015). Kullanılan veri türlerinin birbirlerine üstünlükleri ve sınırlılıkları bulunmaktadır. Her biri

araştırmacının amacına en uygun şekilde hizmet edecek biçimde kullanılmalıdır. Öte yandan süreç ve sonuç verilerinin birlikte kullanılmasının, ölçülmesi hedeflenen örtük özelliğin daha bütüncül bir bakış açısıyla anlaşılmasına katkı sunacağı düşünülmektedir. Bununla birlikte KPÇ becerisinden elde edilen puan yorumlamalarının geçerliği artırılabilir.

Çalışmada kullanılan PISA verisinin, performansın altında yatan bilişsel süreçle yönelik bilgiler vermesi, yeni yöntemlerin keşfedilmesi, yeni modelleme çalışmalarının yapılması gibi pek çok konuda faydalı bir kaynak olduğu söylenebilir. Bu araştırma aynı zamanda büyük veri sağlaması sebebiyle veri madenciliği (makine öğrenmesi) tekniklerinin uygulanması için de uygun bir kaynaktır (Yu ve ark., 2012). Bununla birlikte bu tekniklerin bir uzantısı olan eğitsel veri madenciliği alanında yapılmış çeşitli çalışmalar bulunmaktadır (örn Gobert ve ark., 2015; Pejić ve Molcer, 2021; Sinharay, 2016; Sinharay ve ark., 2019; Stadler ve ark., 2019). Bilinen pek çok makine öğrenme algoritması içinden sınıflandırma ve regresyon ağaçları (SRA) oldukça yüksek anlaşılabilirliğe sahip olmalarıyla oldukça popülerdir (West, 2012, s. 280). İlk olarak Breiman ve arkadaşları tarafından 1980'lerin ortalarında (Breiman ve ark., 1984) ortaya atılan SRA'lar bağımlı değişkenin değerini bir dizi bağımsız değişkenden yararlanarak tahmin etmeyi içeren parametrik olmayan regresyon yöntemlerindendir (Strobl, 2013, s. 683). SRA'lar tüm ölçek türlerindeki değişkenleri ele alabilmeleri, eksik verilerle başa çıkabilmeleri, değişken önemini değerlendirmek için ölçütler sunabilmeleri (Breiman ve ark., 1984, s. 216, Steinberg ve Colla, 1995, s. 25), bağımsız değişkenlerdeki uç değerlere karşı sağlam olmaları, bağımsız değişkenlerin monoton dönüşümlerinden etkilenmemeleri, büyük örneklemelerde de hesaplama yapabilmeleri, değişkenlerin önceden seçilmesine gerek olmadan ilgisiz bağımsız değişkenlerle başa çıkabilmeleri (Breiman ve ark., 1984, s. 253; Hastie ve ark., 2009, s. 351; Steinberg ve Colla, 1995, s. 23-24) gibi çeşitli avantajlara sahiptir. Bu özelliklerle SRA'lar makine öğrenme alanında önemli ölçüde popüler öğrenme yöntemi olarak bilinirler (Hastie ve ark., 2009, 352, Loh, 2014a; West, 2012, s. 280). Eldeki

çalışmada kullanılan veri setlerindeki değişkenlerden birçoğunun normal dağılım göstermemesi, doğrusal olmayan ilişkileri içermesi gibi durumlardan dolayı regresyon ağaçlarının ve metnin devamında bahsedilecek bazı makine öğrenme algoritmalarının kullanımının avantaj sağlayacağı düşünülmüştür.

Belirtilen avantajlarıyla birlikte ağaçların bazı sınırlılıkları da söz konusudur. Bunlardan biri ağaçların kararsızlığıdır. Buna göre ağaçlar öğrenme verisindeki çok küçük bir değişiklikte bile oldukça farklı bölünmelere yol açabilirler (Hastie ve ark., 2009, s. 312). Bu kararsızlık nedeniyle tekli ağaçlardan elde edilen tahminler yüksek değişkenlik gösterirler (Strobl, Malley ve ark., 2009). Buna çözüm olarak ağaçları yapı taşı olarak kullanan torbalama, rastgele orman ve artırma gibi topluluk öğrenme yöntemlerinin tekli ağaçların tahmin performansını önemli ölçüde artırabileceği ifade edilmektedir (James ve ark., 2013, s. 316). Bu sebeple çalışmada tekli regresyon ağaçlarına ek olarak torbalama, rastgele orman ve gradyan artırma yöntemleri de kullanılmıştır. Bununla birlikte rastgele orman algoritmasında Breiman ve Cutler'ın (2008) önerdiği şekliyle değişken önemi hesaplanırken istatistik, ağaç sayısına göre elde edilen standart hataya göre standartlaştırılır. Ancak bu ölçekleme işleminin Strobl ve Zeileis (2008) tarafından, testin davranışında bazı beklenmedik durumlara yol açtığı bulunmuştur. Bu sebeple çalışmada rastgele orman analizlerinde değişken öneminin hem ölçeklenmiş hem de ölçeklenmemiş istatistiğe göre elde edilen sıralamaları da incelenmiştir.

SRA'lar doğası gereği (ayrıntılı açıklama Bölüm 2'de sunulmuştur) daha fazla olası bölünme noktasına sahip (daha çok kategorili veya sürekli) bağımsız değişkenlerin lehine değişken seçim yanlılığına yol açar (Hothorn ve ark., 2006). Bu durum Strobl ve arkadaşları (2007) tarafından gösterilmiştir. Aynı zamanda algoritmanın istatistiksel anlamlılık kavramına sahip olmaması da bir sınırlılık olarak görülmektedir (Mingers, 1987). Bu sorunlarla başa çıkabilmek için parametrik olmayan ve ağaç tabanlı regresyon modellerini koşullu çıkarım teorisiyle bütünleştiren (Hothorn ve ark., 2015) koşullu çıkarım ağaçları algoritması Hothorn ve arkadaşları (2006) tarafından önerilmiştir. Bu çalışmada kullanılan

değişkenler de farklı ölçek türlerindedir. Bazı değişkenlerin ranjları oldukça geniştir. Dolayısıyla bu çalışmada koşullu çıkarım ağaçları, koşullu çıkarım ağaçlarına dayalı torbalama, rastgele orman ve gradyan artırma algoritmaları da kullanılmıştır. Öte yandan Strobl ve arkadaşları (2008) rastgele orman algoritması için korelasyonlu yordayıcı değişkenlerin varlığında oluşan değişken seçim yanlılığını göstermişlerdir. Araştırmada rastgele orman algoritmasında ilişkili yordayıcıların varlığında bunların önem değerlerinin yapay bir şekilde arttığı raporlanmıştır (Strobl ve ark., 2008). Bunun için araştırmacılar koşullu çıkarım ağaçlarına dayalı rastgele orman algoritması için regresyon modellerindeki beta katsayıları anlamında koşullu olacak şekilde (Hothorn ve Zeileis, 2023, s. 55) hesaplama yapan versiyonda değişken öneminin hesaplanmasını önermişlerdir. Bu çalışmada kullanılan veri setlerindeki (özellikle süreç verisinde) bazı bağımsız değişkenler arasında yüksek korelasyon bulunmaktadır. Dolayısıyla çalışmada koşullu çıkarım ağaçlarına dayalı rastgele orman algoritması için değişken önem düzeyi sıralamaları koşullu olan ve koşullu olmayan versiyonlar için ayrı ayrı incelenmiştir.

Araştırmanın Amacı ve Önemi

Önemli bir 21. yüzyıl becerisi olan KPÇ'ye ilişkin araştırmalar incelendiğinde süreç ve sonuç verilerinin bir arada kullanıldığı çalışmaların sınırlı olduğu görülmüştür (örn. Scherer ve ark., 2015; Zhan ve Qiao, 2022; Wüstenberg ve ark., 2012). Aynı zamanda KPÇ, süreç verisi ve makine öğrenme algoritmalarının bir arada kullanıldığı çalışmalar sınırlı olmakla birlikte bu kapsamda değerlendirilebilecek Qiao ve Jiao (2016) ve Pejic ve Molcer (2021) çalışmalarında sınıflandırma ağaçlarına dayalı yöntemleri ele almışlardır. Öte yandan çalışmada kullanılan makine öğrenme algoritmalarının ayrı ayrı olarak eğitim alanındaki bazı çalışmalarda kullanıldığı görülmüştür. Buna regresyon ağaçları için Aydoğan ve Gelbal (2022), Blanch ve Aluja (2013), Gomes ve arkadaşları (2021), Gomes ve Jelihovschi (2020); topluluk öğrenme yöntemleri için Güre ve arkadaşları (2020), Sinharay (2016), Sinharay ve arkadaşları (2019); koşullu çıkarım ağaçlarına dayalı yöntemler için Gomes ve arkadaşları (2020) örnek verilebilir. Bununla birlikte bilindiği

kadarıyla KPÇ'ye yönelik süreç ve sonuç verilerini birlikte içeren, eğitim ile ilgili bir veri setinde tekli regresyon ağaçları, topluluk öğrenme yöntemleri ve koşullu çıkarım ağaçlarına dayalı yöntemlerin bir arada ele alındığı bir çalışmaya rastlanmamıştır. Aynı zamanda rastgele orman algoritması gibi önemli bir teknik için değişken öneminin hesaplanmasında standartlaştırma işleminin testin gücü bağlamında etkileri (Strobl ve Zeileis, 2008) ve ilişkili yordayıcıların varlığında oluşan değişken seçim yanlılığını engellemeye yönelik önerilen koşullu değişken önemi sınıflandırma ağaçlarına dayalı olarak araştırılmıştır (Strobl ve ark., 2008). Ancak bu önerilen yöntemlerin regresyon ağaçlarındaki davranışının nasıl olacağı ve değişken önem sıralamalarına etkisi merak konusudur. Dolayısıyla çalışmada PISA 2012 Türkiye verisinde KPÇ'ye ilişkin klima kontrolü ünitesinin birinci maddesine ait günlük dosyasından elde edilen süreç verilerinin, farklı bilişsel ve duyuşsal yapıları içeren sonuç verilerinin ayrı ayrı ve birlikte ele alınarak KPÇ ile ilişkilerinin regresyon ağaçları, topluluk öğrenme yöntemleri (torbalama, rastgele orman, gradyan artırma) ve koşullu çıkarım ağaçlarına dayalı yöntemlerle (koşullu çıkarım ağaçları, koşullu çıkarım ağaçlarına dayalı torbalama, rastgele orman, gradyan artırma) incelenmesi ve değişken önem sıralamalarının tespit edilmesi amaçlanmıştır. Bu amaca bağlı olarak elde edilen bulgularla, eğitsel bir veri setinde farklı makine öğrenme algoritmalarının performanslarının karşılaştırılması, verideki esas ilişkilerin daha ayrıntılı biçimde ortaya konulması, KPÇ'ye ilişkin süreç ve sonuç verilerinin bir arada ele alınmasıyla yapıya ilişkin daha derinlemesine yorum yapılması, KPÇ'yi ölçen testler için etkin biçimlendirici geri bildirim mekanizmalarının işlemlerine katkı sunulması, test geliştirme süreçlerine yönelik düzenlemelere ilişkin bilgi sağlanması açısından alan yazına katkıda bulunulması hedeflenmiştir.

Araştırma Problemi

PISA 2012 klima kontrolü ünitesi birinci maddesine ait süreç verilerinin ve çeşitli sonuç verilerinin; tekli regresyon ağacı, torbalama, rastgele orman, gradyan artırma, koşullu çıkarım ağaçları, koşullu çıkarım ağaçlarına dayalı torbalama, rastgele orman, gradyan

artırma algoritmaları analizlerine göre karmaşık problem çözme becerisini yordama düzeyleri nedir?

Alt Problemler

1. Karmaşık problem çözme becerisine ilişkin elde edilen süreç verilerinin, tekli regresyon ağacı, torbalama, rastgele orman, gradyan artırma, koşullu çıkarım ağaçları, koşullu çıkarım ağaçlarına dayalı torbalama, rastgele orman, gradyan artırma algoritmaları kullanılarak ilgili beceriyi yordama performansı nedir ve hangi yöntem daha iyi performans sergilemiştir?
2. Çalışmada ele alınan sonuç verilerinin, tekli regresyon ağacı, torbalama, rastgele orman, gradyan artırma, koşullu çıkarım ağaçları, koşullu çıkarım ağaçlarına dayalı torbalama, rastgele orman, gradyan artırma algoritmaları kullanılarak ilgili beceriyi yordama performansı nedir ve hangi yöntem daha iyi performans sergilemiştir?
3. Süreç ve sonuç verilerinin birlikte ele alındığı durumda oluşan veri setindeki değişkenlerin, tekli regresyon ağacı, torbalama, rastgele orman, gradyan artırma, koşullu çıkarım ağaçları, koşullu çıkarım ağaçlarına dayalı torbalama, rastgele orman, gradyan artırma algoritmaları kullanılarak ilgili beceriyi yordama performansı nedir ve hangi yöntem daha iyi performans sergilemiştir?

Çalışmada kullanılan veri setlerinin (süreç/sonuç/süreç+sonuç) yordama performansları incelendiğinde;

- a) *Tekli regresyon ağacı ile topluluk öğrenme yöntemlerinden (torbalama, rastgele orman, gradyan artırma)*
- b) *Tekli regresyon ağacı ile koşullu çıkarım ağacı yöntemlerinden*
- c) *Topluluk öğrenme yöntemleri ile koşullu çıkarım ağaçlarına dayalı topluluk öğrenme yöntemlerinden*

daha iyi sonuç veren algoritma hangisidir?

4. Karmaşık problem çözme becerisini yordamada değişkenlerin önem düzeyi sıralamaları;

- a) Süreç verisinde
- b) Sonuç verisinde
- c) Süreç+sonuç verisinde

tüm algoritmalar için nasıldır?

5. Rastgele orman algoritmasında ölçeklenmiş ve ölçeklenmemiş versiyonlarına göre değişken önem düzeyi sıralamaları;

- a) Süreç verisinde
- b) Sonuç verisinde
- c) Süreç+sonuç verisinde

nasıldır?

6. Koşullu çıkarım ağaçlarına dayalı rastgele orman algoritmasında koşullu ve koşullu olmayan versiyonlarına göre değişken önem düzeyi sıralamaları;

- a) Süreç verisinde
- b) Sonuç verisinde
- c) Süreç+sonuç verisinde

nasıldır?

Sayıtlılar

Değerlendirmeye katılan öğrencilerin gerçek performanslarını yansıtacak şekilde katılım sağladığı varsayılmaktadır.

Sınırlılıklar

Çalışmada, PISA 2012 problem çözme değerlendirmesinde yer alan Klima Kontrolü ünitesindeki birinci maddeye ait günlük dosyası kullanılmıştır.

Tanımlar

Karmaşık problem çözme: Dinamik olan (yani kullanıcının müdahalelerinin ve/veya zamanın bir fonksiyonu olarak değişen) ve ortamın kurallar bütünü, tamamı olmasa da bir kısmının, ancak bu süreçte kazanılan bilgilerin başarılı bir şekilde keşfedilmesi ve bütünleştirilmesiyle ortaya çıkarılabildiği görev ortamlarıyla başarılı bir etkileşimdir (Buchner, 1995, s.14).

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde karmaşık problem çözme becerisi, regresyon ağaçları, topluluk öğrenme algoritmaları ve koşullu çıkarım ağaçlarına dayalı makine öğrenme algoritmalarıyla ilgili kuramsal çerçeve oluşturularak bu konu alanlarında yapılan araştırmalara yer verilmiştir.

Karmaşık Problem Çözme Becerisi

Karmaşık problem çözme (KPÇ) günlük yaşamda ve eğitimde önemli bir yetkinlik olup, bu alanlarda karşılaşılan çeşitli zorluklarla başa çıkabilmek için gerekli bir beceridir (Eichmann ve ark., 2019; Herde ve ark., 2016). Buna göre öğrencilerin 21. yüzyılda karşılaşılabilecekleri, basit problemlerle yakalanamayan, dinamik ve şeffaf olmayan durumları içeren KPÇ görevleriyle karşılaşmalarının önemli olduğu vurgulanmaktadır (Wüstenberg ve ark., 2014). Bu beceri, temelde belirli bir başlangıç durumu ve amaçlanan hedef durumu arasındaki engelleri azaltma biçiminde ortaya çıkar (Funke, 2012). Bununla birlikte Funke'ye (2003) göre karmaşık problemlerin özellikleri, basit problemlerin özelliklerinden oldukça farklıdır. Bu bağlamda karmaşıklık, ilgili değişkenler arasındaki bağlantı, durumun dinamikleri, şeffaf olmama ve birden fazla hedefin aynı anda dikkate alınması gereken durumları içermesi (polytely) karmaşık problemlerin beş temel özelliğidir. İlk olarak problem durumunun karmaşıklığı bir problem yapısındaki değişkenlerin sayısıdır. İkinci olarak, bu değişkenler arasında bağlantılar söz konusudur. Üçüncü olarak, durumun dinamikleri özelliği karmaşık problem sistemine yapılan müdahalelerin değişkenler arasındaki ilişkileri değiştirebilmesini ifade eder. Bu değişim amaçlanmayan süreçleri de harekete geçirebilir. Dahası bazı değişkenlerin durumu zaman içinde kendi kendine de değişebilir (eigendynamics-iç dinamikler). Dördüncü olarak, şeffaf olmama özelliği olası hedefler ve değişkenler hakkında gerekli tüm bilgilerin verilmemesi anlamına gelir. Beşinci olarak, polytely özelliği ise karmaşık problem durumunda önceliklerin tanımlanmasını

gerektirecek, birbirleriyle çelişebilen birden fazla hedef oluşunu içerir (Akt., Funke, 2012). Bu özellikleri de vurgulayan bir tanım Buchner (1995) tarafından yapılmıştır:

Dinamik olan (yani kullanıcının müdahalelerinin ve/veya zamanın bir fonksiyonu olarak değişen) ve ortamın kurallar bütününün, tamamı olmasa da bir kısmının, ancak bu süreçte kazanılan bilgilerin başarılı bir şekilde keşfedilmesi ve bütünleştirilmesiyle ortaya çıkarılabildiği görev ortamlarıyla başarılı bir etkileşimdir. (s.14).

Buchner'ın tanımı, problem durumu hakkında bilgi edinilmesini ve edinilen bu bilginin problemi çözmek için uygulanmasını gerektirir (Herde ve ark., 2016). KPÇ performansı için esas olan bu iki durum Funke (2001) tarafından tanımlanmıştır. Buna göre bilgi edinme aşamasında katılımcının, değişkenler arasındaki ilişkileri ve sistemin dinamiklerini anlaması gerekir. Bilgi uygulamada ise ilk aşamada sistem hakkında edinilen bilginin, soruda belirtilmiş hedef durumuna ulaşmak için kullanılması söz konusudur. Bilgi edinimi aşamasının, bilgi uygulama aşaması için gerekli ve yeterli bir koşul olduğuna dair nedensel bir ilişki olduğu yaygın bir varsayımdır (Funke, 2001). KPÇ'nin yapı geçerliğinin araştırıldığı Greiff, Fischer ve arkadaşları (2013) tarafından yapılan çalışmada bilgi edinme ve bilgi uygulama boyutu arasında yüksek ilişki olduğu bulunmuştur. Benzer kapsamda Greiff, Wüstenberg ve arkadaşları (2013) tarafından yapılan başka bir çalışmada da yapının bilgi edinme ve bilgi uygulama olmak üzere iki boyutlu olduğu sonucuna ulaşılmıştır. Wüstenberg ve arkadaşları (2012) tarafından yapılan başka bir araştırmada iki boyutlu yapı doğrulanmış ve değişkenler arasındaki ilişki kurallarının bilinmesinin, bu kuralları uygulamak için ön koşul olduğu ifade edilmiştir.

KPÇ becerisinin ölçülmesinde iki farklı yaklaşım bulunmaktadır. Doğrusal yapısal eşitlik yaklaşımı (DYE) ve sınırlı durum otomatları (SDO) olarak ifade edilen bu iki yaklaşımda değişkenler ve etkileri arasındaki ilişkinin modellenmesi söz konusudur. DYE nicel değişkenler arasındaki doğrusal ilişkileri modellemek için bir çerçeve sağlarken, SDO nitel değişkenler arasındaki ilişkileri modellemek için kullanılır (Funke, 2001). Kahve tüketiminin uyanıklık ve susuzluk üzerindeki etkisinin incelenmesini içeren bir problem

durumu DYE, bir cep telefonu veya bilet otomatının düğmelerinin etkilediği durumları içeren bir problem SDO yaklaşımına örnektir (Greiff, Fischer ve ark., 2015). DYE yaklaşımında, bilgi edinme aşamasında girdi değişkenlerinin değiştirilebilen değerlerinin çıktı değişkenlerinde oluşturduğu değişikliklerden yararlanarak sistemin nedensel yapısı hakkındaki ilişkiye ulaşılmaya çalışılır. SDO yaklaşımında da bilgi edinme aşaması benzerlik gösterir. DYE kapsamında bilgi uygulaması aşamasında girdi değişkenleri yardımıyla çıktı değişkenleri için soruda belirtilmiş hedef duruma ulaşılması gerekir. SDO durumunda ise, başlangıç durumundan hedef duruma ulaşmak için bir yol bulunması bilgi uygulaması aşamasını temsil eder (Funke, 2001). SDO'ya dayalı problemlerde bir müdahalenin sonucu, DYE'ye dayalı problemlerde olduğu gibi, miktarla değil sistemin yeni bir durumuyla ifade edilir (OECD, 2014a, s. 34). Funke (2001) tarafından tanımlanan bu iki yaklaşımla karmaşık problemler ortak çerçevelerde ifade edilebilmiş ve buna bağlı olarak KPÇ araştırmalarında karşılaştırılabilirlik sorunu çözülerek, yeni problem durumlarının tasarlanması daha esnek hale gelmiştir (Greiff, Fischer ve ark., 2015). Ancak Funke'nin (2001) yaklaşımı karşılaştırılabilirlik sorununu çözmüş olsa da ölçeklenebilirlik sorununu çözmemiştir. Bu bağlamda KPÇ'nin ölçümünün yapıldığı ilk dönemlerde tek bir görevin kullanılması testin psikometrik özellikleri açısından çeşitli sınırlılıklara yol açmıştır (Greiff, Fischer ve ark., 2015). Bu sebeple bahsedilen sınırlılıklara çözüm olarak birden fazla görevin kullanıldığı çoklu karmaşık sistemler önerilmiştir. Bu amaçla MicroDYN ve MicroFIN yaklaşımları tanımlanmıştır. Buna göre MicroDYN yaklaşımı DYE'ye dayalı problemlerden, MicroFIN yaklaşımı SDO'ya dayanan problemlerden oluşan setlerdir (Greiff, Fischer ve ark., 2015).

Modern eğitim bağlamlarında önemli bir konu olan KPÇ, PISA ve PIAAC gibi uluslararası değerlendirme programlarına dahil edilmiştir (Greiff ve Fischer, 2013). PISA 2012 döngüsünde KPÇ becerisi ölçülmüştür. Araştırmada bu beceri "yaratıcı problem çözme" olarak etiketlenmiştir (OECD, 2014a). Bu doğrultuda 15 tanesi statik (sorunun çözümü için gerekli tüm bilgilerin başlangıçta verilmesi durumu), 27 tanesi etkileşimli

(gerekli tüm bilgilerin başlangıçta verilmediği ve süreçte katılımcı tarafından keşfedilmesi beklenen durum) olan ve dört farklı problem çözme sürecini temsil eden (keşfetme ve anlama, temsil etme ve formüle etme, planlama ve uygulama, izleme ve yansıtma) 42 maddeyle problem çözme becerisi değerlendirilmiştir (OECD, 2014a, s. 29, 79). Araştırmada statik ve etkileşimli maddelerdeki performansın güçlü bir ilişkide olduğu raporlanmıştır (OECD, 2014a, s. 81). Buna ek olarak, bilgi edinme ile ilgili görevler PISA kapsamında “keşfetme ve anlama” veya “temsil etme ve formüle etme” süreçlerini içerirken, bilgi uygulama ile ilgili görevler “planlama ve uygulama” sürecine karşılık gelir. “İzleme ve yansıtma” süreçlerini içeren görevler ise hem bilgi edinme hem de bilgi uygulama süreçlerini içerirler (OECD, 2014a, s.84). Araştırmadaki etkileşimli problemleri içeren ünitelerden dördü MicroDYN, altısı MicroFIN türünde karmaşık problemleri içermektedir. Bunlardan “Lost, Climate Control, Aquarium, Advertising, Remote Controls, Drink Machine, Clock, Vitamins, Tickets, Smiley” üniteleri karmaşık problem çözmeye dayalı görevler içerirken; “Robot Cleaner, Traffic, New Furniture, Blocks Game, Communications” üniteleri statik türden görevleri içermektedir. Buna ek olarak üç maddeden oluşan “Lights” ünitesinin 1. ve 3. maddeleri karmaşık, 2. maddesi statik problem türündendir (OECD, 2014b, s. 417). Problem çözme değerlendirmesinde bulunan 16 ünite, her biri 20 dakikada tamamlanacak şekilde dört küme halinde gruplandırılmıştır. Her katılımcıya, katılımcının bilgisayar tabanlı okuma veya matematik değerlendirmesine katılıp katılmadığına göre bir ya da iki küme verilmiştir. Her iki durumda da bilgisayar tabanlı testler 40 dakika sürmüştür (OECD, 2014a, s. 32). Buna ek olarak test tasarımından dolayı öğrenciler farklı soru setlerini aldıkları için örneklemdaki her öğrenci toplam madde havuzunun yalnızca bir kısmını görmüştür (OECD, 2014a, s. 150).

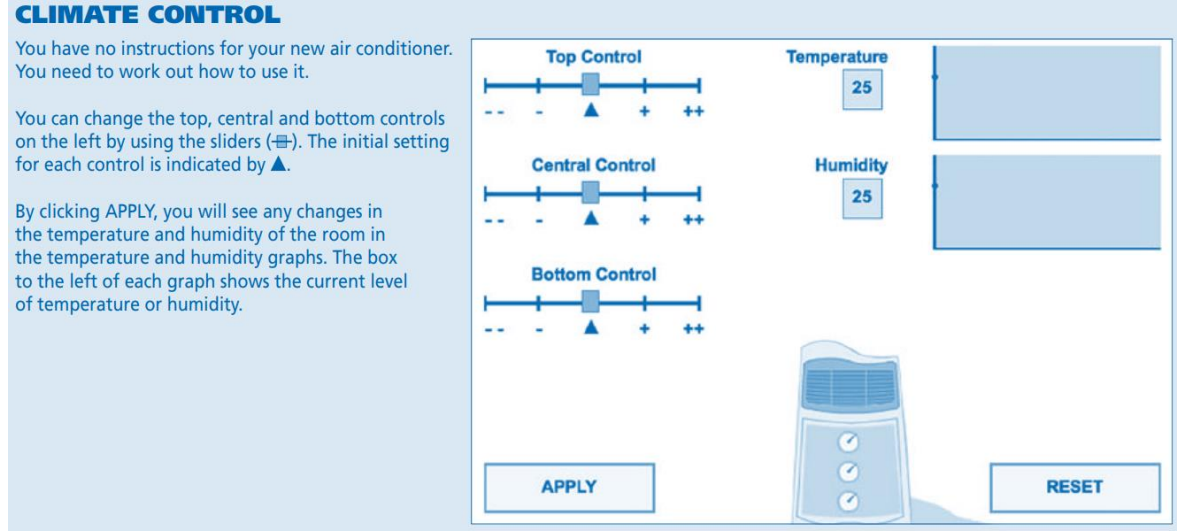
Hem karmaşık hem de statik problemleri içeren PISA 2012 değerlendirmesinde yapılan problem çözme tanımının klasik KPÇ tanımlarıyla arasındaki benzerlik dikkat çekmektedir. Buna göre PISA 2012’de problem çözme şu şekilde tanımlanmıştır: “Bireyin, bir çözüm yönteminin doğrudan belli olmadığı problem durumlarını anlamak ve çözmek için

bilişsel süreci işletebilme kapasitesidir. Yapıcı ve yansıtıcı bir vatandaş olarak kişinin, potansiyeline ulaşmak için bu tür durumlarla ilgilenme isteğini içerir (OECD, 2014a, s. 30)". Bu tanımın ilk cümlesinde geçen ifadelerin sadece sonucun değerlendirilmesinin ötesinde, öğrencilerin problem çözme sürecinde gerçekleştirdiklerinin (örneğin kullanılan stratejiler) ölçülmesini vurguladığı ve bu yönüyle de PISA 2012 problem çözme tanımının, klasik KPÇ tanımlarıyla yakından benzerlik gösterdiği ifade edilmektedir (Greiff, Wüstenberg ve ark., 2015). Bu doğrultuda PISA 2012 döngüsünde değerlendirilen, KPÇ becerisinin teorik çerçevesiyle uyumlu (Greiff, Wüstenberg ve ark., 2015) ve kapsamlı bir günlük dosyasına sahip olan klima kontrolü ünitesine ait ilk maddeye ilişkin veriler çalışmada kullanılmıştır. Bu üniteye ait sorular MicroDYN türündendir (OECD, 2014a, s. 38) ve alana özel ön bilgi gerektirmezler (Greiff ve ark., 2018). KPÇ'nin önemli yönlerini temsil eden MicroDYN problemlerinde bilinmeyen bir sistemdeki nedensel ilişkilerin keşfi ve kontrolü temel görevlerdir (Funke, 2010; OECD, 2014a, s. 34). Yalnızca girdi değişkenlerinin manipüle edilebildiği bu problemlerde girdilerin değiştirilebilen değerlerine göre çıktı değişkenlerin değerleri etkilenir (Funke, 2010; Goode ve Beckmann, 2010). Çalışmada kullanılan klima kontrolü ünitesinde de KPÇ'nin yapısına uygun olarak iki aşama bulunur. Buna göre bilgi edinme aşamasında bireylerin en fazla üç girdi değişkenini kontrol etmesi gerekir (OECD, 2014a, s. 34). Burada bireyler sistemi serbestçe keşfederler. Örneğin, bireyler bu aşamada sistemi sıfırlayıp yeniden keşif yapmaya başlayabilirler (Funke, 2010). Ardından keşfedilen kurala ilişkin bilginin gösterilmesi gerekir (OECD, 2014a, s. 34). Bu işlem diyagram çizimiyle gerçekleştirilir. Zihinsel modelin dışsallaştırılması olarak da ifade edilebilen bu çizim, edinilen nedensel bilginin değerlendirilmesini sağlar (Funke, 2010). İkinci aşama olan bilgi uygulama aşamasında ise bireylerin uygun girdi düzeylerini seçerek çıktı değerlerinde verilen hedef değerlere ulaşmaları için sistemi kontrol etmeleri istenir (Funke, 2010). Bu aşamada bilgi edinme ve bilgi uygulama boyutları arasındaki kavramsal bağımsızlığı sağlamak amacıyla doğru diyagram çizimi kullanıcının ekranında görüntülenir (Greiff ve ark., 2018). Çalışmada tercih edilen ünitenin 1. maddesi bilgi edinme, 2. maddesi bilgi

uygulama boyutu ile ilgilidir (OECD, 2014a, s. 38). Bu açıklamalar doğrultusunda ilgili üniteye ilişkin ekran görüntüsü Şekil 1’de verilmiştir³.

Şekil 1

Klima Kontrolü Ünitesi Bilgilendirme Ekranı (OECD, 2014a, s. 37)



Şekil 1’e göre kullanma kılavuzu olmayan klimanın nasıl çalıştığını bulmak için sol taraftaki üst, orta ve alt kontrol çubuklarının (—) kullanılabileceği ifade edilmektedir. Her bir ayarın başlangıçtaki konumu (▲) ile gösterilmektedir. “Uygula” tuşuna tıklandığında odanın sıcaklık ve nem değerlerindeki değişiklik, sıcaklık ve nem grafiklerinde görülebilir. Her grafiğin solundaki kutu, o andaki sıcaklık ve nem seviyesini göstermektedir.

Şekil 2’de gösterilen ilk soruda öğrencilerin kontrol çubuklarını sağa sola kaydırarak her bir çubuğun sıcaklık ve nem değerlerini değiştirip değiştirmediğini bulmaları istenir. Bu soru problem çözme süreçlerinden temsil etme ve formüle etme basamağındadır (OECD, 2014a, s. 38).

³ Ünitenin Türkçe’ye çevrilmiş hali erişime açık olmadığı için maddelerin orijinallerine ait görüntüler verilmiştir.

Şekil 2

Klima Kontrolü Ünitesi Birinci Sorusu (OECD, 2014a, s. 38)

Question 1: CLIMATE CONTROL CP025Q01

Find whether each control influences temperature and humidity by changing the sliders. You can start again by clicking RESET.

Draw lines in the diagram on the right to show what each control influences. To draw a line, click on a control and then click on either Temperature or Humidity. You can remove any line by clicking on it.

Top Control

Central Control

Bottom Control

Temperature

Humidity

?

→

Yönergeye göre “sıfırla” seçeneğine tıklayarak yeniden deneme yapılabilir. Her bir çubuğun neyi etkilediğinin şekilde oklar çizilerek gösterilmesi istenir. Bir ok çizmek için üst, orta ya da alt kontrol çubuklarından birine tıklanması ve sonra sıcaklık ya da nem kutucuklarına tıklanması gerekmektedir. Herhangi bir ok üzerine tıklanarak kaldırılabilir. Bu soruda öğrencinin keşif yapması için herhangi bir deneme sınırlaması yoktur. Diyagram doğru bir şekilde çizilirse tam puan, her seferinde yalnızca bir girdi değişkeni değiştirilerek değişkenler arasındaki ilişkiler araştırılır ancak diyagram doğru bir şekilde çizilmezse kısmi puan alınmaktadır (OECD, 2014a, s. 38).

Öğrencilerden klimanın sıcaklığını ve nemini belirlenen hedef seviyeye ayarlamalarının istendiği ikinci soru Şekil 3’te verilmiştir. Bu soru problem çözme süreçlerinden planlama ve uygulama basamağındadır (OECD, 2014a, s. 39).

Şekil 3

Klima Kontrolü Ünitesi İkinci Sorusu (OECD, 2014a, s. 38)

Question 2: CLIMATE CONTROL CP025Q02

The correct relationship between the three controls, Temperature and Humidity is shown on the right.

Use the controls to set the temperature and humidity to the target levels. **Do this in a maximum of four steps.** The target levels are shown by the red bands across the Temperature and Humidity graphs. The range of values for each target level is 18-20 and is shown to the left of each red band. **You can only click APPLY four times and there is no RESET button.**

Top Control

Central Control

Bottom Control

Temperature

Humidity

?

→

Şekil 3'e göre üç girdi değişkeni (üst, orta, alt kontrol çubukları) ve iki çıktı değişkeni (sıcaklık ve nem) arasındaki doğru ilişki soruda gösterilmektedir (öğrenciler test sırasında bir önceki soruya geri dönemezler). Sıcaklık ve nemin hedeflenen seviyeye ayarlanması için kontrol çubukları kullanılmalıdır. En fazla dört denemeye izin verilen soruda, hedef seviyeler sıcaklık ve nem grafiklerinde kırmızı bantlarla gösterilmektedir. Her hedef düzeyi için değer aralığı 18-20'dir ve değerler her kırmızı bandın solunda gösterilir. Bu soruda "sıfırla" düğmesi bulunmamaktadır. Öğrenci dört defalık deneme sayısını aşmadan sıcaklık ve nemi hedef seviyelerine ulaştırırsa tam puan, dört deneme sayısı içinde iki çıktıyı da hedef seviyelerine yaklaştırabilirse (her ikisi için de tam olarak istenen seviyeye ulaşamaz) kısmi puan almaktadır (OECD, 2014a, s. 39).

Klima kontrolü ünitesine ilişkin 1. madde pek çok KPÇ değerlendirmesine konu olmuştur (örn. Chen ve ark., 2019; Eichmann ve ark., 2020; Greiff, Wüstenberg ve ark., 2015; Han ve ark., 2019; Pejić ve Molcer, 2021; Xu ve ark., 2018). Greiff, Wüstenberg ve arkadaşlarına (2015) göre de klima kontrolü ünitesi KPÇ becerisini iyi bir şekilde temsil etmektedir. Bu sebeplerden dolayı çalışmanın amacına bağlı olarak klima kontrolü ünitesinin 1. maddesine ait günlük dosyasından elde edilen değişkenler süreç verisi olarak kullanılmıştır. Oluşturulan değişkenlere ilişkin açıklamalara aşağıda yer verilmiştir.

Çalışmada Oluşturulan Süreç Verisi Değişkenleri

Süreç verisinin analizini içeren çalışmalarda sıklıkla ***görevde geçirilen toplam süre*** değişkeninin kullanıldığı görülmüştür. Bu değişken, görevin başlamasından tamamlanmasına kadar geçen süreyi ifade etmektedir. Tepki süreleri bilgisayar tabanlı testlerde otomatik olarak toplanan ve bilişsel süreçleri araştırmak için doğal bir veri türüdür (De Boeck ve Jeon, 2019; Toton ve Maynes, 2019). Görevin tamamlanma sürecinin genel bir özelliği olan sürenin (Goldhammer ve ark., 2017) etkisi için çok çeşitli çalışmalar ve farklı bulgular vardır. Yapılan çalışmalar incelendiğinde süre bilgisinin KPÇ performansının daha detaylı anlaşılmasında (Greiff ve ark., 2016; Scherer ve ark., 2015), ölçülmesi hedeflenen yapıya ilişkin tahmin doğruluğunun artırılması için yanıt doğruluğu ve yanıt süresinin birlikte

modellenmesinde (Bolsinova ve Molenaar, 2018; Klotzke ve Fox, 2019; Reis Costa ve ark., 2012; Wang ve ark., 2019), sınavdan önce sorularla ilgili ön bilgisi olan bireylerin belirlenmesinde (Toton ve Maynes, 2019), hızlı tahmin davranışının tespit edilmesinde (Lee ve ark., 2019), hız-doğruluk dengesinin ortaya çıkarılması için geliştirilen modelleme çalışmalarında (Guo ve ark., 2020) kullanıldığı görülmüştür. Özellikle KPÇ ve süre arasındaki ilişkilerin incelendiği çalışmalarda çeşitli sonuçların ortaya çıktığı dikkat çekmiştir. Örneğin, Greiff ve arkadaşları (2016) çalışmasında KPÇ becerisi ile görevde geçirilen süre arasında ters “U” şeklinde bir ilişki bulunmuştur. Başka bir deyişle, görevde geçirilen çok az ve çok fazla süre düşük KPÇ performansı ile ilişkilendirilmiştir. Öte yandan Scherer ve arkadaşları (2015) çalışmalarında KPÇ performansı ve süre arasında pozitif orta düzey bir ilişki bulmuşlardır. Bu bağlamda performansla çeşitli şekillerde ilişkili olduğu görülen süre değişkeni çalışmaya dahil edilmiştir.

Dikkate alınan bir diğer değişken, problemin çözümü için oldukça önemli olan **strateji (VOTAT) puanıdır**. Çalışmada kullanılan klima kontrolü ünitesine ait birinci problemin çözümünde Tschirgi (1980) tarafından ortaya atılan “her seferinde bir şeyi değiştir (VOTAT; vary-one-thing-at-a-time)” stratejisinin kullanılması gerekmektedir (Herde ve ark., 2016). Bu stratejiye göre olumsuz bir sonuçla karşılaşıldığında, olumsuz sonuca sebep olduğu düşünülen değişkenin değiştirilmesi, diğerlerinin sabit tutulması söz konusudur (Tschirgi, 1980). Greiff, Wüstenberg ve arkadaşları (2015) çalışmalarının sonucunda VOTAT stratejisinin problem çözme alanındaki daha geniş stratejik yeterliklerin geçerli bir göstergesi olduğunu ifade etmişlerdir. VOTAT stratejisi KPÇ araştırmalarında büyük ilgi görmüştür ve bu stratejinin uygulanması genel KPÇ performansının yüksek olmasıyla doğrudan ilişkilidir (Greiff ve ark., 2018). Bu stratejiyi uygulamanın, PISA 2012 problem çözme değerlendirmesi kapsamında, birey ve ülke düzeyinde hem klima kontrolü maddesindeki performansla hem de genel problem çözme performansı ile yüksek düzeyde pozitif ilişkili olduğu bulunmuştur (Greiff, Wüstenberg ve ark., 2015). Yine çeşitli çalışmalarda bu strateji ve KPÇ becerisi arasında pozitif ilişki bulunmuştur (Greiff ve ark.,

2016; Stadler ve ark., 2019). Aynı zamanda KPÇ performansını içeren çeşitli çalışmalarda bu stratejinin önemli olduğu raporlanmıştır (Chen ve ark., 2019; Pejić ve Molcer, 2021). Wüstenberg ve arkadaşları (2012) KPÇ becerisinin yapısını araştırdıkları çalışmada VOTAT kullanmadan çözüme yönelik doğru bir model oluşturma olasılığının, en kolay madde hariç, ortalama %3.4 olduğunu bulmuşlardır. Bu açıklamalar doğrultusunda çalışma kapsamında VOTAT stratejisine ilişkin bir değişken oluşturulması önemli görülmüştür. Bireyin bu stratejiyi uygulamış olması için iki girdi değişkeninin (üst-orta-alt kontrol çubukları) konumu “0” da iken, diğer değişkenin konumunun hareket ettirilmesi gerekir. Buna göre ilgili değişken için yapılan kısmi puanlama şu şekildedir: Üç girdi değişkeninden hiçbirini için VOTAT stratejisi uygulanmamışsa 0, yalnızca bir tanesi için uygulanmışsa 1, iki tanesi için uygulanmışsa 2, üç tanesi için de uygulanmışsa 3 puan verilmiştir.

Çeşitli süreç verileri bireylerin karmaşık bir problemi çözme sürecinde gösterdiği planlama ve düzenleme davranışları hakkında da bilgi verebilir. Bu kapsamda değerlendirilebilecek **göreve başlamadan önce geçirilen süre/bekleme süresi** bilgisi çeşitli çalışmalarda kullanılmıştır (Klotzke ve Fox, 2019; Qiao ve Jiao, 2018). Bu değişken, öğrenci tarafından gerçekleştirilen ilk eylem ile görevin başladığı an arasındaki farkı ifade etmektedir. Örneğin, Chen ve arkadaşları (2019) geliştirdikleri yeni istatistiksel modelleme yaklaşımına göre, KPÇ ortamında görev yönergesini daha hızlı okuyan ve harekete geçen bireylerin problemi daha düşük doğrulukla ve daha kısa sürede tamamlama eğiliminde olduklarını bulmuşlardır. Öte yandan bireylerin gerçekleştirdikleri eylemlerin frekansları ile ilgili de çeşitli süreç verisi değişkenleri elde etmek mümkündür. Bu bağlamda **eylem sayısı** değişkeninin günlük dosyası analizlerinde kullanıldığı görülmektedir (Gobert ve ark., 2015; Teig ve ark., 2020). Bireyin, görev sırasında kaydedilen tüm eylemlerinin sayısı olan bu değişken ile KPÇ becerisi arasında ilişkiler araştırılmıştır. Bu konuda Goldhammer ve arkadaşları (2017) tarafından özetlenen çeşitli araştırmalar dikkat çekmektedir. Örneğin, Naumann ve arkadaşları (2014) PIAAC problem çözme değerlendirmesi verileriyle yaptıkları çalışmada eylem sayısının görev başarısı için önemli bir yordayıcı olduğunu ve

eylem sayısı ile görev başarısı arasında ters U şeklinde ilişki elde ettiklerini raporlamışlardır. Yine aynı çalışmada az, orta ve çok miktarda eylemde bulunmak arasındaki farklılıklar katılımcıların aktiflik düzeyleri, dikkat dağınıklığı ya da görevde ne yapılacağına karıştırılması gibi durumlarla ilişkilendirilmiştir. Öte yandan He ve arkadaşlarının (2019) günlük dosyasından elde ettikleri çeşitli değişkenlerle yaptıkları kümeleme analizinin sonuçlarına göre eylem sayısı ve sürenin kümelemede en baskın özellikler olduğu raporlanmıştır. Aynı zamanda araştırmacılar genel olarak daha uzun süre ve daha fazla eyleme sahip olan bireylerin problem çözme puanlarının daha yüksek olduğunu bulmuşlardır. Kullanılan diğer bir değişken ise bireylerin problemi çözerken sistemi kaç defa sıfırladıklarını içeren **sıfırlama (reset) olaylarının sayısıdır**. Bireylerin yeni baştan işlem yapma sıklıklarının problem çözme performansı ile ilişkili olabileceği düşünüldüğü için ilgili değişken çalışmaya dahil edilmiştir. Chen ve arkadaşları (2019) çalışmalarında sıfırlama tuşunun kullanımının problemi doğru çözme olasılığını artırdığını ve bununla birlikte problemi çözmek için gerekli süreyi de uzattığını ifade etmişlerdir. Süreç ve öğrenci düzeyinde hiyerarşik modelleme yapılan bir başka çalışmada ise sıfırlama sayısının her iki düzeydeki yetenek tahminleriyle ilişkili olduğu bulunmuştur (Liu ve ark., 2018). Oluşturulan bir başka değişken **müdahale sayısıdır**. Soruda yapılan keşif işlemlerinin kaydedilebilmesi için uygula (apply) tuşuna tıklanması gerekmektedir. Buna göre ilgili değişken “uygula” tuşuna tıklanma sayısını ifade eder. Öte yandan sistemde herhangi bir değişiklik yapmadan “uygula” tuşuna tıklanmasını içeren diğer bir değişken **müdahalesiz gözlemdir**. Bu değişken bireylerin sistemdeki dinamikleri keşfetmeleri için en az bir defa tüm değişkenleri “0” da tutarak “uygula” tuşuna tıklamasını ifade etmektedir (Greiff ve ark., 2016). Yapılan bir araştırmada müdahalesiz gözlem yapan ve daha düşük miktarda “uygula” tuşuna tıklayan öğrencilerin KPÇ görevlerini çözmede daha başarılı oldukları bulunmuştur (Greiff ve ark., 2016). Greiff ve arkadaşlarının (2016) yaptığı bu çalışmada puanlama yapılırken, en az bir kez müdahalesiz gözlem yapıp yapmama durumu dikkate alınmıştır. Ancak eldeki çalışmada ilgili işlemin gerçekleştirilme miktarının geniş bir aralıkta olduğu dikkat çekmiştir. Başka bir deyişle bireylerin tüm kontrol çubukları “0” konumunda iken “uygula” tuşuna

basma frekansları oldukça farklılık göstermiştir (0 ile 33 arasında değişen değerler). Aynı eylemin birçok kez tekrarlanması öğrencilerin dikkatiyle ilgili olabilir (Chen ve ark., 2019). Bu sebeple çalışmada ilgili durum için iki farklı değişken oluşturulmuştur. İlk değişken, ilgili işlemi yapıp yapmamayı temsil eden ve 0-1 şeklinde puanlanan **müdahalesiz gözlem puanı** iken diğer değişken ilgili işlemin frekansını ifade eden **müdahalesiz gözlem sayısı** olmuştur. Öte yandan Pejic ve Molcer (2021) tarafından yapılan araştırmada PISA 2012 klima kontrolü ünitesi birinci maddesi için oluşturulan, her bir kontrol çubuğu için toplam “uygula” olaylarının sayısına oranlanmış VOTAT uygulama sayısı değişkenine göre üç kontrol çubuğunu da eşit derecede deneyenlerin görevde en fazla başarıya sahip oldukları bulunmuştur. Buna göre çalışmada sorunun çözümünde kullanılan **her bir kontrol çubuğu (üst-orta-alt) için VOTAT stratejisinin kullanıldığı durumların sayısına** ilişkin üç farklı değişken oluşturulmuştur.

Açıklanan değişkenlere ek olarak veri odaklı bir yaklaşım da izlenmiştir. Bu bağlamda yapıyla ilişkili olabileceği düşünülmüş çeşitli değişkenler de elde edilmiştir. Öncelikle bireylerin VOTAT uygulama miktarlarında çeşitlilik olduğu görülmüş ve bu sebeple her bireyin **VOTAT uygulama sayısı** da başka bir değişken olarak ele alınmıştır. Oluşturulan bir diğer değişken ise **ilk müdahaleye kadar geçen süredir**. Bu değişken ilk “uygula” tuşuna tıklanmasına kadar geçirilen süreyi ifade etmektedir. Keşif yapmaya başlanmasının bir göstergesi olarak ilgili tuşa ilk tıklamaya kadar geçen sürenin bilgi verebileceği düşünülmüştür. Diğer bir değişken **diyagram tuşuna tıklama sayısıdır**. Çalışmada dikkate alınan problemde tam puan alınabilmesi için değişkenler arasındaki ilişkilerin doğru bir şekilde çizilmesi gerekmektedir (OECD, 2014a, s. 34). Bu sebeple ilgili değişkenin ele alınmasının önemli olacağı düşünülmüştür. Bireylerin cevaplarını onaylamadan önce, yaptıklarını kontrol edip etmedikleri hakkında bilgi sağlayabilecek diğer bir değişken **son kontrol süresi** olarak tanımlanmıştır. Bu değişken son eylem ile bir önceki eylemin gerçekleşme zamanı arasındaki farktır. Kullanılan son değişken ise **hızıdır**. İlgili

değişken gerçekleştirilen toplam eylem sayısının, görevde geçirilen toplam süreye oranlanmasıyla elde edilmiştir.

Sonuç olarak eldeki çalışmada oluşturulan süreç verisi değişkenleri; görevde geçirilen toplam süre, strateji (VOTAT) puanı, göreve başlamadan önce geçirilen süre/bekleme süresi, eylem sayısı, sıfırlama olaylarının sayısı, müdahale sayısı, müdahalesiz gözlem puanı, müdahalesiz gözlem sayısı, VOTAT uygulama sayısı, her bir kontrol çubuğu (üst-orta-alt) için VOTAT stratejisinin kullanıldığı durumların sayısı, ilk müdahaleye kadar geçen süre, diyagram tuşuna tıklanma sayısı, son kontrol süresi ve hızdır. Çalışmanın amacına bağlı olarak süreç verileriyle birlikte çeşitli sonuç verilerinin de KPÇ becerisini yordama düzeyleri incelenmiştir. Buna göre kullanılan sonuç verisi değişkenlerine aşağıda yer verilmiştir.

Çalışmada Kullanılan Sonuç Verisi Değişkenleri

Alan yazın incelendiğinde KPÇ becerisiyle ilgili çeşitli sonuç verilerinin kullanıldığı görülmüştür. Örneğin Greiff, Fischer ve arkadaşları (2013) çalışmalarında KPÇ ile akıl yürütme ve akademik başarı arasındaki ilişkiyi araştırmışlardır. Buna göre KPÇ ile akıl yürütme arasında orta düzey ilişki bulunmuş ve KPÇ akademik başarıyı (matematik, fizik, kimya, biyoloji notları) yordamıştır. Akademik başarı ve KPÇ arasındaki ilişkiye ilişkin benzer bir bulgu Wüstenberg ve arkadaşları (2012) tarafından da raporlanmıştır. KPÇ'nin yapısının araştırıldığı Sonnleitner ve arkadaşları (2013) tarafından yapılan çalışmada KPÇ ile akıl yürütme ve matematik, okuduğunu anlama, fen gibi farklı eğitim başarısı göstergeleri arasında anlamlı pozitif ilişkiler bulunmuştur. Buna ek olarak PISA 2012 değerlendirmesinde matematik, fen ve okuduğunu anlama ile genel problem çözme performansı arasında oldukça yüksek korelasyonlar raporlanmıştır (sırasıyla 0.81, 0.78, 0.75) (OECD, 2014a, s. 68). Bahsedilen bulgular ışığında KPÇ'nin çeşitli akademik başarı göstergeleriyle arasındaki ilişki dikkat çekmiştir. PISA 2012 problem çözme değerlendirmesinde öğrencilerin temel problem çözme süreçlerine dayalı olarak genel muhakeme becerilerine, problem çözme süreçlerini düzenleme yeteneğine ve bu süreçteki

istek düzeylerine odaklanmıştır (OECD, 2014a, s. 28). Bununla birlikte KPÇ’de yüksek bir yeterlik düzeyinin farklı alanlardaki problemlerle başarılı bir şekilde başa çıkılabilmesinde etkili olacağı da ifade edilmektedir (Greiff ve ark., 2018). Molnár ve arkadaşları (2013) tarafından yapılan bir başka araştırmada alana özgü problem çözme, tümevarımsal akıl yürütme ve karmaşık problem çözme arasındaki ilişki incelenmiştir. Buna göre alana özgü problem çözme ile karmaşık problem çözme arasında genel olarak 0.35 (kısmi korelasyon değeri 0.26) korelasyon raporlanmıştır. Buna ek olarak eldeki çalışmada dikkate alınan ve problemin çözümünde kilit rol oynayan VOTAT stratejisinin fen bilimlerindeki deneysel yöntemin merkezinde yer aldığı ifade edilmektedir (OECD, 2014a, s. 30). Bu sebeplerle eldeki çalışmada sonuç verisi olarak öğrencilerin PISA 2012 değerlendirmesinde raporlanan **matematik, fen ve okuduğunu anlama puanlarına ait birinci makul değerler** kullanılmıştır.

KPÇ bilişsel becerilerle birlikte motivasyonel, davranışsal ve duyuşsal yönleri de içermektedir (Greiff ve ark., 2018). Öte yandan PISA 2012 problem çözme tanımında geçen “...yapıcı ve yansıtıcı bir vatandaş olarak kişinin, potansiyeline ulaşmak için bu tür durumlara ilgilenme isteğini içerir (OECD, 2014a, s. 30)” ifadesinden hareketle karmaşık bir problemi çözmek için gerekli olabilecek bazı duyuşsal özelliklerin de ele alınmasının önemli olduğu görülmüştür. Örneğin; Scherer ve arkadaşları (2015) çeşitli bilişsel ve motivasyonel yapılarla KPÇ arasında anlamlı ilişkiler bulmuşlardır. Buna paralel olarak azim ve problem çözmeye açıklığın, özellikle en yüksek yeterlik düzeylerinde problem çözme performansı ile güçlü ilişki gösterdiği ifade edilmektedir (OECD, 2014a, s. 111). Öte yandan bir başka çalışmada üç farklı ülkenin PISA 2012 verilerine göre problem çözmeye açıklık ve azim değişkenlerinin KPÇ ile ilişkisi araştırılmıştır (Scherer ve Gustafsson, 2015). Sonuçta, her iki yapının da ülkeler arasında önemli farklılıklar olmayacak şekilde KPÇ ile pozitif ilişkide (0.25-0.36 arasında değişen) olduğu bulunmuştur. Bahsedilen sebeplere bağlı olarak eldeki çalışmada kullanılan diğer sonuç verileri PISA 2012 araştırmasında ölçülen **problem çözmeye açıklık ve azim** değişkenlerine ilişkin puanlar olmuştur. Sonuç olarak

çalışmada kullanılan sonuç verisi değişkenleri; matematik puanı, fen puanı, okuduğunu anlama puanı, problem çözmeye açıklık ve azim puanları olarak sıralanabilir. Değişkenlere ilişkin betimsel istatistikler, dağılım grafikleri ve korelasyon değerleri sırasıyla EK-A ve EK-B'de verilmiştir.

Karmaşık Problem Çözme Becerisinin ve Süreç Verisinin Kullanıldığı Araştırmalar

Wüstenberg ve arkadaşları (2012) çalışmasında KPÇ'nin yapı geçerliğini araştırmışlardır. Almanya'daki bir üniversitenin psikoloji bölümü öğrencileriyle çalışma gerçekleştirilmiştir. Buna göre MicroDYN yaklaşımıyla hazırlanmış KPÇ soruları ve bilgisayar ortamında akıl yürütme yeteneğinin ölçüldüğü bir testle veriler toplanmıştır. Buna ek olarak not ortalaması ve çeşitli demografik bilgiler de elde edilmiştir. Doğrulamalı faktör analizi ve Rasch analiziyle hesaplamalar yapılmıştır. Elde edilen sonuçlara göre, VOTAT kullanmadan doğru model oluşturma olasılığı (en kolay madde hariç) ortalama %3.4 olarak bulunmuştur. Öte yandan KPÇ için bilgi edinme ve bilgi uygulama olarak etiketlenmiş iki boyutlu model doğrulanmıştır. Aynı zamanda KPÇ'nin not ortalamasını yordadığı da ifade edilmiştir.

Greiff, Fischer ve arkadaşları (2013) tarafından yapılan çalışmada KPÇ'nin yapı geçerliği araştırılmıştır. Almanya'daki bir üniversitedeki psikoloji bölümü öğrencileriyle yapılan araştırmada çoklu özellik çoklu yöntem yaklaşımı kullanılmıştır. MicroDYN, MicroFIN ve Genetics Lab yöntemleriyle KPÇ becerisi ölçülmüştür. Buna ek olarak akıl yürütme ve akademik başarının KPÇ ile ilişkisi araştırılmıştır. Sonuç olarak akıl yürütme, KPÇ'nin bilgi edinme ve bilgi uygulama boyutunu yordamıştır. Öte yandan KPÇ doğa bilimlerindeki (matematik, fizik, kimya, biyoloji) akademik başarı varyansını açıklarken, sosyal bilimlerdeki (Almanca, tarih, coğrafya, sosyal bilgiler) başarıyı yordamamıştır. Buna ek olarak KPÇ'nin iki boyutu arasında yüksek bir ilişki olduğu görülmüştür. Sonuç olarak KPÇ'nin yakınsak ve iraksak geçerliğine dair yeterli kanıt bulunduğu vurgulanmıştır.

Greiff, Wüstenberg ve arkadaşları (2013) tarafından yapılan araştırmada KPÇ'nin yapısının araştırılması hedeflenmiştir. Bu bağlamda, bilgi edinme ve bilgi uygulama süreçleri incelenmiştir. Aynı zamanda genel zihinsel yetenek ve ebeveyn eğitimi arasındaki ilişkiler de araştırılmıştır. 5 ila 11. sınıflardaki Macar öğrencilerle yapılmış çalışmada 7 farklı MicroDYN görevi kullanılmıştır. Yapısal eşitlik modelleriyle değişkenler arasındaki ilişkilerin ortaya koyulduğu çalışmada, tüm sınıf düzeylerinde bilgi edinme ve bilgi uygulama olmak üzere iki boyutlu yapı doğrulanmıştır. Bu yapının varlığı gruplar arasında değişmezlik göstermiştir. Örtük ortalama karşılaştırmalarına göre sınıf seviyesi arttıkça iki boyutta gösterilen başarıda kısmen bir artış olmuştur. Ancak bu durum 9. sınıf öğrencileri için geçerli değildir. KPÇ'nin genel zihinsel yetenekle ilişkisi olduğu görülmüştür. Aynı zamanda ebeveyn eğitimi de KPÇ ve genel zihinsel yetenekle ilişkilidir.

Molnár ve arkadaşları (2013) alana özgü problem çözme (AÖPÇ) (özellikle matematik alanı) tümevarımsal akıl yürütme (TAY) ve KPÇ becerileri arasındaki ilişkileri ve bu becerilerin gelişim düzeylerini oldukça kapsayıcı bir örnekleme incelemişlerdir. Buna göre Macaristan'da seçilen ilk ve ortaokullarda eğitim gören 3 ila 11. sınıf arasındaki bireyler (her grupta 300-400 öğrenci) çalışmaya dahil edilmiştir. Tüm öğrenciler birer tane TAY ve AÖPÇ testi çözerken, 5 ila 11. Sınıf öğrencileri ek olarak KPÇ testini çözmüşlerdir. Buna göre TAY ve AÖPÇ için 2769, KPÇ için 788 kişinin verileriyle analizler gerçekleştirilmiştir. TAY ve AÖPÇ testleri çoktan seçmeli ve açık uçlu formatta verilmiştir. KPÇ görevleri ise MicroDYN yaklaşımına göre tasarlanmış yedi görevle ölçülmüştür. Çalışmada elde edilen sonuçlardan bazılarına burada yer verilmiştir. Buna göre her bir beceri sınıf düzeyleri arasında lojistik birer eğriye uygun olacak şekilde gelişim göstermiştir. Üç yapı arasında 0.35 ile 0.44 arasında değişen orta düzeyde ilişkiler bulunmuştur. Dolayısıyla üç yapının ilişkili ancak farklı yapılar olduğu görülmüştür. Alana özgü problem çözme ve KPÇ arasındaki ilişki 0.35 (kısmi korelasyon 0.26) olarak bulunmuştur.

Sonnleitner ve arkadaşları (2013) Genetics Labs problem yaklaşımıyla KPÇ'nin yapısını araştırmışlardır. Lüksemburg'da seçilen bir okuldaki 11. sınıf öğrencileriyle yapılan

araştırmada 12 tane KPÇ sorusu kullanılmıştır. Sonuçta KPÇ'nin matematik, okuduğunu anlama, fen puanı olarak ele alınan eğitim başarısı göstergeleri ve akıl yürütme ile pozitif ilişkili olduğu bulunmuştur. Aynı zamanda çalışmada KPÇ'nin kural tanımlama, kural bilgisi ve kural uygulaması olarak üç boyutlu olduğu bir yapı olarak ya da alt boyutları kapsayan genel bir KPÇ faktörünü içeren hiyerarşik bir yapı olarak kavramsallaştırılabileceği ancak bu iki model arasında net bir tercih yapılamadığı vurgulanmıştır.

Wüstenberg ve arkadaşları (2014) tarafından yapılan araştırmada bireylerin KPÇ bağlamında VOTAT hakkındaki bilgileri, akışkan zekâ düzeyleri ve öğrenme yönelimlerinin VOTAT stratejisini uygulama üzerindeki etkilerinin incelenmesi amaçlanmıştır. Finlandiya'da seçilen okulun 6. ve 9. Sınıf öğrencileriyle araştırma yapılmıştır. Kâğıt-kalem tabanlı bir testle VOTAT bilgisi ölçülürken, VOTAT uygulaması 9 tane MicroDYN tabanlı KPÇ göreviyle ölçülmüştür. Doğrulayıcı faktör analizine dayalı çeşitli modeller kurulmuştur. Sonuçta; öğrenme yönelimi, akışkan zekâ ve VOTAT bilgisinin, VOTAT uygulanması ile önemli düzeyde ilişkili olduğu bulunmuştur. Buna ek olarak VOTAT uygulama değişkeni, öğrenme yönelimi ve VOTAT bilgisi değişkenlerinin KPÇ performansı üzerindeki etkisine aracılık etmiştir.

Greiff, Wüstenberg ve arkadaşları (2015) tarafından yapılan çalışmada PISA 2012 problem çözme değerlendirmesindeki klima kontrolü ünitesinin birinci maddesindeki performans, genel problem çözme performansı ve kullanılan strateji arasındaki ilişkinin araştırılmasında günlük dosyası analizlerinin potansiyelinin ortaya koyulması amaçlanmıştır. 44 ülkeden katılan bireylerin verilerinin kullanıldığı çalışmada VOTAT kullanımı, etkileşim sayısı, yanıt süresi, ilk etkileşime kadar geçen süre, diyagram çizimlerinin sayısı, sıfır tur sayısı, ilk sıfır tura kadar geçen süre değişkenleri oluşturulmuştur. Çalışmada genel KPÇ performansı olarak birinci makul değer kullanılmıştır. Sonuç olarak, VOTAT stratejisini uygulamanın birey ve ülke düzeyinde maddedeki performansla ve genel KPÇ performansı ile yüksek düzeyde ilişkili olduğu bulunmuştur. Aynı zamanda çalışmada VOTAT uygulayan ve başarılı olan, VOTAT'ı

kısmen anlayan ve başarısız olan, VOTAT'ı uygulamayan ve başarısız olan öğrencilerin ortaya çıkarılmış olmasının günlük dosyalarının analizlere kattığı zenginlik olarak vurgulandığı görülmüştür. Bunlara ek olarak, çalışmanın ek materyalinde raporlanan sonuçlara göre, eylem sayısının küçük ve pozitif, diğer oluşturulan değişkenlerin küçük ve negatif etkiye sahip olduğu da bulunmuştur.

Scherer ve arkadaşları (2015) tarafından yapılan çalışmada KPÇ'de görev süresi ve KPÇ yeteneği arasındaki ilişki araştırılmıştır. Süre ve yetenek bilgisine ek olarak çeşitli bilişsel ve motivasyonel yapılar, okul başarısı da çalışmada incelenmiştir. Finlandiya'da seçilen bir okuldaki 9. sınıf öğrencileriyle 9 farklı MicroDYN tabanlı görev üzerinden gerçekleştirilen çalışmada ilişkiler Bayesian tepki süresi MTK modeli ve yapısal eşitlik modelleriyle araştırılmıştır. Sonuç olarak süre ve yetenek arasında orta düzeyde pozitif ilişki bulunmuştur. Muhakeme, süre ve yetenek arasında da pozitif anlamlı ilişkiler olduğu gözlenmiştir. Hedef yönelimi sadece süreyle pozitif ilişki gösterirken, performans yönelimi sadece süre ile negatif ilişki göstermiştir. Öte yandan KPÇ'de yüksek performans gösterenlerin okul notlarının da daha yüksek olduğu görülmüştür.

Scherer ve Gustafsson (2015) araştırmasında azim ve problem çözmeye açıklık değişkenlerinin faktör yapılarını, ülkelere göre ölçme değişmezliğini ve bu yapıların genel problem çözme becerisiyle (hem karmaşık hem de statik maddelerle ölçülmüştür) ilişkisini incelemiştir. Bu amaçla PISA 2012 değerlendirmesine ait veriler kullanılmıştır. Avustralya, Singapur ve Norveç verilerinin kullanıldığı çalışmada ilişkiler yapısal eşitlik modelleriyle araştırılmıştır. Elde edilen sonuçlara göre, problem çözmeye açıklık ve azim yapıları arasında orta düzey (0.47) ilişki bulunmuştur. Öte yandan skalar değişmezlik sağlanmıştır. Her iki yapı ise genel problem çözme ile ülkeler arasında önemli farklılıklar göstermeyecek şekilde pozitif ilişkilidir (0.25-0.36 arasında değişen).

Greiff ve arkadaşları (2016) tarafından yapılan çalışmada KPÇ ile ilgili bilgileri genişletmek adına günlük dosyası verilerinin potansiyelinin araştırılması amaçlanmıştır. Buna bağlı olarak çeşitli davranış göstergeleriyle KPÇ'nin iki alt boyutu (bilgi edinme ve bilgi

uygulama) arasındaki ilişki incelenmiştir. Finlandiya’da gerçekleştirilen bir uygulamayı içeren çalışmada 9 tane MicroDYN temelli KPÇ görevi kullanılmıştır. Çeşitli yapısal eşitlik modelleri ile incelemeler yapılmıştır. Sonuç olarak VOTAT kullanımının her iki boyuttaki performansla pozitif ve güçlü ilişkide olduğu bulunmuştur. Süre ile iki boyut arasında ters U şeklinde bir ilişki olduğu sonucuna ulaşılmıştır. Öte yandan düşük sıklıkta müdahalede bulunan (“uygula” olaylarının sayısı) ve müdahalesiz gözlemlerde (en az bir kez tüm kontrol çubukları 0 değerinde iken “uygula” tuşuna tıklanması durumu) bulunanların görev performansının daha yüksek olduğu vurgulanmıştır.

Greiff ve arkadaşları (2018) tarafından yapılan çalışmada KPÇ’de VOTAT kullanımının derinlemesine araştırılması hedeflenmiştir. Macaristan’daki bir okulda öğrenim gören 6 ila 8. sınıf öğrencilerinden oluşan bir örnekleme 9 farklı görev ile veriler elde edilmiştir. Görevler içinden MicroDYN yaklaşımına uygun olan 6 görevde VOTAT kullanımı araştırmak için örtük sınıf analizi kullanılmıştır. Sonuç olarak 6 farklı sınıf ortaya çıkmıştır. 6. sınıftan 8. sınıfa doğru daha iyi performans gösteren sınıf profillerinde küçük bir ilerleme olmakla birlikte, bu bağlamda cinsiyet farklılığı gözlenmemiştir. VOTAT kullanımı ve genel KPÇ puanı arasında pozitif ilişki bulunmuştur.

Xu ve arkadaşları (2018) tarafından yapılan çalışmada ham günlük dosyası verilerinden yararlanarak benzer davranış kalıplarına sahip grupları keşfetmek ve bilgilendirici olay türlerini ortaya çıkarmak amaçlı tekrarlayan olayları analiz etmek için örtük sınıf analizi ve olay geçmişi analizini birleştiren bir model önerilmesi amaçlanmıştır. PISA 2012 klima kontrolü ünitesi 1. maddesini yanıtlayan bireyler analize dahil edilmiştir. Modellemede “uygula” tuşuna tıklandığında üç kontrol sürgüsünün tüm olası durumlarını içeren olaylar kullanılmıştır. Sonuç olarak önerilen modeller, problem çözme kalıplarına göre bireyleri başarılı bir şekilde gruplayabilmiştir.

Eichmann ve arkadaşları (2019) tarafından yapılan araştırmada KPÇ sürecinde planlama davranışının araştırılması amaçlanmıştır. PISA 2012 problem çözme değerlendirmesine katılan Alman öğrencilere ait 21 tane karmaşık problemin günlük

dosyaları kullanılmıştır. Planlama aralıklarının uzunluğu, en uzun aralığın gecikme süresi ve planlama süresindeki değişim değişkenlerinin oluşturulduğu çalışmada, genelleştirilmiş doğrusal karma modellerle analizler gerçekleştirilmiştir. Sonuçta en uzun aralığın gecikmesi göstergesi için anlamlı negatif etki bulunmuş olması dikkat çekicidir. Başka bir deyişle KPÇ'de erken planlama davranışının faydalı olduğu ifade edilmiştir. Aynı zamanda üç gösterge için de göreve ve etkileşime bağlı etkiler olduğu raporlanmıştır.

Chen ve arkadaşları (2019) tarafından yapılan araştırmada günlük dosyası verilerinden yararlanarak KPÇ'de görevi bitirme süresi ve sorudaki nihai sonucun (başarı-başarısızlık) tahminini sağlayacak istatistiksel model geliştirilmesi amaçlanmıştır. Olay geçmişini analizinin kullanıldığı çalışmada PISA 2012 problem çözme değerlendirmesindeki klima kontrolü ünitesinin birinci maddesine ait veriler kullanılmıştır. Katılımcı tüm ülkelerin dahil edildiği çalışmada herhangi bir zaman noktasına kadar gerçekleşen olayların analiziyle nihai sonuç ve sürenin koşullu dağılımı modellenmiştir. Çalışmada çeşitli süreç verisi değişkenlerine ilişkin dikkat çekici bulgular elde edilmiştir. Buna göre VOTAT stratejisinin nihai sonuç üzerinde pozitif, süre üzerinde negatif etkiye sahip olduğu bulunmuştur. Basit eylemlerde bulunma (örneğin, kontrol sürgüsünü hareket ettirme) göstergesi ve sıklığı sonuç üzerinde pozitif etkiye sahipken, bu eylemlerde bulunmak daha kısa süreye yol açmıştır. Ancak tüm basit eylemlerde bulunduktan sonra bu eylemlerin sıklığının artması süre üzerinde zayıf ancak negatif bir etkiye yol açmıştır. Yönergeyi daha hızlı okuyan bireylerin ise görevi daha düşük doğrulukta ve daha kısa sürede tamamlama eğiliminde olduğu bulunmuştur. Son olarak sıfırlama tuşunun kullanımı ise hem süre hem de sonuç üzerinde pozitif etkiye sahiptir.

Han ve arkadaşları (2019) çalışmasında süreç verilerinden özellik elde edilmesi ve bu özelliklerden yararlanarak performansı yordamada önemli değişkenlerin seçilmesine yönelik iki aşamalı bir yaklaşım önerisinde bulunmuşlardır. Buna göre PISA 2012 problem çözme değerlendirmesi klima kontrolü ünitesinin birinci maddesine ait günlük dosyası ve madde doğruluğu bilgileri kullanılmıştır. Tüm katılımcı ülke ve ekonomilerden toplam 30.224

kişinin verilerinin kullanıldığı araştırmada üç farklı kategoride süreç verisine dayalı özellikler elde edilmiştir. Bu bağlamda n-gram (ardışık eylem dizileri) yaklaşımıyla biagram ve triagramlar (iki ve üç sıralı ardışık eylemi içeren mini eylem dizileri), çeşitli davranış göstergeleri, zamana dayalı 77 özellik elde edilmiştir. Madde performansını yordamada önemli değişkenlerin seçiminde rastgele orman algoritmasıyla analizler gerçekleştirilmiştir. Sonuç olarak pek çok değişkenin etkisi ortaya koyulmuştur. Bunlardan VOTAT puanı (VOTAT stratejisinin kullanıldığı kontrol çubuğu sayısı) ve VOTAT uygulama sıklığı değişkenlerinin önemli değişkenler arasında olması dikkat çekmiştir.

Ren ve arkadaşları (2019) tarafından yapılan araştırmada günlük dosyası verilerinden yararlanarak KPÇ için çoklu hedef dengeleme stratejisine ilişkin ampirik kanıt sunulması hedeflenmiştir. Bu amaçla Avusturya, Japonya, Avustralya, İrlanda, Almanya ve Fransa'nın PISA 2012 problem çözme değerlendirmesindeki bilet ünitesinin üçüncü maddesine ilişkin günlük dosyası verilerinden yararlanılmıştır. Hedef arayışını temsil edecek davranışları içeren rubrik tasarlanmış ve R yazılımında buna göre otomatik puanlama yapan bir algoritma geliştirilmiştir. Kümeleme analizi ve regresyon analizlerinin kullanıldığı çalışma sonucunda bireyler hedef takip göstergelerine göre 10 anlamlı kümeye ayrılmıştır. Oluşturulan göstergelerin genel KPÇ yeterliğini ve görev puanını anlamlı olarak yordadığı sonucuna ulaşılmıştır. Sonuç olarak KPÇ'de çoklu hedef dengeleme stratejisinin değerlendirilmesinin önemine vurgu yapılmıştır.

Stadler ve arkadaşları (2019) tarafından yapılan bir başka çalışmada KPÇ bağlamında bir probleme VOTAT stratejisini uygulayan ancak doğru çözüme ulaşamayan bireylerle, stratejiyi doğru uygulayan ve doğru çözüme ulaşan bireyler arasındaki davranış farklılıklarını incelemek amacıyla günlük dosyalarının analizi gerçekleştirilmiştir. Finlandiya'daki yerel bir uygulamadan elde edilen veriler kullanılmıştır. n-gram yaklaşımıyla puan, süre ve VOTAT kullanımına ilişkin değişkenler elde edilerek veri madenciliği yöntemlerinden ki-kare özellik seçim modeli kullanılmıştır. Sonuçta süreç verisi değişkenleri ile ilgili bir dizi bulguya ulaşılmıştır. Buna göre VOTAT uygulanmasının genel olarak

problemi çözme olasılığını önemli ölçüde artırdığı bulunmuştur. Genel davranış örüntüsü olarak görevde başarılı olanların senaryo üzerinde daha az, model üzerinde daha fazla inceleme yaptıkları görülmüştür. Öte yandan görevde başarılı olamayanların ise bunun tam tersi bir yaklaşım izleyerek bulguları model üzerinde hemen göstermedikleri ve buna bağlı olarak da bilişsel yüklerini artırıp görevin zorluğunu pekiştirdikleri ifade edilmiştir. Öte yandan çalışmada problemi doğru çözemeyenlerin ya stratejiyi anlamadıkları ya da stratejinin kullanımını yönetemedikleri vurgulanmıştır. Buna ek olarak sıfırlama ve yardım tuşlarının kullanımının ek bilgi sağlamadığı için dikkate alınmadığı raporlanmıştır.

Eichmann ve arkadaşları (2020) davranış dizilerinin KPÇ’de başarı ya da başarısızlıkla ilişkisini araştırmışlardır. PISA 2012 problem çözme değerlendirmesindeki klima kontrolü ve bilet ünitelerine ait günlük dosyası verileri analiz edilmiştir. Tam yol dizi analizi ile ilişkilerin araştırıldığı çalışmada tam davranış dizilerinin benzerliklerine göre kümelenecek davranış kalıplarının ortaya çıkarılması sağlanmıştır. Çeşitli davranış göstergeleri elde edilmiştir. Ki-kare testleri ve hiyerarşik kümeleme analizleri yapılmıştır. Genel olarak doğru yanıtlar arasında yüksek oranda hedefe yönelik davranış, yanlış yanıtlar arasında da yüksek oranda hedefe yönelik olmayan keşif gözlenmiştir. Buna ek olarak her iki görevde de sıfırlama tuşunun kullanımı doğru yanıtlarda daha sık görülmüştür.

Pejić ve Molcer (2021) tarafından yapılan araştırmada çeşitli makine öğrenme algoritmalarıyla günlük dosyası verilerinden yararlanarak KPÇ’ye ilişkin performans tahmini yapılması amaçlanmıştır. Buna göre PISA 2012 problem çözme değerlendirmesindeki klima kontrolü ünitesi birinci maddesine ilişkin tüm ülkelerdeki katılımcıların günlük dosyası verileri kullanılmıştır. Çalışmada VOTAT stratejisine ve zamana dayalı çeşitli özellikler elde edilmiştir. İlk olarak, her bir kontrol çubuğu için VOTAT stratejisinin uygulanma sayısının toplam “uygula” tuşuna tıklanma sayısına oranlanmasıyla üç farklı özellik elde edilmiştir. Ayrıca toplam “uygula” tuşuna tıklanma sayısı da başka bir değişken olarak ele alınmıştır. Zamana dayalı olarak ise görevi okumak için harcanan göreceli süre, kontrol çubuklarıyla deney yapmak için harcanan göreceli süre, diyagram çizimi için harcanan göreceli süre, görevde

geçirilen toplam görelî süre ve “uygula” tuşuna tıklanma sayısına göre elde edilen hız değişkenleri oluşturulmuştur. Navie Bayes, lojistik regresyon, derin öğrenme, gradyan artırma, sınıflandırma ağacı ve rastgele orman algoritmalarıyla yapılan analizler karşılaştırmalı olarak incelenmiştir. Sonuç olarak, her üç kontrol çubuğunu eşit şekilde deneme eğiliminde olanların görevde en fazla başarıya sahip bireyler oldukları görülmüştür. VOTAT tabanlı özellikler açısından Navie Bayes ve lojistik regresyon yöntemlerinde en önemli özellik orta kontrol çubuğunun kullanımı ile ilgili özellik iken, derin öğrenme yöntemine göre alt kontrol çubuğunun kullanımı en önemli, orta kontrol çubuğunun kullanımı ile ilgili özellik en az önemli olarak bulunmuştur. Öte yandan genel olarak hem zamana dayalı özellik setiyle hem de VOTAT tabanlı özellik setiyle yapılan tüm analizler iyi performans göstermiştir. Bununla birlikte VOTAT tabanlı özellikleri içeren veri setiyle yapılan analizlerin sonuçları ise diğerine göre daha iyi performans göstermiştir. Genel olarak her iki özellik seti için derin öğrenme algoritması en uygun makine öğrenme algoritması olmuştur.

Tang ve arkadaşları (2021) tarafından yapılan araştırmada süreç verilerini işlemek ve analiz etmek için R yazılımında paket geliştirilmesi hedeflenmiştir. Bu amaçla istatistiksel yöntemler ve makine öğrenme yöntemlerini birlikte uygulayan ProcData paketi geliştirilmiştir. Paketin fonksiyonlarıyla PISA 2012'deki klima kontrolü maddesine ait günlük dosyası analiz edilmiştir. Bu paketteki fonksiyonlarla süreç verilerinden özellik çıkarma, düzensiz bilgileri düzenli sayısal vektörlere dönüştürme, sinir ağı tabanlı modellerden tahminler yapma işlemleri gerçekleştirilebileceği ifade edilmiştir.

Zhan ve Qiao (2022) çalışmalarında süreç verilerinden eylemle ilgili olanların tanısal sınıflandırma yaklaşımıyla analiz edilmesine imkân tanıyan madde genişletme yöntemi önermişlerdir. Bunun için PISA 2012 problem çözme değerlendirmesindeki bilek ünitesinden bir soruya ait ABD, Singapur, Türkiye ve Avusturya verilerini kullanmışlardır. Çalışmada sonuç verileri (madde doğruluğu) kısmi puan modeliyle analiz edilmiştir. Ardından bu modelle kestirilen yetenek puanıyla en iyi uyumu sağlayan tanısal sınıflandırma modelinden

kestirilen yetenek karşılaştırılmıştır. Q matrisi hazırlanırken oluşturulan eylem dizileri hayali maddeler olarak ele alınmıştır. Bu eylem dizilerini oluşturmak için gereken problem çözme becerileri de örtük özellikler olarak değerlendirilmiştir. Geliştirilen yöntem hem yetenek kestirimini hem de sınıflandırmayı başarılı bir şekilde gerçekleştirmiştir. Buna ek olarak süreç verileriyle yapılan kestirimin sadece sonuç verisine dayalı bir kestirime göre bireyin yetenek düzeyini daha doğru yansıttığı sonucuna ulaşılmıştır.

Csányi ve Molnár (2023) çalışmalarında KPÇ'ye dayalı bir testte gösterilen çabayı, günlük verileri ve öz bildirim anketlerinden elde edilen verilerle araştırmışlardır. Türkiye'deki bir üniversitenin birinci sınıf öğrencileriyle gerçekleştirilen çalışmada k-ortalamlar kümeleme analizi, korelasyon değerleri ve tekrarlı ölçümler için ANOVA ile bulgular elde edilmiştir. Buna göre testte gösterilen çaba anlamında 3 grup bulunmuştur. Öz bildirim verilerine göre elde edilen sınavdaki çaba ve performans arasındaki ilişki gerçek sınav davranışına kıyasla daha zayıf bulunmuştur. Tıklama sayısı performansı tahmin etmede en önemli yordayıcıdır. Süre ile performans arasında ise bir ilişki bulunmamıştır.

Fu, Chen ve arkadaşları (2023) araştırmalarında eylem dizilerinin (görevin tamamlanması için sergilenen bir dizi eylem ya da durum geçişi) iki kategorili olarak modellenmesini sağlayacak bir ve iki parametrelili eylem dizisi modeli önermişlerdir. Bu amaçla hem PISA 2012 problem çözme değerlendirmesindeki bilek ünitesinin verileri kullanılmış hem de simülasyon çalışması yapılmıştır. Geliştirilen modellerin sıralı tepki modeliyle karşılaştırması yapılarak performansı değerlendirilmiştir. Elde edilen sonuçlara göre, geliştirilen modellerin yetenek tahminindeki performansı sıralı tepki modeliyle uyumlu iken, hesaplama süresi açısından daha hızlıdır. Buna ek olarak iki modelin parametreleri yüksek doğrulukta kestirdiği bulunmuştur. Ayrıca eylem dizisi uzunluğu arttıkça kestirimlerin doğruluğunun da arttığı görülmüştür. Sonuçta geliştirilen modellerin eylem dizisi verilerinin analizinde etkili olduğu ifade edilmiştir.

Fu, Zhan ve arkadaşları (2023) çalışmasında eylem süresi ve eylem dizilerinin eylem düzeyinde birlikte modellenmesini gerçekleştirmişlerdir. Buna göre eylem dizileri için sıralı

tepki modeli eylem süresi için de yeni bir lognormal eylem süresi modeli kullanılmıştır. PISA 2012 problem çözme değerlendirmesindeki bilek görevine ait veriler kullanılmıştır. Sonuçta geliştirilen modelin parametrelerinin yorumlanabildiği, tahminlerin doğru olduğu bulunmuştur. Çalışmada eylem dizileri ve sürelerinin birlikte ele alınmasının davranış kalıpları hakkında daha derinlemesine bilgi verdiğine dikkat çekilmiştir.

Xu ve arkadaşları (2024) çalışmalarında çift kümeleme (biclustering) yaklaşımıyla süreç verilerindeki eylem ve zamana dayalı özelliklere göre bireyleri eş zamanlı gruplandırmayı amaçlamışlardır. Bu amaçla PISA 2012 problem çözme değerlendirmesindeki bilek ünitesine ait bir görevin günlük dosyaları kullanılmıştır. Buna ek olarak bireylerin sorudaki performanslarıyla, oluşturulan kümeler arasındaki ilişkiler de incelenmiştir. Buna göre ABD, Singapur, Avusturya ve Türkiye'den 3760 öğrencinin verileri kullanılmıştır. Ham süreç verilerinden zamana dayalı 4 özellik, eylem dizilerine dayalı 32 özellik çıkarılmıştır. Çift kümeleme algoritmalarından BIMAX (ikili kapsama-maksimum/binary inclusion-maximal), FABIA (Laplace önceliği ile çift kümelemeli kazanç için faktör analizi/factor analysis for bicluster acquisition with Laplace prior), Spectral çift kümeleme, BCCC (Cheng ve Church çift kümeleme algoritması (Cheng and Church's biclustering algorithm) kullanılmıştır. Sonuçlar aynı zamanda tek modlu yaklaşımlardan k-ortalamlar kümeleme analiziyle de karşılaştırılmıştır. Yöntemlerin performansları VAR (çift küme içi varyans/within-biclustervariance), MSR (artık kareler ortalaması/mean squared residue) ve sanal hata (virtual error) metriklerine göre değerlendirilmiştir. Buna göre VAR metriğine göre tüm çift kümeleme yaklaşımları, tek modlu yaklaşıma kıyasla daha iyi performans göstermiştir. Başka bir deyişle çift kümeleme yaklaşımlarının daha homojen yapılar üretme eğiliminde olduğu ifade edilmiştir. Bununla birlikte çift kümeleme yaklaşımları kendi içinde karşılaştırıldığında VAR metriğine göre BIMAX, MSR ve sanal hata metriklerine göre BCCC daha iyi performans göstermiştir. Çalışmada BCCC'ye göre elde edilen kümelerle maddedeki performans arasında çeşitli ilişkiler de bulunmuştur. Buna göre maddeden kısmi puan alanların eylem dizisi örüntüleri orta düzeyde homojenlik gösterirken,

yanlış yanıt veren bireylerin belirli eylemlere takılıp kaldıkları veya sıklıkla yanlış butonları denedikleri görülmüştür. Buna ek olarak zamana dayalı özelliklerin de davranışların anlaşılmasına önemli katkısı olduğu ifade edilmiştir. Sonuç olarak, çift kümeleme yaklaşımlarının problem çözmedeki davranış kalıpları hakkında daha derinlemesine bilgi verme potansiyeline vurgu yapılmıştır.

KPÇ ve süreç verisi ile ilgili araştırmalar birlikte değerlendirildiğinde bazı ortak hedefleri paylaştıkları dikkat çekmiştir. Buna göre; KPÇ'nin yapısı ve bu yapının incelenmesinde günlük dosyalarının potansiyelinin incelenmesi, KPÇ sorularını çözmede kullanılan stratejiler, günlük verilerinden elde edilen bilgilerin çeşitli amaçlarla modellenmesi, ardışık olay dizilerinin modellenmesi, makine öğrenme algoritmalarının karşılaştırılması ile ilgili başlıklarda araştırmaların olduğu görülmüştür. Erken dönemli araştırmalar genellikle yapının anlaşılmasıyla ilgiliyken yakın dönemli araştırmalarda yeni modelleme çalışmalarının olması dikkat çekmektedir. Öte yandan bu çalışmada olduğu gibi KPÇ, süreç verisi ve regresyon ağaçları temelli makine öğrenme algoritmalarının bir arada ele alındığı bir çalışmaya bilindiği kadarıyla rastlanmamıştır.

Süreç ve Sonuç Verilerinin Farklı Yapılarla ve Bağlamlarda Kullanıldığı Araştırmalar

Deane'nin (2014) çalışması süreç ve ürün verilerinin birlikte kullanıldığı bir çalışma olarak dikkat çekmiştir. Kompozisyon puanlarının yazı kalitesi, okuma, dinleme puanlarıyla ilişkisinin belirlenmesinin amaçlandığı çalışmada Educational Testing Service (ETS) tarafından yapılan CBAL (Cognitively Based Assessments of, for, and as Learning) araştırması kapsamında ABD örneklemine ait veriler kullanılmıştır. Bu amaçla biri ikna edici diğeri edebi formatta olan iki kompozisyon değerlendirmesi kullanılmıştır. 10 tane ürün (e-rater sisteminden elde edilen) ve 7 tane süreç (günlük dosyalarından elde edilen) özelliği elde edilmiştir. Her bir veri türü için açıklayıcı faktör analizleri yapılmıştır. Ardından ürün, süreç faktörlerinin ve kaynak kullanımı ile ilgili iki özelliğin yazma kalitesini yordama düzeyleri incelenmiştir. Çeşitli bulguların elde edildiği çalışmada süreç ve ürün özelliklerinin,

görev türüne göre bazı farklılıklar içermekle birlikte, istikrarlı faktör yapılarına (her bir veri seti 3 faktörlü olarak bulunmuştur) sahip olduğu bulunmuştur. Süreç ve ürün özelliklerinin birlikte ele alınarak yapıldığı regresyon analizlerinde ürün özelliklerinin daha yüksek regresyon katsayılarına sahip olduğu görülmüştür. Bununla birlikte çalışmada süreç faktörleri ve kaynak kullanımı ile ilgili özelliklerin, yalnızca ürün faktörleri tarafından doğrudan yakalanamayan yönlerin ortaya çıkarılmasına imkân sağladığı ve buna bağlı olarak da süreç ve ürün özelliklerinin birlikte yazma profillerini ortaya koymak için kullanılabileceğine dikkat çekilmiştir.

Süreç ve ürün özelliklerinin birlikte kullanıldığı Zhang ve Deane (2015) tarafından yapılan çalışmada bu iki veri türünün yazma performansı ile ilişkisi incelenmiştir. ETS-CBAL değerlendirmesi kapsamında ABD örneklemine ait verilerin kullanıldığı çalışmada Cömert Hediye formuna ait bilgilerden yararlanılmıştır. 29 süreç ve 10 tane ürün özelliği elde edilmiştir. Faktör yapıları ve yordama düzeylerinin incelendiği çalışmada, süreç özellikleri insan-puanlayıcı (human-rater) tarafından yapılan puanlamaları tahmin etmede ürün verilerine göre daha sınırlı performans göstermişlerdir. Öte yandan elde edilen sonuç yorumlanırken, insan-puanlayıcı tarafından yapılan puanlamaların süreç özelliklerinden ziyade ürün özelliklerini içeren puanlama anahtarına göre yapıldığının da göz önünde bulundurulması gerektiği vurgulanmıştır.

Scalise ve Clarke Midura (2018) çalışmasında Şehirde Yeni Bir Kurbağa Var isimli simülatif bir video oyununun günlük dosyası verilerinden yararlanarak bilimsel sorgulama becerisinin incelenmesini amaçlamışlardır. Çalışmada tek boyutlu kısmi puan modeli, iki boyutlu kısmi puan modeli, Bayes ağları ve çok boyutlu MTK modelini içeren hibrit bir yaklaşım olan çok boyutlu MTK-Bayes ile modellemeler yapılmıştır. Sonuçta, tek ve iki boyutlu modelden iki boyutlu olan daha iyi uyum göstermiştir. Çok boyutlu MTK-Bayes modeline ise bazı süreç verileri dahil edilmiştir. Süreç verilerinin eklenmesiyle bilimsel sorgulama becerisinin her iki boyutunda da yapılan kestirimlerin güvenilirlikleri artmıştır.

Tang ve arkadaşları (2020) çalışmasında günlük dosyalarından çok boyutlu ölçekleme yoluyla değişken elde edilmesini amaçlamışlardır. Hem simülasyon çalışması yapılmış hem de PIAAC 2012 bilgisayar tabanlı problem çözme değerlendirmesine ilişkin verilerle analizler gerçekleştirilmiştir. Çalışmada çeşitli süreç değişkenleri elde edilmiştir. Her süreç değişkeni için görevde başarılı olup olunmadığına dair bilgi veren iki kategorili yanıtlar da mevcuttur. Özellikler ve yanıt doğruluğunu karşılaştırmak için aritmetik, okuduğunu anlama puanı, demografik özellikler ve bazı arka plan anketlerinden elde edilen özellikler kullanılmıştır. Başka bir deyişle bahsedilen bu değişkenlerin özellikler ve yanıt doğruluğuna dayalı tahminleri incelenmiştir. Sonuçta, süreç verisinden elde edilen özelliklerle yapılan kestirimin, yanıt doğruluğuna dayalı yapılanlara göre daha iyi performans gösterdiği bulunmuştur. Kullanılan yöntemin de bireylerin performanslarıyla ilgili daha ayrıntılı bilgi içerdiği ve makul düzeyde yorumlanabilirliğe sahip olduğu vurgulanmıştır.

Reis Costa ve arkadaşları (2021) çalışmasında PISA 2012 bilgisayar tabanlı matematik değerlendirmesindeki maddelerde harcanan süre ile maddeye verilen yanıtları birlikte modellemeyi ve bu modelin ülkelere göre ölçme değişmezliğini incelemeyi amaçlamışlardır. Tüm ülkelerin verilerinin dahil edildiği analizlerde 10 madde kullanılmıştır. Çeşitli modellerin kurulduğu çalışmada temel olarak, yanıt sürelerinin dahil edilmesiyle yetenek tahmininin kesinliğinin arttığı vurgulanmıştır. Öte yandan zayıf ölçme değişmezliği bulunmuştur.

Zhang ve arkadaşları (2023) çalışmalarında süreç verisinden elde edilen bilgilerin puanlamaya dahil edilerek yetenek kestirimi sağlayan yöntem geliştirmeyi ve bu yöntemin güvenilirliğini araştırmayı hedeflemişlerdir. Bu amaçla hem simülasyon verisi hem de PIAAC 2012 problem çözme verisi kullanılmıştır. Temel olarak çok boyutlu ölçeklemeye dayanan Rao-Blackwellization yaklaşımıyla istatistiksel formül geliştirilmiştir. Puanlama için ardışık eylem dizileri ve eylem türleri kullanılmıştır. Sonuçta geliştirilen otomatik puanlama yönteminden elde edilen puanların genelleştirilmiş kısmi puan modeliyle yapılan yetenek kestirimlerine göre daha yüksek güvenilirlikte olduğu gösterilmiştir. Buna ek olarak, süreç

verisinin puanlamaya dahil edilmesiyle, daha kısa testlerde de güvenilir yetenek kestirimleri yapılabileceği vurgulanmıştır.

Süreç ve sonuç verilerinin bir arada ele alındığı çalışmalar incelendiğinde genel olarak süreç verilerinin dahil edildiği kestirimlerin, yalnızca sonuç verilerine göre gerçekleştirilenlere kıyasla daha yüksek kesinlik ve doğrulukta olduğuna yönelik bulguların olduğu dikkat çekmektedir. Öte yandan bu iki veri türünün birlikte kullanımının önerildiği çalışmalarda süreç verilerinin de dahil edilmesiyle ölçülen örtük özelliğe ilişkin daha detaylı ve farklı bilgiler elde edilebileceğine vurgu yapılmıştır.

Veri Madenciliği (Makine Öğrenme Algoritmaları)

Günümüzde çeşitli kararlar yapay zekâ ve veri bilimindeki gelişmelerle birlikte karmaşık otomatik sistemler vasıtasıyla alınmaktadır (Redelmeier ve ark., 2020). Bu bağlamda genellikle çeşitli girdi verilerine dayalı olarak bir kararın/sonucun tahmin edilmesi yaklaşımıyla çalışan makine öğrenme modellerinin çalışma prensiplerinin anlaşılması giderek daha önemli hale gelmektedir (Redelmeier ve ark., 2020). Bu kapsamda dikkat çeken bir konu alanı olarak veri madenciliği ise büyük veri setlerini analiz etmek amacıyla bilgisayar bilimleri ve istatistikte geliştirilen teknikler için kullanılan genel bir kavramdır. İstatistiksel öğrenme, örüntü tanıma ve makine öğrenmesi terimleri de bu kavramla aynı anlamda kullanılmaktadır (Strobl, 2013, s. 678). Öte yandan veri madenciliğinin daha özelleşmiş bir alanı olarak görülebilecek eğitsel veri madenciliği, eğitim alanından gelen ham verilerin yine eğitim araştırmalarında kullanılabilecek faydalı bilgilere dönüştürülmesi sürecini kapsar (Romero ve Ventura, 2010). Romero ve Ventura (2010) tarafından yapılan kapsamlı derleme çalışmasındaki sınıflandırmaya göre eğitsel veri madenciliğinin çeşitli amaçlarla kullanıldığı görülmüştür. Buna göre, öğrencilerde gerçekleşen öğrenmeleri anlamlandırmayı sağlamak amacıyla verilerin analizi ve görselleştirilmesi, eğitimcilere karar alma süreçlerinde destek sağlayacak geri bildirimlerin sağlanması, öğrencilere uyarlanmış öğrenme içerikleri, bağlantılar ve aktivitelerin önerilmesi, öğrenci performansının tahmin

edilmesi, öğrenci modellemesiyle bireylere ilişkin bilişsel modellerinin geliştirilmesi, kopya, hatalı eylemler, düşük motivasyon, oyun oynama, kötüye kullanma vb. gibi istenmeyen öğrenci davranışlarının tespit edilmesi, bireyselleştirilmiş öğrenme sistemleri oluşturmak gibi amaçlarla öğrencilerin gruplandırılması, sosyal ağ analiziyle bireyler arasındaki ilişkilerin incelenmesi, kavram haritalarının oluşturulması, eğitim yazılımlarının geliştirilmesi, eğitim alanında çeşitli planlama faaliyetleriyle geleneksel eğitim sürecinin geliştirilmesi alanlarında veri madenciliğinin kullanıldığı ifade edilmiştir. Veri madenciliği teknikleri verilerin dağılımları hakkında parametrik modellere kıyasla çok daha az varsayımda bulunmaları bakımından oldukça esnek ve genellikle daha gerçekçidirler (Strobl, 2013, s. 679). Bu teknikler doğrulayıcı bir yaklaşıma sahip denetimli öğrenme teknikleri ve daha keşifsel bir yaklaşım içeren denetimsiz öğrenme teknikleri olmak üzere iki gruba ayrılırlar (Qiao ve Jiao, 2018). Buna göre eğitim setindeki verilerin, girdi vektörlerini (bağımsız değişkenler) ve bunlara karşılık gelen hedef vektörleri (bağımlı değişken) içermesi denetimli öğrenme tekniklerini kapsar. Denetimsiz öğrenme tekniklerinde ise eğitim verileri, karşılık gelen hedef değerler olmadan yalnızca girdi vektörlerini içerir. Denetimli öğrenme tekniklerinde, istenen çıktının sonlu sayıda ayrık kategoriden ya da sürekli değişkenden oluşma durumuna göre sırasıyla sınıflandırma ve regresyon problemi ortaya çıkar. Denetimsiz öğrenme tekniklerinde ise, girdi uzayındaki verilerin dağılımının belirlenmesi, görselleştirme amacıyla verilerin boyutunun indirgenmesi veya verideki benzer grupların keşfedilmesi söz konusudur (Bishop, 2006, s.3). Çeşitli avantajları bulunan bu teknikler büyük verilerle ve aynı zamanda çok sayıda yordayıcı içeren durumlarla otomatik değişken seçimi yaklaşımı sayesinde başa çıkma imkanı sunar. Öte yandan değişkenler arasındaki oldukça karmaşık ilişki kuralları da ortaya çıkarılabilir (Strobl, 2013, s. 683). Bu anlamda veri madenciliği yaklaşımlarının karmaşık gerçek dünya problemlerine cevap vermede avantaj sağladıkları söylenebilir.

Sınıflandırma ve Regresyon Ağaçları

İlk olarak Breiman ve arkadaşları tarafından 1984 yılında ortaya atılan sınıflandırma ve regresyon ağaçları kategorik ve sürekli değişkenleri içeren yordama problemlerinde kullanılmaktadır. Regresyon ve sınıflandırmayı içeren ağaç tabanlı yöntemler temelde tahmin uzayını basit bölümlere/bölgelere ayıran ve her bir bölüme basit bir model uydurarak oluşturulan tahmin modelleridir (Loh, 2010; James ve ark., 2013, s. 303). Ağaç tabanlı algoritmalar temelde bir değerin belirli bir değerden büyük veya küçük olmasıyla ilgili bilgiyi kullanır (LeDell, 2018). Bu doğrultuda bir gözlem için bağımlı değişkene ilişkin tahminde bulunulurken, bağımlı değişkenin türüne göre değişmekle birlikte, genellikle gözlemin ait olduğu bölgedeki eğitim verilerinin ortalaması veya modu kullanılır. Bu yöntemlerde işlemler gerçekleştirilirken kullanılan bölme kuralları bir ağaç şeklinde gösterilebildiği için “karar ağacı” yöntemleri olarak da bilinirler (James ve ark., 2013, s. 303). Öte yandan sınıflandırma ve regresyon ağaçları (SRA) ve ileriki bölümlerde bahsedilecek olan birkaç ağaçtan elde edilen tahminlerin çeşitli şekillerde birleştirilmesiyle kestirim yapan topluluk öğrenme yöntemleri için “özyinelemeli bölünme” terimi de kullanılmaktadır. Bu terim, örneklemin belirli kurallarla bölündükten sonra yordayıcı değişkenlere göre tekrar tekrar gruplara bölündüğünü vurgular (Strobl, 2013, s. 683). Bu sayede gruplar kendi içinde bağımlı değişken açısından daha fazla homojen ve gruplar arasında daha fazla heterojen olacak şekilde aşamalı olarak daha küçük gruplara bölünürler (Ma, 2018, s. 12). Bu yöntemlere göre bağımlı değişken sürekli ise regresyon, kategorik ise sınıflandırma ağaçları elde edilir (Hastie ve ark., 2009, s. 10). Temel olarak hem sınıflandırma hem de regresyon ağaçları ortak istatistiksel ilkeleri paylaşmakla birlikte bazı ayrıntılarda farklılık gösterirler (Ma, 2018, s. 15). İlk olarak, sınıflandırma ağaçlarında safsızlık ölçütü olarak 0-1 arasında değişen ve bir sınıfın düğümdeki gözlem sayısının toplam gözlem sayısına oranı yoluyla elde edilen Gini indeksi veya entropi değeri kullanılırken, regresyon ağaçlarında bu ölçüt üst sınırı olmayan hata kareler toplamıdır (James ve ark., 2013, s. 312; Ma, 2018, s. 48). İkinci olarak sınıflandırma ağaçlarında aynı terminal düğüme ait eğitim veri setindeki gözlemlerde en sık

görülen yanıt sınıfı tahmin edilirken, regresyon ağaçlarında bu gözlemlerin ortalaması tahmin edilen değer olarak kullanılır (James ve ark., 2013, s. 311). Son olarak, sınıflandırma ağaçlarında model değerlendirme metrikleri olarak doğruluk, karmaşıklık matrisinden hesaplanan hassasiyet, duyarlılık ve F1 puanı kullanılırken, regresyon ağaçlarında bu amaçla sürekli değişkenler için uygun olan ortalama mutlak hata, ortalama karesel hata ve/veya açıklanan varyans gibi metrikler kullanılır. Denetimli öğrenme teknikleri kapsamındaki SRA'lar bağımlı değişkenin değerini bir dizi bağımsız değişkenden yararlanarak tahmin etmeyi içeren parametrik olmayan regresyon yöntemlerindendir (Strobl, 2013, s. 683). Dolayısıyla bu algoritmaların veride normal dağılım gerekliliği ile ilgili varsayımları yoktur (LeDell, 2018). Bu teknikle bağımlı değişkeni etkileyen en önemli değişkenler ve bunlar arasındaki etkileşimler tespit edilir (Yohannes ve Webb, 1999, s. 1). Bu bağlamda SRA'larda geleneksel istatistiksel tekniklerde olduğu gibi bağımsız değişkenler arasındaki karmaşık etkileşimlerin önceden belirlenmesine gerek yoktur. Bu sayede olası yanlış belirlleme sorunlarının önüne geçilerek geleneksel tekniklerin kolaylıkla başa çıkamadığı karmaşık ve doğrusal olmayan ilişkiler tespit edilebilir (Ma, 2018, s. 12; Strobl, Malley ve ark., 2009). Öte yandan regresyon ağaçlarının performansı doğrusal olmayan ilişkiler içeren durumlarda, doğrusal regresyona göre çok daha doğru olabileceken eğer değişkenler arasında iyi derecede doğrusal ilişki varsa performansı biraz daha düşük olma eğilimindedir (Breiman ve ark., 1984, s. 264).

Karar ağaçları çok zaman alan veri ön işleme gibi adımları gerektirmeden doğrudan verilere uygulanabilmeye en yakın olan yöntemdir (Hastie ve ark., 2009, s. 352). Buna göre karar ağaçları tüm ölçek türlerindeki değişkenleri ele alabilmeleri, eksik verilerle başa çıkabilmeleri (yedek/vekil (surrogate) bölünmelerle), değişken önemini değerlendirmek için ölçütler sunabilmeleri (Breiman ve ark., 1984, s. 216, Steinberg ve Colla, 1995, s. 25), bağımsız değişkenlerdeki uç değerlere karşı sağlam olmaları (genellikle bölünmeler aykırı olmayan değerlerde gerçekleşirken, bağımlı değişkendeki aykırı değerler genellikle ağaçta sonraki bölünmeleri etkilemeyecek küçük düğümlere atanır ve bu şekilde izole edilir),

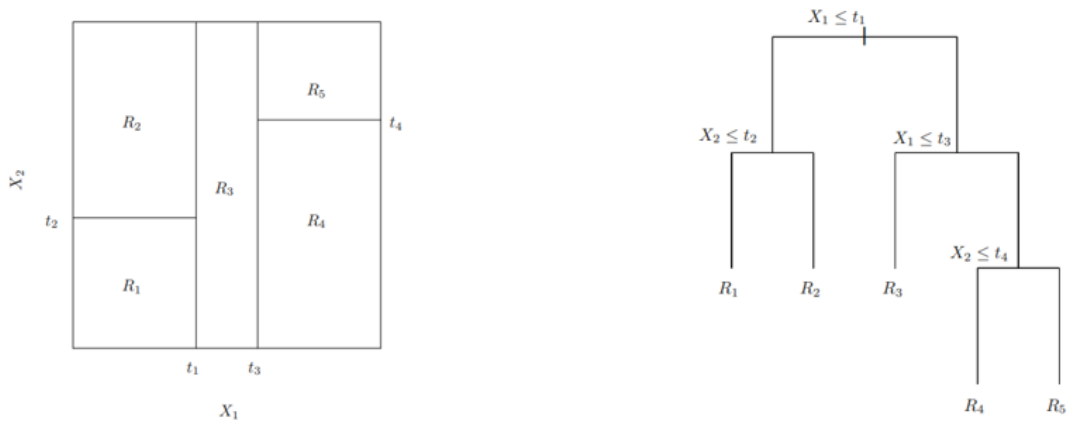
bağımsız değişkenlerin monoton dönüşümlerinden etkilenmemeleri (bağımsız değişkenler için karekök, logaritma vb. dönüşümlerden uygun olanlarını denemeye gerek yoktur), büyük örneklerde de hesaplama yapabilmeleri, bölme kurallarını belirlemek için aşamalı bir yöntem kullanmaları sebebiyle değişkenlerin önceden seçilmesine gerek olmadan ilgisiz bağımsız değişkenlerle başa çıkabilmeleri (Breiman ve ark., 1984, s. 253; Hastie ve ark., 2009, s. 351; Steinberg ve Colla, 1995, s. 23-24), yorumlanmalarının kolay olması (James ve ark., 2013, s. 303) gibi çeşitli avantajlara sahiptir. Bu özelliklerle karar ağaçları makine öğrenme alanında önemli ölçüde popüler öğrenme yöntemi olarak bilinirler (Hastie ve ark., 2009, 352, Loh, 2014a; West, 2012, s. 280).

Regresyon Ağaçlarının Temel Elemanları

Bu bölümde, çalışmada dikkate alınan, regresyon ağaçlarının yapısında bulunan temel elemanlar tanıtılmıştır. Bu amaçla terimler Şekil 4'te verilen örnek bir ağaç (James ve ark., 2013, s. 308) üzerinden açıklanmıştır.

Şekil 4

Örnek Bir Ağaç (Sol: İki boyutlu örnek üzerinde özyinelemeli ikili bölünme çıktısı. Sağ: Soldaki bölünmeye karşılık gelen ağaç)



SRA oluşturulurken yordayıcı uzayı Şekil 4'ün solunda gösterilen örnekte olduğu gibi j farklı örtüşmeyen bölgeye ayrılır. Buna göre bölme işleminin gerçekleştiği her alt grup (sağdaki şekilde $X_j < t_k$ biçiminde gösterilen) *düğüm* olarak isimlendirilir. Burada X_j bağımsız

değişkenleri, t_k ise kesme noktalarını temsil eder. SRA'larda bir bağımsız değişken ağaçta birden fazla dalda bölünmede görülebileceği gibi hiç görünmeyebilir. Çünkü her bir düğümde en iyi bölünme tüm değişkenler dikkate alınarak araştırılır. Öte yandan şekilde de görüldüğü gibi regresyon ağaçlarında yalnızca ikili bölünmeler gerçekleşir. Bu durum anlaşılması ve yorumlanması daha kolay ağaçlar elde etmeyi sağlar. Aynı zamanda bu durum, tek düğümden ikiden fazla sayıda bölünme olması durumunda verilerin çok hızlı şekilde parçalanarak bir sonraki bölünme için yetersiz veri kalması riskini de önlemiş olur. Tersten çizilen ağacın en tepesindeki düğüme *kök düğüm* denir. Buna göre $X_1 \leq t_1$ kök düğüme karşılık gelir. Bir bölünmede oluşan iki alt düğüme *çocuk düğümler*, bu bölünmenin gerçekleştiği düğüme *ebeveyn düğüm* denir. Örneğin $X_2 \leq t_2$ ve $X_1 \leq t_3$, $X_1 \leq t_1$ ebeveyn düğüme ait sırasıyla sol ve sağ çocuk düğümlerdir. Ağaçlarda bölünmeler gerçekleşirken, eğer koşul sağlanıyorsa gözlemler soldaki düğüme, sağlanmıyorsa sağdaki düğüme gider. Düğümleri birbirine bağlayan parçalar ise *dal* olarak isimlendirilir. Çocuk düğümlere daha fazla bölünmeyen düğümlere ise *terminal düğüm* ya da *yaprak* adı verilir. Şekilde verilen ağaca göre R_1 , R_2 , R_3 , R_4 , R_5 ile gösterilen kısımlar ağacın terminal düğümleri ya da yapraklarıdır. Her yapraktaki sayı regresyon ağaçlarında, buraya düşen gözlemlerin bağımlı değişkene ilişkin ortalamasıdır. Her bir terminal düğümden tahmin edilen değer sabittir (Breiman ve ark., 1984, s. 228; Genuer ve Poggi, 2020, s. 9, 12; Hastie ve ark., 2009, s. 311; Ma, 2018, s. 19, 22, 49; James ve ark., 2013, s. 304-306).

Regresyon Ağaçlarının Oluşturulmasındaki İstatistiksel Temeller

Regresyon ağaçları, regresyon problemlerine çözüm için bir alternatif sunmaktadır. Bu yöntem bazı durumlarda doğrusal regresyon analizinde ortaya çıkarılması zor veri yapıları hakkında ipuçları verebilir. Elbette ki her yöntem gibi uygun bir şekilde kullanıldığında faydalı olabilecek bir analiz çeşididir (Breiman ve ark., 1984, s. 265). Bu amaçla regresyon ağaçları oluşturulurken arka planda gerçekleşen işlemleri daha iyi anlayabilmek adına aşağıda yöntemin istatistiksel temellerine dair açıklamalara yer verilmiştir.

Bir regresyon ağacı temel olarak iki aşamada oluşturulur. İlk olarak, verileri ayıran en iyi bağımsız değişken ile en iyi kesme noktası bulunur ve veriler ikiye ayrılır. Bu işlem özyinelemeli olarak her alt gruba uygulanır. Bu aşama alt gruplar daha önceden belirlenmiş minimum gözlem sayısına sahip olana kadar veya safsızlıkta hiçbir iyileşme olmayana kadar devam eder (Therneau ve Atkinson., 2023a, s.3). Ancak bu aşama için ikinci kriter kullanıldığında gereksiz yere derin ağaçlar oluşabilir. Bu sebeple genellikle ilk kriter durdurma kuralı olarak kullanılır (Genuer ve Poggi, 2020, s. 12). İkinci aşamada ise ilk aşamada oluşan karmaşık modelin çapraz doğrulama yöntemiyle budanması söz konusudur (Therneau ve Atkinson., 2023a, s.3). Bölme işlemi gerçekleştirilirken her defasında tüm yordayıcı uzayı bölünmez. Örneğin ilk bölünmeyle ikiye ayrılan uzayda ikinci bölme işlemi, daha önceden oluşmuş iki bölgeden birinde gerçekleşir. Bu şekilde üç bölgeli bir bölünme elde edilir. Sonraki aşamada üç bölgeden biri bölünür ve bu süreç bir durdurma kriterine ulaşana kadar yinelenir (James ve ark., 2013, s. 307). SRA'larda bölünmenin gerçekleşeceği en iyi değişkenlerin ve bunlara ilişkin en iyi kesme noktalarının bulunmasında çeşitli kriterler vardır. Bu seçimlerin yapılması için safsızlık azaltma yaklaşımı kullanılır (Strobl, 2013, s. 684). Burada bahsedilen safsızlık, bir gruptaki bireylerin bağımlı değişkenin farklı değerlerine veya kategorilerine ait olma derecesidir (Ma, 2018, s. 15). Regresyon ağaçlarında kullanılan bölme kriteri, bağımlı değişkene ilişkin düğüm içi kareler toplamıdır ve bir bölmenin iyiliği, kareler toplamında meydana gelen azalma ile ölçülür (Yohannes ve Webb, 1999, s. 27). Buna göre Yohannes ve Webb (1999, s. 27) regresyon ağacının oluşturulma aşamalarını daha ayrıntılı olarak şu şekilde ifade etmiştir:

1. Kök düğümünden başlayarak tüm bağımsız değişkenler üzerinden tüm olası bölünmelerin gerçekleştirilmesi ve her bölünmeye düğüm safsızlık ölçütünün uygulanarak safsızlıktaki azalmanın belirlenmesi.

2. Bölünme iyiliği kriterinin uygulanarak en iyi bölünmenin seçilmesi ve sağ/sol alt düğümlerin oluşturulması.

3. Terminal olmayan her düğüm için 1. ve 2. adımların tekrar uygulanmasıyla mümkün olan en büyük ağacın üretilmesi.

4. En büyük ağaca budama yapılması ve en uygun ağacın elde edilmesi.

Regresyon ağaçlarında hangi değişkenin en iyi bölünmeyi verdiği karar vermeyi sağlayan bölme kriteri düğüm içi varyanstır. Bu bağlamda safsızlık ölçütü olan düğüm içi varyans, hata kareler toplamı alınarak hesaplanır (Ma, 2018, s. 47). Hata kareler toplamı (Sum of Squares) şu şekilde formülize edilir (Therneau ve Atkinson, 2023a, s. 32):

$$SS_T = \sum (y_i - \bar{y})^2 \quad (1)$$

Eşitlik 1'de verilen y_i i bireyinin bağımlı değişkene ilişkin gözlenen değeri iken, $\bar{y}$ bağımlı değişkene ilişkin düğüm ortalamasıdır. Regresyon ağaçlarında bir üst düğüm bölündüğünde, üst düğümün safsızlık ölçüsü alt düğümlerinkine karşılaştırılır. Ana düğüm ve çocuk düğümler arasında safsızlıktaki en büyük azalmayı veren bağımsız değişken bölme değişkeni olarak seçilir (Ma, 2018, s. 47). Bu kriter ise şu şekildedir (Therneau ve Atkinson, 2023a, s. 32):

$$SS_T - (SS_L + SS_R) \quad (2)$$

Burada SS_L ve SS_R sırasıyla sol ve sağ çocuk düğümleri için kareler toplamıdır (Therneau ve Atkinson, 2023a, s. 32). Regresyon ağaçlarında safsızlığın üst sınırı yoktur. Bu sebeple değerlendirme yapılırken bağımlı değişkenin orijinal varyansının referans olarak kullanılması gerekir (Ma, 2018, s. 48).

Yukarıda açıklanan süreçle muhtemelen büyük ve/veya karmaşık bir ağaç oluşturulmuş olur (Therneau ve Atkinson, 2023a, s. 12). Esasında büyük bir ağaçta safsızlıkta önemli bir azalma elde edilebilir (Ma, 2018, s. 48). Öte yandan regresyon ağaçlarında daha az terminal düğüme sahip (yani, daha küçük) ağaçlar basitliği sağlarken, doğruluk daha fazla terminal düğüme sahip (yani, daha büyük) ağaçlarda daha fazladır. Teorik olarak yüzlerce terminal düğüme sahip bir ağaç çok karmaşıkken, tek bir düğüme sahip ağaç basit olmasına karşın oldukça hatalı kabul edilir (Steinberg ve Colla, 1995, s.

175). Bu duruma bağılı olarak oluşturulan büyük ve/veya karmaşık ağaç eğitim setinde iyi tahminler üretirken, test setinde iyi genelleme yapamadığı için performansı düşük olabilir. Bu durum aşırı uyum olarak bilinir. Böyle bir sorunla karşılaşıldığında daha az terminal düğüme sahip bir ağaç, bir miktar sapma gösterse de daha kolay yorumlama ve daha düşük varyans verme eğiliminde olacaktır (James ve ark., 2013, s. 307; Strobl, Malley ve ark., 2009). Esasında ağaçların doğruluk ve varyansları arasında bir ödünleşim (trade-off) ilişkisi vardır. Buna göre daha düşük yanlılığa sahip (daha büyük) ağaçlar yüksek varyans gösterme eğiliminde olurlar (bunun tersi de geçerlidir) (Polikar, 2012, s. 2). Temelde hedeflenen, yorumlanabilirlik ve tahmin doğruluğu arasında denge kurabilen bir ağaç elde etmektir (Steinberg ve Colla, 1995, s. 175). Bunu sağlamak için mümkün olan en büyük ağaç oluşturulduktan sonra bağımsız bir test örneği veya çapraz doğrulama yaklaşımıyla budama yapılır (Yohannes ve Webb, 1999, s. 28). Başka bir deyişle, bu yaklaşımla tahmin doğruluğuna katkıda bulunmayan dallar elenir (Strobl, 2013, s. 687). Bu işlem maliyet-karmaşıklık (cost complexity) ölçütüyle büyük ağaca bir ceza terimi ekleme yoluyla yapılır (Ma, 2018, s. 48). Burada olası tüm ağaçları dikkate almak yerine negatif olmayan bir ayar parametresi α ile indekslenen bir ağaç dizisi ele alınır. α 'nın her değeri için $T \subset T_0$ alt ağacı vardır (James ve ark., s. 308-309). Bu durumda, α (karmaşıklık parametresi) modele başka bir değişken eklemenin maliyetini ölçen 0 ile ∞ arasında bir sayıdır. Buna göre maliyet-karmaşıklık değeri şu şekildedir (Therneau ve Atkinson, 2023a, s. 12):

$$R_\alpha(T) = R(T) + \alpha|T| \quad (3)$$

Eşitlik 3'teki $|T|$ terminal düğümlerin sayısıdır. $R(T)$ ise hata kareler toplamıdır (Therneau ve Atkinson, 2023a, s. 12). α parametresi alt ağacın eğitim verilerine uyumu ile ağacın boyutu arasındaki dengeyi kontrol eder. $\alpha = 0$ ise T alt ağacı T_0 'a yani hiç budanmamış ağaca karşılık gelir. α değeri arttıkça daha küçük bir ağaç elde edilir, küçük α değerleri için bunun tersi söz konusudur (Hastie ve ark., 2009, s. 308; James ve ark., 2013, s. 309). α 'nın tahmini genellikle 5 veya 10 kat çapraz doğrulama ile hesaplanır (Hastie ve

ark., 2009, s. 308). Özetle yukarıda açıklanan süreçlere dayalı olarak bir regresyon ağacı oluşturma süreci James ve arkadaşları (2013, s. 309) tarafından şu şekilde ifade edilmiştir:

1. Eğitim verilerinde özyinelemeli ikili bölme ile her bir terminal düğümde daha önceden belirlenmiş minimum gözlem sayısından (genellikle 5 olarak alınır (Breiman ve ark., 1984, s.233)) daha az gözlem olana kadar devam edilerek büyük bir ağaç elde edilir. Minimum gözlem sayısından daha az gözlem olduğunda süreç durur.

2. En iyi alt ağaçları elde etmek için büyük ağaca maliyet karmaşıklığı hesaplamasıyla budama yapılır.

K-katlı çapraz doğrulama ile α seçilir. Bu amaçla her $k=1, \dots, K$ için:

(2a) k. kattaki eğitim verileri hariç olmak üzere, 1. ve 2. adımlar geri kalan eğitim verilerine uygulanır.

(2b) α 'nın bir fonksiyonu olarak dışarıda bırakılan k. kattaki veriler üzerinden ortalama hata kareler değeri hesaplanır.

Sonuçların ortalaması alınır ve ortalama hatayı en aza indiren α seçilir.

3. Seçilen α değeriyle budanmış ağaç elde edilir.

Ağaç oluşturma işlemi tamamlandıktan sonra her bir terminal düğümde, bu düğümdeki bağımlı değişkene ilişkin sabit bir ortalama değer atanarak tahmin yapılır (Strobl, 2013, s. 688). Başka bir deyişle, test veri setindeki gözlemin ait olduğu terminal düğümde bulunan eğitim veri seti gözlemlerinin ortalaması alınır ve bu değer test veri setindeki gözleme dair bağımlı değişkene ilişkin tahmin edilen değer olur (James ve ark., 2013, s. 307). Bu işlemle bir regresyon ağacı parçalı ve sabit bir tahmin fonksiyonu oluşturmuş olur (Loh, 2014a, Strobl, 2013, s. 688).

Regresyon Ağaçlarında Eksik Değer Mekanizması

Çalışmada kullanılan rpart (Therneau ve Atkinson, 2023b) paketinde bağımlı değişkene ilişkin bir değere sahip olmayan ya da hiçbir yordayıcı değişken için değere sahip

olmayan gözlemler silinir. Bunun haricinde bağımlı değişken ve en az bir bağımsız değişken için değere sahip olan tüm bireyler modellemeye dahil edilir (Therneau ve Atkinson, 2023a, s. 18, 23). Popüler bir yöntem olarak regresyon ağaçlarında eksik verilerin ele alınmasında bir ana düğümü bölmek için seçilen bağımsız değişkenlerin alternatiflerine karşılık gelen vekil (surrogate) bölünmeler kullanılır (Ma, 2018, s. 69). Kullanılan pakette de bu yönteme göre hesaplamalar yapılır. Bu yaklaşımla eksik değere sahip bir gözlemin değeri diğer bağımsız değişkenler kullanılarak tahmin edilir. Örneğin, $yaş < 40$, $yaş \geq 40$ olacak şekilde bir bölünme seçildiğini varsayalım. Yaş değişkeninde eksik değere sahip gözlemlerin neye göre bölüneceğine karar vermek için diğer bağımsız değişkenler kullanılarak $yaş < 40$, $yaş \geq 40$ bölünmesini en iyi şekilde taklit edebilecek değişkeni seçmek için özyineleme olmadan bölme algoritması yeniden uygulanır. Bölme değişkeni eksik olan gözlem daha sonra ilk vekil değişken kullanılarak sınıflandırılır. İlk vekil değişken eksikse ikinci vekilin, bu da eksikse üçüncü vekilin kullanılması şeklinde bir süreç devam eder (Therneau ve Atkinson, 2023a, s. 18-19).

Değişken Önemi

Kullanılan rpart (Therneau ve Atkinson, 2023b) paketinde değişken önemi, değişkenlerin birincil ve vekil olarak görüldüğü bölünmelere göre belirlenen iki değer toplamasıyla hesaplanır. Buna göre ilk değer, X_j bağımsız değişkeninin birincil değişken olduğu her bölünme için heterojenlikteki azalmanın (bölünme iyiliği) toplamıdır. İkinci değer, X_j 'nin vekil olduğu bölünmelerde ortaya çıkan heterojenlikteki azalmanın toplamının düzeltilmiş uyuma göre ağırlıklandırılmasıyla elde edilir. Bu değer bölünme iyiliği ile düzeltilmiş uyumun çarpılmasıyla bulunur. Düzeltilmiş uyum şu şekilde hesaplanır:

$$\begin{cases} (n_{X_j - n_{maj}}) / (n_t - n_{maj}) & n_{X_j} > n_{maj} \text{ ise} \\ 0 & n_{X_j} \leq n_{maj} \text{ ise} \end{cases} \quad (4)$$

4 numaralı ifadede n_{X_j} ve n_{maj} sırasıyla X_j 'yi içeren vekil bölünme ve tüm gözlemleri en büyük boyuttaki alt düğüme gönderen çoğunluk kuralına göre bölünme sonucu yönlendirilen gözlemlerin sayısıdır. Paket çıktısında hesaplanan değerler, toplamaları 100 olacak şekilde

ölçeklendirilir ve yuvarlanmış değerleri gösterilir. Önem değeri %1'den az olan değişkenler gösterilmez (Genuer ve Poggi, 2020, s.25; Therneau ve Atkinson, 2023a, s.12).

Ağaçların Kararsızlığı

Regresyon ağaçlarının pek çok avantajı olmakla birlikte bazı sınırlılıkları da söz konusudur. Bunlardan en önemlisi yüksek değişkenlik/kararsızlık göstermeleridir (Genuer ve Poggi, 2020, s. 1). Başka bir deyişle öğrenme verisindeki çok küçük bir değişiklik bile oldukça farklı bölünmelere yol açarak yorumlamada bazı karışıklıklara sebep olabilir (Hastie ve ark., 2009, s. 312). Özyinelemeli bölünmede, bölünmedeki her bir değişken ve seçilen her kesme noktası, sonrasında yapılan bölünmeleri etkiler. Ayrıca, bölünme değişkeninin ve kesme noktasının seçimi eğitim veri setindeki bireylerin belirli dağılımına büyük ölçüde bağlıdır. Buna göre eğitim veri setindeki küçük bir değişikliğe bağlı olarak ilk kesme noktası ve ilk bölünme değişkeni değişirse tamamen farklı bir ağaç elde edilebilir. Ya da ilk bölünmedeki bir hatanın etkisi geri kalan tüm bölünmelere yansır (Hastie ve ark., 2009, s. 312; Strobl, Malley ve ark., 2009). Bu kararsızlık nedeniyle tekli ağaçlardan elde edilen tahminler yüksek değişkenlik gösterirler (Strobl, Malley ve ark., 2009). Sürecin hiyerarşik doğasından kaynaklanan bu soruna (Hastie ve ark., 2009, s. 312) çözüm olarak tahmin için bir yerine birden fazla ağacın kullanılmasıyla elde edilen sonuçların ortalamasının alınmasına dayalı topluluk öğrenme yöntemleri önerilmektedir (Strobl, 2013, s. 690). Ağaçları yapı taşı olarak kullanan torbalama, rastgele orman ve artırma gibi topluluk öğrenme yöntemleri tekli ağaçların tahmin performansını önemli ölçüde artırabilir (James ve ark., 2013, s. 316). Buna ek olarak topluluk öğrenme yöntemleriyle tekli ağaçların parçalı sabit formundan kaynaklanan, yordayıcı değişkenlerdeki küçük değişiklikler için bile yapılan tahminin oldukça farklı olabilme riski de azaltılmış olur. Başka bir deyişle topluluk öğrenme yöntemleri karar sınırlarının daha düzgün (smooth) olmasını sağlar (Strobl, Malley ve ark., 2009).

Ağaçlar parametrik karşılıklarının ele alabileceğinden çok daha fazla bağımsız değişkenle başa çıkabilir. Parametrik tekniklerde bağımsız değişken sayısının artmasıyla

birlikte gerekli olan örneklem büyüklüğü vb. gibi karşılanması gereken bazı durumlar söz konusudur. Karar ağaçları ise her bir düğümde aynı anda tek bir değişkenle çalıştığı için fazla sayıda bağımsız değişkeni ele alabilir. Bununla birlikte ağaçların bazı ek sınırlılıkları da bulunmaktadır. Örneğin karar ağaçlarında standart hata, regresyon katsayısı, gerekli örneklem büyüklüğü gibi bazı istatistiksel kavramlar söz konusu değildir. Bu anlamda olasılığa dayalı modeller olmayıp, fonksiyonel formun belirlenmesinde bilgisayar programına bağlı olmaları açısından algoritmiktirler (Ma, 2018, s. 61, Sinharay, 2016). Bu bağlamda yüksek tahmin doğruluğuna sahip olmakla birlikte tek tek yordayıcıların etkilerinin doğrudan yorumlanamayışı açısından çoğu makine öğrenme algoritması “kara kutu” tahmin yöntemleri olarak bilinir (Mayr ve ark., 2014). Buna karşılık değişkenler arasındaki ilişkilerin matematiksel denklemler aracılığıyla gösterilememesi, karar ağaçlarının sağladığı doğrudan yorumlama ve sunduğu anlaşılabilirlik sayesinde her zaman olumsuz bir özellik olarak görülmeyebilir (Ma, 2018, s. 61). Öte yandan bu duruma belli bir derecede çözüm sunması için çeşitli algoritmalar değişkenlerin önem düzeylerini belirten ölçütler sunmaktadır (Hofner ve ark., 2011). Bazı algoritmalar için arka planda işleyen süreç hakkında daha açıklayıcı bilgiler edinmeyi sağlayan araştırmalar da bulunmaktadır (bkz. Philipp ve ark., 2016; Philipp ve ark., 2018; Schneider ve ark., 2023). Özel olarak gradyan artırma algoritması için bilindik istatistiksel sonuçlar sağlayan yaklaşımlar da geliştirilmiştir (bkz. Bühlmann ve Hothorn, 2007). Bahsedilen sınırlılıklara ek olarak makine öğrenme algoritmalarının hesaplama sürelerinin geleneksel tekniklere göre (özellikle büyük veri setlerinde) daha uzun olabileceği de söylenmelidir (Sinharay, 2016).

Regresyon Ağaçlarının Kullanıldığı Araştırmalar

Blanch ve Aluja (2013) tarafından yapılan araştırmada kişilik, zekâ ve akademik başarı arasındaki ilişkilerin regresyon ağacıyla ortaya koyulması hedeflenmiştir. Buna göre 818 ortaöğretim öğrencisiyle yapılan çalışmada yetenek, kişilik ve dört öğrenme alanındaki akademik performans (dil (ispanyolca, katalanca), fen (matematik ve doğa)) regresyon ağaçlarıyla modellenmiştir. Sonuçta, yeteneğin (kızlar ve erkeklerde) akademik

performansı açıklamada kişiliğe göre daha önemli olduğu sonucuna ulaşılmıştır. Özellikle kızlar için sözel ve hesaplama yetenekler, erkekler için soyut akıl yürütme ve hesaplama yetenekler akademik başarı için kişiliğe göre daha belirleyici yordayıcılar olarak bulunmuştur.

Gomes ve Jelihovschi (2020) tarafından yapılan araştırmada büyük ve karmaşık bir veri setinde regresyon ağacı yönteminin kullanılabilirliğinin gösterilmesi amaçlanmıştır. Bu amaçla Brezilya 2011 ulusal ortaöğretim sınavına katılan 3.670.089 kişinin verileriyle sınavdaki genel başarıya ilişkin tahmin yapılmıştır. Farklı ölçek türlerinde 53 bağımsız değişken kullanılmıştır. Sonuçta, regresyon ağacı yönteminin bağımlı değişkendeki varyansı tahmin etmenin ötesinde, değişkenler arasında ortaya koyduğu ilişkiler sayesinde teorik bilgiyi de genişletmeye imkân sağlayabileceği vurgulanmıştır.

Gomes ve arkadaşları (2021) tarafından yapılan araştırmada karmaşık özellikte eğitime ilişkin bir veri setinde çoklu doğrusal regresyon ve regresyon ağacı yöntemlerinin performanslarının karşılaştırılması amaçlanmıştır. Bu amaçla Brezilya ulusal ortaöğretim sınavına katılan 3.670.089 kişinin verileri 53 yordayıcıyla beraber analiz edilmiştir. Bağımlı değişken olarak okuduğunu anlama puanı kullanılmıştır. Çalışmanın sonucunda regresyon ağacı analizinin farklı ölçek türlerine sahip karmaşık veri kümelerinde kullanılabilmesi, doğrusal olmayan ilişkileri ortaya çıkarabilmesi gibi avantajlarına vurgu yapılmıştır.

Aydoğan ve Gelbal (2022) tarafından yapılan araştırmada fen başarısını yordayan değişkenlerin regresyon ağacıyla ortaya koyulması amaçlanmıştır. Buna göre TIMSS 2015 8. Sınıf fen bilimleri testinden elde edilen makul değerlerin ortalaması, öğrenci, öğretmen ve okul anketlerinden elde edilen veriler kullanılmıştır. 218 okuldan 6079 öğrenci ve 220 öğretmenin verisinin kullanıldığı araştırmada, verideki çok düzeyli yapıyı daha iyi ortaya koyabilmek adına sınıflandırma ve regresyon ağacı temelli RE-EM (Random Effect-Expectation Maximization) algoritmasıyla analizler yapılmıştır. Sonuç olarak on yedi yordayıcıdan on ikisinin fen başarısını anlamlı düzeyde yordadığı tespit edilmiş ve anlamlı çıkan değişkenlerle fen başarısı arasındaki ilişkiler yorumlanmıştır.

Yukarıda verilen araştırmalar incelendiğinde regresyon ağaçlarının farklı alanlardaki akademik başarı, kişilik, zekâ tahmininde kullanıldığına ve büyük veri setlerinde de uygulanabilir olduğuna dair çalışmalar olduğu söylenebilir. Öte yandan bu çalışmada olduğu gibi regresyon ağaçlarının KPÇ bağlamında ele alındığı bir çalışmaya bilindiği kadarıyla rastlanmamıştır.

Topluluk Öğrenme Yöntemleri

Ağaçların kararsızlık problemine çözüm olarak sunulan topluluk öğrenmede temel olarak daha basit modellerden oluşan ağaç topluluğunun güçlü yönlerinin çeşitli yöntemlerle birleştirilmesiyle tahmin modeli oluşturulması söz konusudur (Hastie ve ark., 2009, s. 605; Strobl, 2013, s. 690). Bu yöntemler pek çok makine öğrenme problemini ele almakla birlikte başlangıçta tahmin modelinin varyansını azaltarak doğruluğunu artırmak amaçlı geliştirilmişlerdir. Ortalama almanın varyans azaltıcı bir etkisi olduğu bilinmektedir. Dolayısıyla topluluk öğrenme yöntemlerinde nispeten benzer yanlılığa sahip bir dizi ağaç oluşturulduktan sonra varyansı azaltmak için her bir ağaçtan yapılan tahminlerin, örneğin regresyon ağaçlarında ortalama alınmasıyla, birleştirilerek yeni bir tahmin elde edilmesi amaçlanır. Tahminlerin ortalamasının alınması aynı zamanda hata bileşenlerinin de ortalamasının alınmasına yol açtığı için hatayı da azaltır (Polikar, 2012, s. 1-3). Topluluk öğrenme yöntemlerinin uygulaması bir anlamda tahmin için uzman grubunun kullanılmasına benzetilmiştir. Bu benzetmeye göre farklı bilgi/deneyim geçmişine sahip uzmanların tek bir uzmana göre daha doğru tahmin yapma olasılığının daha yüksek olması beklenir (Sinharay, 2016). Öte yandan genel olarak sağladığı tahmin doğruluğu avantajıyla birlikte bu yöntemlerin doğrudan ağaç şeması elde edilmesine imkan tanımaması açısından tekli ağaçların sunduğu yorumlanabilirlikle rekabet edemeyeceği de bilinmelidir (Ahn, 2014). Ancak bu sınırlılığı aşmak için topluluk öğrenme yöntemlerinde değişken öneminin yorumlanmasına ilişkin çeşitli ölçütler geliştirilmiştir. Verilen açıklamalara dayalı olarak metnin devamında, çalışmada kullanılan topluluk öğrenme yöntemlerinden torbalama, rastgele orman ve gradyan artırma algoritmalarına yönelik açıklamalara yer verilmiştir.

Torbalama (Bagging)

Bir dizi bağımsız ağacın büyütülmesine dayanan torbalama (Liaw ve Wiener, 2002) Breiman (1996) tarafından önerilmiştir. *Bootstrap aggregating* ifadesinin kısaltması olarak bilinen torbalama, eğitim veri setinden önyükleme (bootstrap) ile elde edilen alt kümelerden yapılan tahminlerin birleştirilmesiyle gerçekleştirilir. Birleştirme işlemi eğer bağımlı değişken kategorik ise çoğunluk oylamasıyla yapılırken, sürekli ise ortalama alınır (Breiman, 1996). Bootstrap örnekleme; n boyutlu öğrenme verisi $\mathcal{L}_n$ için bu verideki her bir gözlemin (X_i, Y_i) her çekilişte seçilme olasılığı $1/n$ olacak şekilde rastgele çekilmesidir (Genuer ve Poggi, 2020, s. 35). Topluluk öğrenme yöntemlerinde karar sınırlarının düzgünleştirilmesiyle parçalı sabit formda olan tekli ağaçlara göre daha esnek bir fonksiyonel form elde edilir (Strobl, Malley ve ark., 2009). Algoritma adımları şu şekilde sıralanabilir (James ve ark., 2013, s. 317):

1. Tekrarlı olarak (B defa-ağaç sayısı) eğitim kümesinden önyükleme ile rastgele örneklem çekilir. Bu örneklemelerden yararlanarak ağaçlar oluşturulur.
2. Oluşturulan ağaçlarla kestirimler yapılır ve her bir tekli regresyon ağacından elde edilen tahminlerin ortalaması alınır. Bu işleme ait formül şu şekildedir:

$$\widehat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \widehat{f}^{*b}(x) \quad (5)$$

Torbalama işleminde budama yapılmaz.

Torba Dışı (Out-of-Bag) Hata Kestirimi

Torbalamada, doğrulama kümesi ya da çapraz doğrulamaya gerek kalmadan hata kestirimi yapılabilir. Ayrı bir test kümesi ihtiyacını ortadan kaldıran torba dışı hata kestirimi yönteminde her bir önyüklemeli eğitim veri setine göre oluşturulmuş ağaçta gözlemlerin yaklaşık üçte ikisi kullanılır. Dışarıda kalan gözlemlerin üçte biri ise torba dışı gözlem olarak isimlendirilir. Örneğin, birinci gözleme ilişkin tahminde bulunurken bu gözlemin torba dışı gözlem olduğu ağaçlar kullanılır. Daha sonra birinci gözlem için elde edilen yaklaşık $B/3$

sayıda tahminin ortalaması alınır. Benzer şekilde n gözlemin tamamı için torba dışı tahmin yapılır ve buradan genel torba dışı hata kestirimi (hata kareler ortalaması) hesaplanır (Breiman, 2001; James ve ark., 2013, s. 317-318).

Değişken Önemi

Torbalamada tekli ağaçlardaki gibi basit ve yorumlanabilir tek bir ağaç modeli elde edilemez. Bu sebeple değişken önem değerleriyle tahminin genel bir özeti elde edilebilir. Buna göre yordayıcılarda meydana gelen bölünmeler sonucu elde edilen hata kareler ortalamasındaki (ağaç sayısına göre ortalama alınır) azalma değişken öneminin göstergesidir. Dolayısıyla hata kareler ortalamasında en büyük azalmayı sağlayan yordayıcılar daha önemli olarak kabul edilir (James ve ark., 2013, s. 319).

Rastgele Orman

Torbalama yöntemine ek bir rastgelelik eklenmesini içeren rastgele orman algoritması (Liaw ve Wiener, 2002) Breiman (2001) tarafından önerilmiştir. Buna göre torbalamadaki gibi önyüklemeli örneklem kullanmaya ek olarak her bir düğüm, o düğümde rastgele seçilen yordayıcılardan oluşan alt küme içinden seçilen en iyi yordayıcı kullanılarak bölünür. Ormandaki tüm ağaçlar bağımsız ve aynı dağılıma göre çekilen örneklemelerden elde edilir (Breiman, 2001). Bu algoritmada temel prensip, değişkenlerin rastgele seçimiyle varyans çok artırılmadan ağaçlar arasındaki korelasyon azaltılarak torbalama algoritmasıyla elde edilen azalan varyansın iyileştirilmesidir (Hastie ve ark., 2009, s. 588). Rastgele değişken seçimi, tüm değişkenlerin birlikte ele alındığı durumda göz ardı edilebilecek değişkenlerin topluluğa girmesine imkân tanıyarak gözden kaçabilecek değişken etkilerinin ortaya çıkarılmasını sağlayabilir (Strobl, Malley ve ark., 2009). Bu algoritmada budama işlemini yapılmaz (Breiman, 2001). Öte yandan genetik ve biyoloji gibi alanlarda sıklıkla karşılaşılan yordayıcı sayısının gözlem sayısından çok daha büyük olduğu yüksek boyutlu problemlerde de (örneğin, yordayıcı olarak pek çok (yüzlerce) genetik ve çevresel değişkenin olduğu bir durumda hastalık tahmini yapmak için nispeten daha az

sayıda hasta olması) rastgele orman algoritması her bölünmede değişkenlerin rastgele alt kümelerinin dikkate alınması sayesinde kullanılabilir (Genuer ve Poggi, 2020, s.2; Hastie ve ark., 2009, s. 649; Strobl ve ark., 2007; Strobl, Malley ve ark., 2009). Bu gibi yüksek boyutlu problemlerde elbette ki öncelikle ön değişken seçimi yapıp ardından geleneksel sınıflandırma ve regresyona dayalı modeller kullanılabilir. Ancak bunun yerine rastgele orman, artırma, destek vektör makineleri vb. gibi makine öğrenme yöntemlerinin kullanılması daha tutarlı sonuçlar vereceği gibi aynı zamanda tahmin ve değişken seçimi işlemlerinin eşzamanlı yapılabilmesi anlamında avantajlı görünmektedir (Hothorn ve Bühlmann, 2006). Bu çerçeveye göre rastgele orman algoritmasında çok sayıda bağımsız değişken, temel bileşen analizi veya faktör analizi gibi boyut azaltma tekniklerine gerek kalmadan aynı anda ele alınabilir (Strobl, Malley ve ark., 2009). Bunlara ek olarak rastgele orman algoritması hem sınıflandırma hem de regresyon problemlerinde kullanılabilmesi, bir ya da iki tane ayarlama parametresi içermesiyle kolay uygulanabilmesi, eğitim ve tahmin süreçlerinin nispeten hızlı olması, hata tahmini için torba dışı örneklem kullanması gibi avantajlara sahiptir (Cutler ve ark., 2012, s. 157). Bu bağlamda algoritmanın adımları şu şekilde ifade edilebilir (Liaw ve Wiener, 2002):

1. Orijinal verilerden toplulukta büyütülecek ağaç sayısı (ntree parametresi) kadar önyüklemeye örneklem çekilir.

2. Her bir örnekte budanmamış ağaçlar oluşturulur. Bu aşamada her bir düğümde kullanılacak rastgele değişken sayısı (mtry parametresi) belirlenir ve en iyi bölünme seçilir. Bu parametrenin varsayılan değeri regresyon ağaçlarında toplam değişken sayısının üçte biridir⁴ (Hastie ve ark., 2009, s. 592). Ayrıca bu parametrenin değeri ağaçtaki tüm düğümler için sabittir (Genuer ve Poggi, 2020, s. 40).

3. Büyütülen tüm ağaçlardan elde edilen tahminlerin ortalaması alınarak kestirim yapılır.

⁴ Sınıflandırma ağaçlarında bu değer, değişken sayısının kareköküdür (Hastie ve ark., 2009, s. 592).

Rastgele orman algoritmasında hem tahmin hatasını hem de değişken önemini hesaplamak için, yukarıda açıklanan, torba dışı hata kestirimi yaklaşımı kullanılır (Cutler ve ark., 2012, s. 164).

Değişken Önemi

Topluluk öğrenme yöntemlerinde her değişken, oluşan farklı ağaçlar sayesinde diğer değişkenlerle farklı biçimlerde ele alınır. Bu sayede yordayıcıların bağımlı değişken üzerindeki etkisi daha iyi görülür (Strobl, Malley ve ark., 2009). Dolayısıyla rastgele orman algoritması esasında kestirim yapılması amacıyla geliştirilmiş olsa da zaman içinde yordayıcıların kestirim için uygunluğunu değerlendirmek adına popüler bir yöntem haline gelmiştir (Debeer ve Strobl, 2020). Bununla birlikte tekli sınıflandırma ve regresyon ağaçlarında, özyinelemeli değişken seçiminin neden olduğu aşamalı regresyon yöntemindekine benzer sıra etkisi görüldüğü bilinmektedir. Çünkü her bölünmede seçilen değişken kendinden önce seçilen değişkenlere göre ancak sonraki değişkenlerden bağımsız olarak seçilir. Bu etki değişken öneminin hesaplanmasında topluluktaki tüm ağaçlar üzerinden ortalama alınmasıyla yoluyla ortadan kaldırılır (Strobl, Malley ve ark., 2009).

Yordayıcıların, bağımlı değişken üzerindeki etkilerinin sıralanmasını içeren değişken önemi (Genuer ve Poggi, 2020, s. 57) düğüm safsızlığındaki ortalama azalmaya (Gini önemi) veya doğruluktaki ortalama azalmaya (permütasyon önemi) göre hesaplanır (Breiman, 2001; Breiman ve Cutler, 2003). Gini önemi, değişkendeki bölünmeden kaynaklanan düğüm safsızlığındaki toplam azalmadır. Bu işlemde tüm ağaçların ortalaması alınır. Düğüm safsızlığı regresyon ağaçlarında hata kareler toplamıyla ölçülür (Liaw ve Wiener, 2022, s. 6). Ancak Strobl ve arkadaşları (2007) bağımsız değişkenlerin kategori sayıları veya ölçek türleri değiştiğinde Gini öneminin yanlış olduğunu göstermişlerdir. Buna ek olarak permütasyon öneminin değişken öneminde kullanılan en yaygın ölçüt olduğu bilinmektedir (Genuer ve Poggi, 2020, s. 57). Bu sebeple metnin devamında, çalışmada kullanılan permütasyon önemiyle ilgili açıklamalara yer verilmiştir.

Permütasyon önemi, etkisi araştırılan değişkenin değerlerinin torba dışı örnekleme rastgele permütasyona tabi tutulmasıyla meydana gelen hatadaki ortalama artıştır (Genuer ve Poggi, 2020, s. 57-58). Öncelikle orijinal veriden torba dışı gözlemlere ilişkin hata kareler ortalaması hesaplanır. Sonra etkisi araştırılan bağımsız değişken X_j 'nin Y bağımlı değişkeni ile olan orijinal ilişkisi permütasyona tabi tutularak bozulur. Permütasyona uğramış X_j değişkeni, permütasyona uğratılmamış diğer bağımsız değişkenlerle birlikte tahmin için kullanılır. Ardından hata kareler ortalaması kaydedilir. Bu işlem tüm bağımsız değişkenler için tekrarlanır. Eğer bağımsız değişken önemli ise permütasyona uğratıldıktan sonra tahmin doğruluğunun önemli ölçüde azalması beklenir. Bu yaklaşıma bağlı olarak değişken önemi tüm ağaçlar üzerinden ortalaması alınan, X_j değişkeninin permütasyonundan sonra ve önce hesaplanan tahmin doğruluğu arasındaki farktır (Liaw ve Wiener, 2022, s. 6; Strobl ve Zeileis, 2008). Permütasyona dayalı değişken öneminin (PI-Permutation Importance) hesaplanması Debeer ve Strobl (2020) tarafından şu şekilde formülize edilmiştir:

$R^{(t)}$ ve $R_{(k)}^{(t)}$ sırasıyla X_k bağımsız değişkeninin torba dışı değerlerinin permütasyonundan önce ve sonra torba dışı örneklem $\beta^{(t)}$ 'ya dayalı t ağacının tahmin hatası olsun. Bu durumda:

$$R^{(t)} = \sum_{i \in \beta^{(t)}} \frac{(\hat{y}_i^{(t)} - y_i)^2}{|\beta^{(t)}|} \quad (6)$$

$$R_{(k)}^{(t)} = \sum_{i \in \beta^{(t)}} \frac{(\hat{y}_{i(k)}^{(t)} - y_i)^2}{|\beta^{(t)}|} \quad (7)$$

Eşitlik 6 permütasyondan önceki tahmin, eşitlik 7 permütasyondan sonraki tahmindir. Paydadaki değer ise t ağacı için torba dışı örneklem sayısıdır. Bu durumda bir ağaç için permütasyon önemi şu şekildedir:

$$PI_{(k)}^{(t)} = R_{(k)}^{(t)} - R^{(t)} \quad (8)$$

Her değişken için ham değişken önem puanı ise tüm ağaçlar üzerinden ortalama alınarak hesaplanır. Burada n_{tree} toplam ağaç sayısı olmak üzere:

$$PI_{(k)} = \frac{\sum_{t=1}^{ntree} PI_{(k)}^{(t)}}{ntree} \quad (9)$$

Buna ek olarak Breiman ve Cutler (2003) tarafından değişken öneminin bir standart hataya göre normalleştirilmesine dayanan z puanı önerilmiştir. Buna göre her bir değişkenin önemi $PI_{(k)}^{(t)}$, σ standart sapmaya sahipse, ortalama önem $PI_{(k)}$, $\frac{\sigma}{\sqrt{ntree}}$ olarak hesaplanan standart hataya sahiptir. Ortalama önemin, belirtilen standart hataya bölünmesiyle ölçeklenmiş değişken önemi puanı elde edilir (Strobl ve Zeileis, 2008). Burada değişken önem ölçümlerinden elde edilen sayısal değerlerin mutlak değerlerine göre herhangi bir kıyaslama yapılmaması gerektiği de vurgulanmalıdır. Büyük ölçüde ayarlama parametrelerinin seçimine bağlı olan bu sayısal değerler için yalnızca değişkenlerin önem sıralamalarının yorumlanması tavsiye edilmektedir (Strobl, Malley ve ark., 2009).

Gradyan Artırma (Gradient Boosting)

Güçlü bir teknik olan gradyan artırmanın temel prensibi birçok zayıf öğrenicinin (learner)⁵ sonuçlarının birleştirilmesiyle güçlü bir öğrenici oluşturmaktır (Hastie ve ark., 2009, s. 337). Bu algoritma Friedman (2001, 2002) tarafından geliştirilmiştir. Yöntemde ardışık oluşturulan ağaçlarda daha önceki aşamalarda yanlış tahmin edilen kısımlara ekstra ağırlık⁶ verilerek tahminin iyileştirilmesi hedeflenir (Liaw ve Wiener, 2002). Yani, her bir ağaç bir önceki ağaçtan elde edilen artıklara göre oluşturulur. Buna göre bir ağaçta iyi kestirim yapılamayan gözlemler daha büyük artıklara sahip olur. Bir sonra oluşturulan ağaçta ise bu tahmini iyileştirebilmek için, bahsedilen gözlemin ağırlığı daha fazla olur (West, 2012, s. 289). Bu anlamda yöntem torbalama ve rastgele orman algoritmalarından ayrılır. Torbalama ve rastgele ormanda önyüklemeli örneklem çekme yapılırken artırma da böyle bir durum

⁵ Makine öğrenme alanında temel öğrenici kavramı sıklıkla kullanılır. Buna göre temel öğrenici, genellikle topluluk öğrenme modeli içindeki bireysel öğrenicilerin her biridir. Bunlar genellikle karar ağacı, sinir ağı vb. makine öğrenme algoritmaları tarafından eğitim verilerinden üretilir (Zhou, 2009, s. 270). Zayıf öğrenici ise rastgele tahminden yalnızca çok az düşük hataya sahip sınıflandırıcılar üreten öğrenme algoritmasıdır. Güçlü öğrenici ise yeterli eğitim verisi olduğu takdirde küçük hataya sahip sınıflandırıcılar üretir (Ferreira ve Figueiredo, 2012, s. 35).

⁶ Pek çok artırma algoritması vardır. Bunlar ağırlıkları hesaplama şekillerine göre farklılık gösterirler (El Morr ve ark., 2022, s. 431).

söz konusu değildir. Artırmada her ağaç orijinal veri setinin değiştirilmiş bir biçimine uydurulur. Başka bir deyişle rastgele orman ve torbalamada ağaçlar birbirlerinden bağımsızken, artırmada bağımlıdır. Rastgele ormanda varyansın azaltılması için güçlü öğrenciler bağımsız ve özdeş dağılıma sahip bir şekilde birleştirilirken, artırmada zayıf bir öğrenci güçlü bir topluluğa dönüştürülür (Ferreira ve Figueiredo, 2012, s. 41; Friedman ve ark., 2000; James ve ark., 2013, s.321). Bu bağlamda artırma algoritmasının seri, rastgele ormanların paralel çalıştığı düşünülebilir (West, 2012, s. 289). Rastgele orman algoritması birbirinden öğrenmeyen ve eşit öneme sahip uzmanlar grubunun kullanılmasına benzetilirken, artırma bilgi geçmişleri rastgele ormanlara göre daha benzer bir uzman grubu kullanılmasına benzetilir. Bu benzetmeye göre artırmada, uzman 2'nin uzman 1'in hatalarından, uzman 3'ün uzman 2'nin hatalarından öğrendiği (böyle devam eden) bir süreç vardır (Sinharay, 2016; Sinharay ve ark., 2019). Artırma algoritmaları içinden en popüler olanı gradyan artırmadır (Sinharay ve ark., 2019). Buna bağlı olarak metnin devamında regresyon ağaçlarını temel öğrenci olarak kullanan gradyan artırma algoritmasıyla ilgili açıklamalara yer verilmiştir.

Artırmada da diğer topluluk öğrenme yöntemlerinde olduğu gibi çok sayıda karar ağacından yapılan $\hat{f}^1, \dots, \hat{f}^B$ (B ağaç sayısıdır) tahminlerinin birleştirilmesi gerekir. Süreç sadece birkaç terminal düğümü olan küçük bir regresyon ağacıyla başlar. Her ağaçtaki bölünme sayısı d parametresi ile kontrol edilir. Oluşturulan ağaç, önceki adımda oluşan ağaçtaki artıklara uydurulur ve iyi tahmin yapılamayan gözlemlerde $\hat{f}$ yavaşça iyileştirilir. Ardından algoritmanın öğrenme hızını kontrol eden küçültme (λ -shrinkage) parametresi ile güncelleme yapılır. Bu parametrenin mümkün olduğunca küçük olması daha yavaş öğrenen bir model ortaya koyacağı için daha iyi performans elde edilmesini sağlar (James ve ark., 2013, s. 321-322). Bununla birlikte iterasyon sayısı (ağaç sayısı) ve küçültme parametresi arasında bir denge vardır. Küçültme parametresinin değeri azaldıkça iterasyon sayısı artar. Ancak bu parametrenin değerinin 0.125'ten küçük olduğu durumlarda tahmin

performansında azalan bir kazanç ile iyileşmenin devam ettiği gösterilmiştir (Friedman, 2001). Algoritma özet olarak şu şekilde açıklanabilir (James ve ark., 2013, s. 323):

1. Eğitim veri setindeki tüm i gözlemleri için $\hat{f}(x) = 0$ ve $r_i = y_i$ olacak şekilde ayarlanır.

2. $b=1, \dots, B$ için aşağıdaki adımlar tekrarlanır:

(a) Eğitim verilerine (X, r) d kadar bölünmeye sahip (yani, $d+1$ terminal düğümlü) $\hat{f}^b$ ağacı uydurulur.

(b) Oluşturulan yeni ağacın küçültme parametresi ile ağırlıklandırılmış versiyonu önceki ağaca eklenerek $\hat{f}$ güncellenir:

$$\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x)$$

(c) Artıklar güncellenir:

$$r_i \leftarrow r_i - \lambda \hat{f}^b(x_i)$$

3. Artırılmış modelden kestirim yapılır:

$$\hat{f}(x) = \sum_{b=1}^B \lambda \hat{f}^b(x)$$

Artırma algoritmasında iterasyon sayısı ve küçültme parametresinin ayarlamalarla optimum değerlerinin bulunması gerekir (Friedman, 2001). Ek olarak her ağaçtaki bölünme sayısının da ayarlanması gerektiği ifade edilmiştir (James ve ark., 2013, s. 322). İterasyon sayısı topluluktaki ağaç sayısıdır ve çapraz doğrulamayla seçilir. Çalışmada kullanılan gbm (Ridgeway ve ark., 2024) paketindeki *n.trees* argümanına karşılık gelen iterasyon sayısının varsayılan değeri 100'dür. *shrinkage* argümanı ile ifade edilen küçültme parametresinin varsayılan değeri 0.1'dir. Son olarak her ağaçtaki bölünme sayısı olan etkileşim derinliği ise topluluğun karmaşıklığını kontrol eder. Genellikle 1 değeri yani sadece iki bölünmesi olan kütükler iyi performans gösterir. Kütüklerde her ağaç bir değişken içerdiği için kütük topluluğu eklemeli bir modele karşılık gelir. Pakette *interaction.depth* argümanına karşılık

gelen bu parametrenin varsayılan değeri eklemeli modeli temsil edecek şekilde 1'e ayarlanmıştır (James ve ark., 2013, s. 322; Ridgeway ve ark., 2024).

Yukarıda genel hatlarıyla özetlenen artırma algoritması, gbm (Ridgeway ve ark., 2024) paketinde Friedman (2002) tarafından performansı artırdığı bulunan rastgele alt örnekleme adımının eklenmesini de içermektedir. Bu sebeple pakette rastgele alt örneklem kullanılarak kayıp fonksiyonunun (hata kareler toplamı) beklenen değeri, $E\left(z\left(y, \hat{f}(x)\right) \mid x\right)$, kestirilir. Dolayısıyla aşağıda Ridgeway'e (2024) göre bazı notasyonlarla birlikte pakette kullanılan algoritma açıklanmıştır.

Tahmin problemlerinde esas olarak bazı kayıp fonksiyonlarının, $\Psi(y, f)$, beklenen değerini en aza indiren $\hat{f}(x)$ fonksiyonunun bulunması söz konusudur. Pakette kayıp fonksiyonu *distribution* argümanı ile belirlenir. Regresyon problemleri için hata kareler toplamına karşılık gelen "gaussian" seçeneği kullanılır. Buna karşılık gelen notasyon şu şekilde gösterilir:

$$\Psi(y_i, f(x_i)) = \sum_{i=1}^N (y_i - f(x_i))^2$$

Bu durumda artırma işleminin genel formu:

$$\hat{f}(x) \leftarrow \hat{f}(x) + E(z(y, \hat{f}(x)) \mid x)$$

Herhangi bir artırma algoritmasının güncelleme adımında λ eklenir:

$$\hat{f}(x) \leftarrow \hat{f}(x) + \lambda E(z(y, \hat{f}(x)) \mid x)$$

Verilen açıklamalara bağlı olarak gbm paketinde uygulandığı haliyle artırma algoritması şu şekildedir:

Algoritma çalıştırılmaya başlamadan önce aşağıdaki argümanlara ilişkin değerlerin/ifadelerin belirlenmesi gerekir:

Kayıp fonksiyonu (distribution)

İterasyon sayısı, T (n.trees)

Etkileşim derinliği, K (interaction.depth)

Küçültme (öğrenme oranı) parametresi, λ (shrinkage)

Alt örnekleme oranı, p (bag.fraction) (varsayılan değeri 0.5'tir)

$\hat{f}(x) = \operatorname{argmin}_{\rho} \sum_{i=1}^N \Psi(y_i, \rho)$ olacak şekilde sabit bir $\hat{f}(x)$ ile başlanır. Her bir $t=1, \dots, T$ için:

1. Negatif gradyan hesaplanır

$$z_i = -\frac{\partial}{\partial f(x_i)} \Psi(y_i, f(x_i)) | f(x_i) = \hat{f}(x_i)$$

2. Veri kümesinden rastgele $p \times N$ gözlem seçilir

3. Sadece rastgele seçilen gözlemler kullanılarak K terminal düğümlü regresyon ağacı uydurulur, $g(x)=E(z|x)$.

4. Sadece seçilen rastgele gözlemler kullanılarak terminal düğümler için tahminler $\rho_1, \dots, \rho_K$ olarak hesaplanır

$$\rho_k = \operatorname{argmin}_{\rho} \sum_{x_i \in S_k} \Psi(y_i, \hat{f}(x_i) + \rho)$$

Burada S_k , k terminal düğümünü tanımlayan x_s kümesidir.

5. $\hat{f}(x)$ şu şekilde güncellenir

$$\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \rho_{k(x)}$$

Burada $k(x)$, x yordayıcılı bir gözlemin gönderileceği terminal düğümün indeksidir.

Değişkenlerin Göreli Önemi

Ağaç tabanlı yöntemler için Friedman (2001) tarafından geliştirilen x_j bağımsız değişkeninin göreli etkisi şu şekildedir (Ridgeway, 2024):

$$\hat{f}_j^2 = \sum_{x_j \text{ deki bölünmeler}} I_t^2$$

Buradaki I_t^2 , ilgili düğümde x_j üzerinde bölünme sonucu elde edilen ampirik iyileşmedir. Friedman'ın (2001) değişken görelî etkisinin artırma modelleri için geliştirdiği uzantısına göre, artırma algoritması tarafından oluşturulan tüm ağaçlardan elde edilen sonuçların ortalaması alınır (Ridgeway, 2024).

Kısmi Bağımlılık Grafikleri

Kısmi bağımlılık grafikleri çeşitli makine öğrenme algoritmalarından elde edilen sonuçların yorumlanmasında kullanılır. Bu grafiklerde bir bağımsız değişkenin bağımlı değişkenle ilişkisi diğer bağımsız değişkenler dikkate alınarak hesaplanır (Friedman, 2001).

Topluluk Öğrenme Yöntemlerinin Kullanıldığı Araştırmalar

Strobl ve Zeileis (2008) rastgele orman algoritmasındaki değişken önemi hesaplamalarından biri olan permütasyon önemi testinin farklı koşullar altındaki davranışını incelemiştir. Bu amaçla Breiman ve Cutler (2003) testinde yapılan, ortalama önemin ağaç sayısına bağlı elde edilen standart hataya göre normalleştirilmesinin etkisi araştırılmıştır. Simülasyon deneylerini içeren çalışmada testin ampirik gücü yordayıcı doğruluğu, ormandaki ağaç sayısı ve örneklem büyüklüğünün bir fonksiyonu olarak gösterilmiştir. İncelemeler standartlaştırılmamış ortalama önem (ham önem), ortalamanın standart hatası, standartlaştırılmış önem (z-puanı) ve güç eğrilerine göre yapılmıştır. 1000 tekrarlı her yinelemede sadece bir tane iki kategorili değişkenin bağımlı değişkenle ilgili olduğu beş yordayıcı içeren 100, 200 ve 500 örneklem büyüklüğünde veriler üretilmiştir. Rastgele orman parametreleri olarak toplam ağaç sayısı 100, 200 ve 500 iken bölünme başına seçilen rastgele değişken sayısı 2 olarak ayarlanmıştır. Grafiklerden elde edilen sonuçlara göre ölçeklemenin yapıldığı durumda testin gücünün, herhangi bir istatistiksel testten beklenenin aksine, örneklem büyüklüğüyle artmadığı görülmüştür. Ek olarak gücün (ağaç sayısına kıyasla) büyük örneklemelerde sıfır olduğu bulunmuştur. Sonuçta, ölçeklemenin keyfi bir ayar parametresi olan ağaç sayısına göre yapılmış olmasından

kaynaklı olarak testin davranışında istenmeyen etkilere yol açtığı ifade edilmiştir. Buna göre ham önem puanının daha iyi istatistiksel özelliklere sahip olduğu vurgulanmıştır.

Stekhoven ve Bühlmann (2012) parametrik olmayan, farklı ölçek türlerindeki değişkenlerin olduğu eksik değerlere sahip veri setleriyle başa çıkabilen ve rastgele orman algoritmasına dayanan bir eksik değer tamamlama yöntemi geliştirmeyi amaçlamışlardır. Yöntemde temel olarak gözlenen değerlerle modelin eğitilmesi ve buna göre eksik değerlerin kestirilmesi söz konusudur. Geliştirilen MissForest yönteminin test edilmesi için biyoloji alanından sadece sürekli, sadece kategorik ve karma tipteki değişkenleri içeren 10 farklı veri seti kullanılmıştır. Bu veri setleri %10, %20 ve %30 oranında tamamen rastgele kayıp değer içerecek şekilde düzenlenmiştir. MissForest yöntemiyle karşılaştırmalar için sadece sürekli türde değişkenlere sahip verilerde k-en yakın komşu ve MissPALasso (Missingness Pattern Alternating Lasso) yöntemleri kullanılmıştır. Sadece kategorik ve karma tipteki değişkenleri içeren veri setleri için ise k-en yakın komşu ve MICE (Multivariate Imputation by Chained Equations) yöntemleri kullanılmıştır. Bahsedilen bu dört yöntemden sürekli veriler için normalleştirilmiş hata kareler ortalamasının karekökü, kategorik veriler için ise yanlış sınıflandırılmış girdilerin oranı hesaplanmıştır. Değerlendirme yapılırken bu hata oranları ile MissForest yönteminde torba dışı kestirime göre hesaplanan imputasyon hatası Wilcoxon testiyle karşılaştırılmıştır. Sonuç olarak MissForest yönteminin aynı anda farklı ölçek türlerindeki değişkenlerde bulunan eksik değerlerle başa çıkmada daha başarılı olduğu görülmüştür. Yöntem özellikle karmaşık etkileşimler ve doğrusal olmayan ilişkiler içeren veri setlerinde daha iyi performans göstermiştir. Hesaplama süresi açısından uygun olduğu belirtilen yöntemin parametre ayarı gerektirmemesinin ve parametrik olmamasının önemi vurgulanmıştır. Öte yandan ayrı bir test setine ihtiyaç duymadan torba dışı kestirimlerle hata tahmini yapabilmesi ve yüksek boyutlu verilerle başa çıkabilmesi gibi özelliklerin avantaj sağladığı ifade edilmiştir.

Sinharay (2016) tarafından yapılan kapsamlı bir araştırmada çeşitli makine öğrenme yöntemleri ve geleneksel yöntemlerin performansı üç farklı durum için karşılaştırılmış ve

sonuçlar pratikteki önemi açısından değerlendirmiştir. Birinci analizde madde özelliklerinden (konu düzeyi, bilişsel alan, okuma yükü) faydalanarak madde parametrelerinin (madde güçlüğü, madde eğimi) tahmin edilmesi amaçlanmıştır. Bu amaçla TIMSS 2007 4. Sınıf matematik değerlendirmesine ait 166 adet iki kategorili madde içeren veri seti kullanılmıştır. Çoklu doğrusal regresyon, tekli regresyon ağacı, rastgele orman, gradyan artırma yöntemlerine göre hesaplamalar yapılmıştır. Sonuçta her iki madde parametresi için en iyi sonuçlar rastgele orman algoritmasından elde edilmiştir. Ardından artırma iyi sonuç vermiştir. Tekli regresyon ağacı da çoklu doğrusal regresyondan biraz daha iyi performans göstermiştir. Ancak makine öğrenme yöntemlerinin sağladığı iyileşmenin oldukça mütevazı olduğuna dikkat çekilmiştir. Bu iyileşmenin madde parametrelerinin tahmini bağlamında pratik etkisinin olmadığına ilişkin yorumda bulunulmuştur. İkinci analize göre lise terkinin tahmini yapılmıştır. Buna göre Florida'da bulunan bir bölgedeki 41 liseden okulu bırakma riski taşıyan 2374 öğrencinin verileri ele alınmıştır. Burada bağımlı değişken okulu bırakma durumu (0-1) iken birey ve okul düzeyinde bazı bağımsız değişkenler bulunmaktadır. Lojistik regresyon, çok düzeyli lojistik regresyon, tekli sınıflandırma ağacı, rastgele orman, gradyan artırma algoritmaları kullanılmıştır. Sonuçta artırma ve rastgele ormanın lojistik regresyon ve çok düzeyli lojistik regresyona göre biraz daha iyi performans gösterdiği görülmüştür. Rastgele ormanlar en iyi sonuçları vermiştir. Tekli sınıflandırma ağacı lojistik regresyondan biraz daha iyi performans göstermiştir. Denetimli öğrenme tekniklerinin geleneksel karşılıklarına kıyasla pratik etkisi bağlamında (katsayılar arasındaki küçük bir farkın bile yüz binlerce lise öğrencisinin okulu bırakmasına yönelik tahmin sağlaması açısından) daha iyi performans gösterdiği vurgulanmıştır. Üçüncü analizde ise ülke genelindeki bir elektronik kompozisyon değerlendirmesinden elde edilen 929 katılımcının verileri kullanılmıştır. Bağımlı değişken değerlendirmecilerin verdiği puanlar (1-4) iken 49 bağımsız değişkenle analizler gerçekleştirilmiştir. Hesaplamalar hem küçük ($n=100$) hem de büyük ($n=620$) eğitim verisinde yapılmıştır. Doğrusal regresyon, kümülatif lojistik regresyon, tekli sınıflandırma ağacı, rastgele orman ve gradyan artırma yöntemleri kullanılmıştır. Sonuçlara göre artırma

ve rastgele ormanlar en iyi performansı göstermiştir. Bunun pratikte anlamlı olduğu ifade edilmiştir. Küçük eğitim veri setleri için doğrusal regresyona göre denetimli öğrenme tekniklerinin sağladığı iyileşmenin daha büyük eğitim setine göre çok daha fazla olduğu görülmüştür. Özetle tüm analizler dikkate alındığında denetimli öğrenme yöntemlerinin bazılarının geleneksel karşılıklarına göre biraz daha iyi performans (oldukça mütevazı iyileşme olmakla birlikte) gösterdiği bulunmuştur. Buna ek olarak denetimli öğrenme tekniklerinden rastgele orman en iyi performansı göstermiştir. Bunu artırma algoritması takip etmiştir.

Qiao ve Jiao (2018) araştırmasında dört adet denetimli öğrenme tekniği (sınıflandırma ağacı, gradyan artırma, destek vektör makinesi, rastgele orman) ve iki denetimsiz öğrenme tekniğinin (kendini organize eden haritalar, k-ortalamalar kümeleme analizi) karşılaştırılması hedeflenmiştir. Bu amaçla PISA 2012 problem çözme değerlendirmesinde kullanılan karmaşık problem çözme becerisine dayalı Tickets ünitesinin birinci maddesine ait günlük dosyası verileri kullanılmıştır. Günlük dosyasından zamana ve eyleme dayalı özellikler elde edilmiştir. Buna göre toplam süre, ilk eyleme kadar geçen süre, son eylem için harcanan süre ve deney için harcanan süre değişkenleri oluşturulmuştur. Öte yandan eyleme dayalı özellikler kümesi ardışık eylemlerin birlikte kodlanmasına dayalı n-gram yaklaşımıyla oluşturulmuştur. Buna göre 12 unigram (tekli eylemler), 18 biagram (iki ardışık eylem içeren özellikler), 2 adet four-gram (dört adet ardışık eylem içeren özellikler) oluşturulmuştur. Bağımlı değişken olarak üç kategorili olarak kodlanmış (0-1-2) madde puanı kullanılmıştır. Elde edilen sonuçlara göre tüm makine öğrenme yöntemleri iyi performans sergilemiştir. Denetimli öğrenme teknikleri içinden gradyan artırma biraz daha iyi sonuç verirken, denetimsiz öğrenme teknikleri benzer tahmin performansı sergilemiştir. Öte yandan incelenen değişken önemi değerlerine göre, süre ile ilgili özelliklerin sınıflandırma için önemli olmadığı dikkat çekmiştir.

Sinharay ve arkadaşları (2019) çalışmalarında bazı süreç ve ürün özelliklerini içeren veri setlerinde gradyan artırma ve doğrusal regresyonun karşılaştırılmasını

amaçlamışlardır. Buna göre 2013 CBAL değerlendirmesine ait pilot testin 6. ve 9. Sınıflara uygulanan kısmıyla analizler gerçekleştirilmiştir. İki adet ikna edici türden kompozisyona ait tuş vuruşu kayıtları ve çeşitli ürün verilerinin kullanıldığı araştırmada örneklem büyüklükleri sırasıyla (veri çıkarma işleminden sonra) 823 ve 825'tir. Bağımlı değişken olarak en az iki kişi tarafından yapılan madde puanlaması dikkate alınmıştır. ETS tarafından geliştirilen tuş vuruşu günlük kaydı programıyla süreç özellikleri, e-rater otomatik puanlama sisteminden de ürün özellikleri elde edilmiştir. Çalışmada tüm analizler süreç verisi, ürün verisi ve her ikisini de içeren veri setleriyle ayrı ayrı gerçekleştirilmiştir. Temel sonuç olarak, gradyan artırmanın doğrusal regresyona göre bir miktar (mütevazi) iyileşme sağladığı vurgulanmıştır. Buna ek olarak performans tahmininde süreç özelliklerinin ürün özelliklerine göre marjinal bir katkısı olmadığı görülmüştür. Sadece ürün özelliklerini içeren veri setleriyle yapılan tahminin sonuçları ile süreç ve ürün özelliklerinin birlikte ele alındığı durumun sonuçları yaklaşık olarak benzer kalmıştır.

Güre ve arkadaşları (2020) çalışmasında çok katmanlı algılayıcı yapay sinir ağları ve rastgele orman yöntemleri (sınıflandırma ağaçlarına dayalı) karşılaştırılmıştır. Bu amaçla PISA 2015 matematik Türkiye değerlendirmesine ait veriler ($n=4422$) kullanılmıştır. 25 bağımsız değişkenin kullanıldığı araştırmada, bağımlı değişken değerlendirmeye ait on farklı olası değerlerin ortalamasının düşük-orta-yüksek olacak şekilde kategorilere ayrılmasıyla oluşturulmuştur. Hata kareler ortalamasının kökü, ortalama mutlak hata, bağıl mutlak hata, bağıl karesel hatanın kökü, gözlenen ve gerçek değerler arası korelasyona göre tahmin performansları karşılaştırılmıştır. Sonuçta rastgele orman algoritmasının daha iyi performans sergilediği bulunmuştur. Buna ek olarak Türkiye'deki öğrencilerin matematik performansını etkileyen en önemli değişkenin başarıya yönelik kaygı düzeyleri olduğu görülmüştür.

Man (2024) çalışmasında sanal test ortamlarında kopya davranışını tespit etmek için yanıt süreleri, göz izleme verileri, madde yanıtları, sınav kaygısı ve kişilik özellikleri gibi çoklu veri kaynaklarının makine öğrenme yöntemleriyle kullanımını incelemiştir. Bu amaçla

335 üniversite öğrencisinin farklı koşullar altında 10 soruluk bir muhakeme testini yanıtlaması gerekmiştir. Toplam 298 kişinin verilerinin kullanıldığı araştırmada 93 kişi ön bilgiye sahip olmayan, 98 kişi teste benzer alıştırma soruları alan, 107 kişi benzer sınav soruları ve cevaplarını alan grupta yer almıştır. Toplam 103 değişkenin elde edildiği araştırmada ön bilgi tespiti için k-ortalamlar, k-en yakın komşu ve rastgele orman algoritmaları kullanılmıştır. Analizlere geçmeden önce rastgele orman algoritmasıyla ön değişken seçimi yapılmıştır. Aynı zamanda çalışmada biyometrik özellikler ve sınav kaygısı göstergelerinin ön bilgi tespitindeki etkinliklerini ortaya çıkarabilmek için bu değişkenlerin varlığında ve yokluğunda rastgele orman algoritmasıyla analizler tekrarlanmıştır. Elde edilen sonuçlara göre rastgele orman ön bilgisi olanları tespit etmede daha başarılı olmuştur. Biyometrik özellikler ve sınav kaygısı bilgisinin eklenmesiyle rastgele ormanın hassasiyeti artmıştır. Yani bu özelliklerin ön bilgi tespitinde önemli olduğu bulunmuştur. Sonuçta farklı veri kaynaklarının (süreç verileri, biyometrik veriler, psikometrik veriler) ön bilgi tespitinde birlikte ele alınmasının ve bu tip bir analizde rastgele orman yönteminin kullanımının avantajlarına vurgu yapılmıştır.

Saidani ve arkadaşları (2024) tarafından yapılan çalışmada çevrimiçi video öğrenme ortamlarındaki öğrenci etkinliğine ilişkin verilerle akademik performansın tahmin edilmesinde topluluk öğrenme yöntemlerine dayalı bir model önerilmesi amaçlanmıştır. Buna göre 326 kişiye ait çevrimiçi öğrenme ortamı Moodle'da öğrenci etkinliğini gösteren veriler, öğrenci video etkileşim verileri ve öğrencilerin akademik verilerinden oluşan 20 özellik çalışmanın bağımsız değişkenlerini oluşturmuştur. Araştırmada rastgele orman, lojistik regresyon, Gaussian Naive Bayes, ekstra ağaç sınıflandırıcı, stokastik gradyan inişi, destek vektör makinesi, k-en yakın komşu, karar ağacı, gradyan artırma algoritmaları kullanılmıştır. Önerilen modelin değerlendirilmesi için 20 adet değişkenden oluşan orijinal özellik seti ve evrişimli sinir ağlarından (convolutional neural network) yararlanılarak elde edilen özellik setinden oluşan veri kümeleri için tüm topluluk öğrenme yöntemleriyle ayrı ayrı analizler gerçekleştirilmiştir. Ardından bireysel olarak iyi performans gösteren

algoritmaların ikili kombinasyonlarından elde edilen modeller üzerinde önerilen yöntem uygulanmıştır. Bu yöntem iki topluluk öğrenme algoritmasından elde edilen sonuçların birleştirilerek tek bir tahmin yapılmasını içerir. Elde edilen sonuçlara göre rastgele orman algoritması orijinal özellik setinde ve evrişim özellik setinde sırasıyla %85 ve %87 ile en yüksek doğruluk oranına sahip olmuştur. İkili olarak birleştirilmiş modellerden elde edilen metrik değerlerine göre ise orijinal veri seti için en iyi performansı rastgele orman+destek vektör makinesi bileşimi göstermiştir. Bu bileşik model evrişim özellik setinde %98.99 doğruluk oranıyla üstün performans göstermiştir. Önerilen modele göre rastgele orman ve destek vektör makinesi yöntemlerinden ayrı ayrı tahmin edilen sınıf olasılıkların ortalaması alınır. Ortalama alındıktan sonra en yüksek olasılığa sahip sınıf nihai tahmin olarak kullanılır. Sonuçta evrişimli sinir ağlarından elde edilen özellikler kullanılarak rastgele orman+destek vektör makinesi bileşimine göre yapılan tahminin en iyi performansı gösterdiği ifade edilmiştir.

Zhen ve Zhu (2024) çalışmasında sınavlarda kopya çeken bireyleri tespit etmek için hibrit bir makine öğrenme yaklaşımı önermişlerdir. Bunun için 3280 katılımcıya ait 253 çoktan seçmeli soruya verilen yanıtlar ve yanıt sürelerini içeren bir veri seti kullanılmıştır. Bu veri setinin içinde kopya çekme davranışında bulunduğu tespit edilen 94 kişi yer almaktadır. Çalışmada Naive Bayes, doğrusal diskriminant analizi, Gauss süreci, destek vektör makinesi, karar ağacı, rastgele orman, Extreme gradyan artırma (XGBoost), AdaBoost, lojistik regresyon, k-en yakın komşu, çok katmanlı algılayıcı ve TabNet algoritmaları kullanılmıştır. İlk olarak tüm algoritmalar sadece yanıtları, sadece yanıt sürelerini ve her ikisini de içeren üç farklı veri setine uygulanmıştır. Ardından karmaşık derin sinir ağı modeli olan TabNet ile tüm algoritmalar ikiye bölünmüş ve birleştirilerek hibrit modeller oluşturulmuş ve performansları incelenmiştir. Son olarak önerilen hibrit yöntem simülasyon çalışmasıyla test edilmiş ve aynı zamanda Stacking2 modeli ile de kıyaslanmıştır. Geliştirilen hibrit yönteme göre her bir algorithmadan elde edilen tahmin ağırlıklı ortalama alma yoluyla birleştirilir. Üç veri setinde bireysel olarak ele alınan algoritmalar içinden en iyi

sonuçları TabNet vermiştir. Genel olarak yanıt süresini içeren veri seti tüm algoritmalarda daha iyi performans göstermiştir. Hibrit modellerden TabNet-AdaBoost 0.92 AUC (area under the receiver operating characteristic curve) değeriyle üstün performans göstermiştir. Sonuç olarak kopya tespiti için makine öğrenme modellerini sinir ağı modelleriyle birleştirmenin önemine vurgu yapılmıştır.

Topluluk öğrenme yöntemlerine ilişkin araştırmalar birlikte ele alındığında kayıp veri tamamlamaya yönelik yöntem geliştirilmesine, topluluk öğrenme yöntemlerinin geleneksel karşılıklarıyla kıyaslanmasına, denetimli ve denetimsiz öğrenme yöntemlerinin karşılaştırılmasına, farklı topluluk öğrenme yöntemlerinin birleştirilerek hibrit modeller önerilmesine, kopya tespitinde topluluk öğrenme yöntemlerinin kullanımının incelenmesine ilişkin çalışmaların olduğu görülmüştür. Öte yandan çalışmalarda rastgele orman algoritmasının öne çıkan performansı da dikkat çekmiştir.

Koşullu Çıkarım Ağaçları

SRA'lar bölünme değişkenini ve kesme noktasını seçerken bir bilgi ölçütünü maksimize edecek (regresyon ağaçları için hata kareler toplamı) tüm olası bölünmeler üzerinden arama gerçekleştirir (Hothorn ve ark., 2006). Ancak bu yaklaşım hem daha fazla olası bölünme noktasına sahip (daha çok kategorili veya sürekli) bağımsız değişkenlerin lehine değişken seçim yanlılığına hem de aşırı uyuma yol açar (Hothorn ve ark., 2006). Bu kapsamda bir özyinelemeli bölünme algoritmasında bağımlı değişken ve bağımsız değişkenler arasında hiçbir ilişkinin olmadığı durum için, herhangi bir bağımsız değişkenin seçilme olasılığı ölçek türü veya eksik değer miktarından bağımsız olarak m adet yordayıcının tamamı için $1/m$ ise bu algoritma yansızdır (Hothorn ve ark., 2006). Değişken seçim yanlılığının varlığına ilişkin bir bulgu Strobl ve arkadaşları (2007) tarafından elde edilmiştir. SRA'lardaki aşırı uyum sorunuyla başa çıkabilmek için budama yapılır. Ancak bu noktada algoritmanın istatistiksel anlamlılık kavramına sahip olmaması ve bu nedenle bilgi ölçütünde önemli ve önemsiz bir iyileşme arasında ayırım yapamadığı sorunu ortaya çıkar

(Mingers, 1987). Bu sorunlarla başa çıkabilmek için parametrik olmayan ve ağaç tabanlı regresyon modellerini koşullu çıkarım teorisiyle bütünleştiren (Hothorn ve ark., 2015) koşullu çıkarım ağaçları algoritması Hothorn ve arkadaşları (2006) tarafından önerilmiştir. Bu algoritma kategorik, sıralı, sürekli, sansürlü, çok değişkenli bağımlı ve bağımsız değişkenleri ele alabilir (Hothorn ve ark., 2015). Algoritmada Strasser ve Weber (1999) tarafından geliştirilmiş permütasyon testleri ile özyinelemeli ikili bölünme yaklaşımları birleştirilmiştir. Buna göre, durdurma kuralı ve bağımlı-bağımsız değişkenler arasındaki ilişkinin anlamlılık düzeyi istatistiksel testlere göre belirlenir. Böyle bir yaklaşıma göre durdurma kuralının belirlenmesiyle oluşturulan koşullu çıkarım modelinin, en iyi tahmini veren budanmış ağaçlarla eşdeğer performansa yol açtığı gösterilmiştir. Dolayısıyla bu algoritmayla hem aşırı uyum problemine hem de sürecin istatistiksel bir altyapıdan yoksun olması sınırlılığına çözüm sunulmuştur (Hothorn ve ark., 2006). Algoritmada özyinelemeli bölünme çerçevesinde bağımlı değişkenin (Y) bağımsız değişkenlere ($X=X_1, \dots, X_m$ olmak üzere m boyutlu) göre koşullu dağılımını içeren regresyon modeli esas alınır. Modelin kurulduğu eğitim veri seti $\mathcal{L}_n$ şu şekilde ifade edilir:

$$\mathcal{L}_n = \{(Y_i, X_{1i}, \dots, X_{mi}); i = 1, \dots, n\}$$

Bu eğitim veri seti için koşullu çıkarıma dayalı özyinelemeli ikili bölünme algoritması, negatif olmayan tamsayı değeri alabilen ağırlıklar ($w=w_1, \dots, w_n$) kullanılarak ifade edilir. Burada ağacın her bir düğümü, ilgili gözlemler düğümün elemanı ise sıfırdan farklı değilse sıfır değerini alan ağırlıklara sahip olurlar. Buna göre algoritma şu şekildedir:

1. Ağırlık vektörü w için, m bağımsız değişkenden herhangi biri ile bağımlı değişken arasında bağımsızlık olduğunu ifade eden global sıfır hipotezi test edilir. Bu hipotez reddedilmezse süreç durur. Eğer reddedilirse bağımlı değişkenle en güçlü ilişkiye sahip (yani, en küçük p değerine sahip) X_{j^*} bağımsız değişkeni bölünme için seçilir.

2. X_{j*} değişkenine ilişkin A^* ve $X_{j*} \setminus A^*$ olmak üzere iki ayrık küme içeren bölme yapabilmek için sol ve sağ ağırlıklar belirlenir. Buna göre $w_{left,i} = w_i I(X_{j*i} \in A^*)$ ve $w_{right,i} = w_i I(X_{j*i} \notin A^*)$ ağırlıklarına göre iki alt grup belirlenir.

3. İlk iki adım sırasıyla sol ve sağ ağırlıklar değiştirilerek tekrarlanır (Hothorn ve ark., 2006)

Özetle algoritmanın birinci adımında değişken seçimi yapılır ve durdurma kriteri belirlenir. İkinci adımında ise bölme kriteri belirlenir. Değişken ve bölme noktası seçimleri SRA'larda eş zamanlı gerçekleştirilir. Buna göre daha fazla kesme noktasına sahip değişkenlerin şans eseri daha iyi bir bölünme kriteri oluşturma olasılığı daha yüksektir. Örneğin, yedi kesme noktasına sahip dört kategorili bir değişkenin, tek kesme noktasına sahip iki kategorili bir değişkenden daha yüksek bir bölünme kriteri üretme ihtimali daha yüksek olacağı için genellikle dört kategorili değişkenin bölünme için seçilmesi söz konusudur (Strobl ve ark., 2007). Bu seçimlerin koşullu çıkarım ağaçlarında art arda yapılması çok sayıda kayıp değere ya da çok sayıda olası bölme noktasına sahip değişkenlerin lehine seçim yanlılığını engellemektedir. Öte yandan 1. Adımda test edilen sıfır hipotezinin daha önceden belirlenmiş α düzeyinde reddedilme durumu dikkate alınır. Eğer reddedilmezse durulur. Algoritmadaki tüm hipotez testleri permütasyon testlerine göre yapılır. Algoritmanın birinci adımında kullanılan bir bağımsız değişken ile bağımlı değişkenin bağımsızlığının test edilmesine yönelik test istatistiği şu şekildedir:

$$T_j(\mathcal{L}_n, w) = vec\left(\sum_{i=1}^n w_i g_j(X_{ji}) h(Y_i, (Y_1, \dots, Y_n))^T\right) \in \mathbb{R}^{pq}$$

Buna göre test istatistiği $\mathcal{L}_n$ eğitim veri setine, w ağırlıklarına, h etki fonksiyonuna ve g dönüşümüne bağlıdır. Bu test istatistiği kullanılmadan önce standartlaştırılır ve maksimumu alınarak skaler bir değere dönüştürülür. Buna göre standartlaştırılmış doğrusal test istatistiği:

$$c_{max}(t, \mu, \Sigma) = \max_{k=1, \dots, pq} \left| \frac{(t - \mu)_k}{\sqrt{(\Sigma)_{kk}}} \right|$$

Burada t ; test istatistiği T 'nin gerçek değeri, μ ; bağımsızlık altında T 'nin beklenen değeri ve $(\Sigma)_{kk}$; kovaryans matrisinin k . köşegen değeridir. Dolayısıyla c , her permütasyon için hesaplanan test istatistiğidir (Hothorn ve ark., 2006). Bu değer analizlerin gerçekleştirildiği paket olan partykit (Hothorn ve Zeileis, 2015; Hothorn ve ark., 2006) içindeki `ctree_control` fonksiyonunda bulunan `teststat` argümanı ile belirtilir (Hothorn ve ark., 2015). Bu argüman değişken seçimi için uygulanacak test istatistiğinin türünü belirtir. Pakette *quadratic* (varsayılan değer) ve *maximum* olmak üzere iki seçenek yer alır (Hothorn ve Zeileis, 2023, s. 10). Varsayılan değer $c_{quad}(t, \mu, \Sigma) = (t - \mu)\Sigma^+(t - \mu)^T$ formülüyle hesaplanır (Hothorn ve ark., 2006). Öte yandan test istatistikleri tüm bağımsız değişkenler aynı ölçekte olmadığında yansız değildir. Bu sebeple değişken seçimindeki yanlılığı engellemek için p değeri hesaplanmalıdır. Bu değer test istatistiğinden faydalanarak elde edilir. Buna göre H_0^j sıfır hipotezi için testin P değeri şu şekildedir (Hothorn ve ark., 2006):

$$P_j = \mathbb{P}_{H_0^j}(c(T_j(\mathcal{L}_n, w), \mu_j, \Sigma_j) \geq c(t_j, \mu_j, \Sigma_j) | S(\mathcal{L}_n, w))$$

Bu değeri elde etmek için paketteki aynı fonksiyonun `testtype` argümanı kullanılır. Bununla test istatistiğinin dağılımı (yani, p değeri) belirlenir. Bonferroni, Monte Carlo, Univariate ve Teststatistic olmak üzere dört farklı değer alabilen bu argümanın ilk üç seçeneğinde kriter olarak p değeri hesaplanır. Teststatistic seçeneğinde ise kriter olarak ham istatistik kullanılır. Varsayılan seçenek olan Bonferroni'de düzeltilmiş p değeri hesaplanır (Hothorn ve Zeileis, 2023, s. 11). Düzeltilmiş p değerinin minimumu önceden belirlenmiş α değerinden küçük ise sıfır hipotezi reddedilir ve süreç devam eder, aksi takdirde algoritma durur. Bu sebeple α ağaç boyutunu belirleyen bir parametre olarak görülür. Dolayısıyla α bir yandan bağımsızlık testlerinde kullanılan eşik değeri olması açısından değişken seçiminde rol oynarken diğer yandan bir durdurma kuralı anlamında ağaç boyutunu belirler (Hothorn ve ark., 2006). Pakette α 'nın varsayılan değeri 0.05'tir. Bununla birlikte bölme yapılması için aşılması gereken eşik *mincriterion* argümanı ile belirtilir. Bunun değeri eğer `testtype` argümanının değeri "TestStatistic" ise test istatistiğinin değerine, `testtype`'ın diğer seçenekleri için ise $1 - \alpha$ 'ya karşılık gelir. Örneğin, *mincriterion*

0.95 ise bu düğümün bölünmesi için p değerinin 0.05'ten küçük olması gerekir. Bu yaklaşımla budama yapılmaksızın doğru büyüklükte ağaç elde edilmiş olur. Bu parametrenin değeri veri setinin büyüklüğüne göre değiştirilebilmekle beraber (0.95 değeri genellikle küçük ve orta büyüklükteki veri setleri için uygundur) hiper-parametre olarak da görülebilir (Hothorn ve Zeileis, 2023, s. 8).

Algoritmanın ikinci adımında gerçekleştirilen bölünme noktasının seçimi için yine permütasyon testi yapılır. Bölünmenin iyiliği aşağıda verilen doğrusal istatistikle ölçülür. X_{j*} değişken uzayının tüm olası A alt kümeleri için bu istatistik :

$$T_{j*}^A(\mathcal{L}_n, w) = \text{vec}\left(\sum_{i=1}^n w_i I(X_{j*i} \in A) h(Y_i, (Y_1, \dots, Y_n))^T\right) \in \mathbb{R}^q$$

Buradan da tüm olası A alt kümelerine göre maksimize edilmiş test istatistiğine sahip bölünmüş A^* belirlenir:

$$A^* = \text{argmax}_A c(t_{j*}^A, \mu_{j*}^A, \Sigma_{j*}^A)$$

İkinci adımda $c(t_{j*}^A, \mu_{j*}^A, \Sigma_{j*}^A)$ 'nin dağılımını (p değerini) hesaplamaya gerek yoktur. Burada gereksiz bölünmeleri engellemek için örneğin, olası bir bölünmenin neden olduğu iki alt grubun ağırlıkları toplamına kısıtlama getirilebilir (Hothorn ve ark., 2006). Bu işlem paketteki *minsplitt* argümanı ile yapılır. Örneğin buna ilişkin değer 20 ise, sol ve sağ alt düğümlerdeki ağırlık toplamının 20'yi aşmaması gerekir (Hothorn ve ark., 2015). Bunlara ek olarak koşullu çıkarım ağaçlarında da eksik veriler, SRA'larda olduğu gibi, vekil bölünmelere göre değerlendirmeye alınabilir (Hothorn ve ark., 2006).

Değişken Önemi

Koşullu çıkarım ağaçları için değişken önemi sıralamaları daha uygun grafiksel gösterim sunması sebebiyle DALEX (Biecek, 2018) paketiyle gerçekleştirilmiştir. Bu paketteki değerler permütasyona dayalı önem hesaplamasına göre elde edilir. Kayıp fonksiyonu olarak regresyon ağaçları için RMSE değeri kullanılır (Biecek ve Burzykowski, 2020).

Koşullu Çıkarım Ağaçlarına Dayalı Torbalama ve Rastgele Orman

Bu algoritmalar kullanılan temel öğrenciler ve kestirimlerin hesaplanma biçimine göre SRA'lara dayalı torbalama ve rastgele orman algoritmalarından farklılık gösterirler. Buna göre temel öğrenciler olarak koşullu çıkarım ağaçlarını kullanırlar. Yani torbalama ve rastgele orman oluşturulurken topluluktaki ağaçlar koşullu çıkarım ağaçlarından oluşur. Ek olarak rastgele orman algoritmasında tahmin için tüm ağaçlardan yapılan kestirimlerin doğrudan ortalaması alınırken, bu algorithmada her bir ağaçtan elde edilen gözlem ağırlıklarının ortalaması alınır (Hothorn ve Zeileis, 2023, s. 2,5). Temel öğrenci olarak koşullu çıkarım ağaçlarını kullanmaları sebebiyle SRA'lara dayalı rastgele orman algoritmasındaki gibi değişken seçim yanlılığına yol açmamaktadırlar. Bu durum Strobl ve arkadaşları (2007) tarafından gösterilmiştir. Bu algoritmalara dayalı analizler R yazılımında bulunan partykit (Hothorn ve Zeileis, 2015; Hothorn ve ark., 2006) paketindeki cforest fonksiyonuyla gerçekleştirilmiştir.

Koşullu Değişken Önemi

SRA'lara dayalı rastgele orman algoritmasındaki değişken seçim yanlılığı esasında tekli ağaçların oluşturulma prensibinden kaynaklanmaktadır (Strobl ve ark., 2007). Bu yanlılıkla başa çıkabilmek için SRA'ları temel öğrenci olarak kullanan randomForest fonksiyonu ve koşullu SRA'ları temel öğrenci olarak kullanan cforest fonksiyonuyla hesaplanan değişken önemi ölçütleri birbirleriyle ilişkisiz yordayıcılar bağlamında karşılaştırılmıştır. Sonuçta randomForest paketindeki Gini önem ölçütünün yanlı olduğu, permütasyon öneminin ise cforest'da bootstrap yerine değiştirmesiz alt örnekleme (varsayılan) ⁷ ile kullanıldığında güvenilir olduğu bulunmuştur (Strobl ve ark., 2007). Değişken seçim yanlılığını ortaya koyma amaçlı bir başka simülasyon da Loh (2014b)

⁷ Önyükleme işleminde (bootstrapping) aynı veri setinden alt kümeler oluşturulurken tüm gözlemler kullanılır ve bir sonraki alt küme oluşturulmadan önce kullanılan gözlem esas veri setine tekrar gönderilir. Buna aynı zamanda "değiştirme ile örnekleme" de denir (El Morr ve ark., 2022, s. 14).

tarafından yapılmıştır. Karşılaştırılan algoritmalar içinden cforest'ın, randomForest ve rpart paketiyle hesaplananlara göre daha düşük yanlışlık gösterdiği bulunmuştur.

Kesme noktası sayısına bağlı değişken seçim yanlışlığına Strobl ve arkadaşları (2007) tarafından çözüm önerisi sunulmuş olsa da ilişkili yordayıcıların varlığı permütasyon öneminde başka bir yanlışlık kaynağı oluşturmaktadır (Strobl, Malley ve ark., 2009). Rastgele orman algoritması için incelenen bu duruma göre ilişkili yordayıcıların önem değerlerinin yapay bir şekilde fazla hesaplanması söz konusudur (Strobl ve ark., 2008). Buna çözüm olarak ise Strobl ve arkadaşları (2008) tarafından değişken önemini daha güvenilir bir şekilde ortaya koyan koşullu permütasyon önemi ölçütü geliştirilmiştir. partykit paketindeki varimp fonksiyonuyla (conditional=TRUE seçeneğiyle) elde edilen koşullu önem puanları regresyon modellerindeki beta katsayıları anlamında koşulludur (Hothorn ve Zeileis, 2023, s. 55).

Önem değerinin yapay şekilde artması durumu Strobl, Hothorn ve arkadaşları (2009) tarafından sahte korelasyona dayanan bir örnekle açıklanmıştır. Örneğe göre 200 öğrencinin okuma puanı bağımlı değişken iken yaş, testin dilinin anadili olarak konuşulup konuşulmadığı ve ayakkabı numarası bağımsız değişkenlerdir. Burada hem test puanı hem de ayakkabı numarası etkili bir yordayıcı olan yaş ile ilişkili olduğu için ortaya çıkan sahte korelasyon sebebiyle ayakkabı numarasıyla test puanı yüksek ilişkide gibi görünmektedir. Bu bağlamda rastgele ormandaki koşulsuz permütasyon önemine göre ayakkabı numarası oldukça yüksek önem puanına sahip olmuştur. Buna karşılık koşullu permütasyon önemine göre hesaplama yapıldığında aynı değişkenin önem puanı olması gerektiği gibi oldukça düşük çıkmıştır (Strobl, Malley ve ark., 2009). Dolayısıyla tek başına değerlendirildiğinde etkili görülen bir değişken, diğer değişkenlerle koşullu olacak biçimde incelendiğinde etkisiz olabilir (Strobl ve ark., 2008). Buna kapsamda metnin devamında Strobl ve arkadaşlarına (2008) göre koşullu permütasyon önemi açıklanmıştır.

Rastgele orman algoritmasındaki koşulsuz permütasyon önemi ve koşullu permütasyon önemi şeması Şekil 5'te verilmiştir.

Şekil 5

Permütasyon Önemi Şeması (Sol: koşulsuz permütasyon önemi. Sağ: koşullu permütasyon önemi)

Y	X_j	Z
y_1	$x_{\pi_j(1),j}$	z_1
$\vdots$	$\vdots$	$\vdots$
y_i	$x_{\pi_j(i),j}$	z_i
$\vdots$	$\vdots$	$\vdots$
y_n	$x_{\pi_j(n),j}$	z_n

Y	X_j	Z
y_1	$x_{\pi_j Z=a(1),j}$	$z_1 = a$
y_3	$x_{\pi_j Z=a(3),j}$	$z_3 = a$
y_{27}	$x_{\pi_j Z=a(27),j}$	$z_{27} = a$
y_6	$x_{\pi_j Z=b(6),j}$	$z_6 = b$
y_{14}	$x_{\pi_j Z=b(14),j}$	$z_{14} = b$
y_{21}	$x_{\pi_j Z=b(21),j}$	$z_{21} = b$
$\vdots$	$\vdots$	$\vdots$

Şekil 5'in sol tarafında verilen koşulsuz permütasyon öneminde yordayıcı X_j 'nin hem bağımlı değişken (Y) hem de geri kalan yordayıcılar olan $Z = X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_p$ 'ye göre permütasyonu ifade edilmektedir. Sağ tarafta verilen koşullu permütasyon öneminde ise X_j değişkeninin yalnızca $Z=z$ olan gözlem grupları içinde permütasyonunun alındığı ifade edilmiştir. Sağdaki şema için koşullama işlemi değişkenler kategorik olduğunda basittir (Strobl ve ark., 2008). Buna karşılık sürekli değişkenler oldukça geniş aralıkta değerlere sahip olabildikleri için koşullandırma sırasında çok seyrek hücreler oluşabilir. Bu da permütasyonu anlamsız hale getirir. Buna çözüm olarak sürekli değişkenler kategorilere ayrılır (Strobl, Hothorn ve ark., 2009). Bu işlem için kesme noktası olarak koşullu çıkarım ağaçlarından elde edilen bölme noktası değerleri kullanılır. Sonuç olarak koşullu permütasyon önemi algoritması şu şekildedir (Strobl ve ark., 2008):

1. Permütasyondan önce torba dışı hata kestirimi yapılır.
2. Koşullandırılacak tüm Z değişkenleri için: ağaçta ilgili değişkeni bölen kesme noktaları elde edilir ve her kesme noktasında veri seti ikiye bölünerek permütasyon şeması oluşturulur.
3. Bu şemaya göre etkisi araştırılan değişkenin değerleri permütasyona tabi tutulur ve permütasyondan sonra torba dışı hata kestirimi yapılır.

4. Permütasyon önceki ve sonraki hata değerlerinin farkı hesaplanır ve ağaç sayısına göre ortalaması alınır.

Algoritmaya göre koşullamada kullanılacak Z değişkenlerinin de belirlenmesi gerekir. Bu değer partykit paketindeki varimp fonksiyonunda bulunan threshold argümanı ile belirlenir. Varsayılan olarak, Z değişkenlerinin her birinin X_j ile korelasyonuna ait test istatistiğinin veya 1-p değerinin 0.2'yi aşması gerekir. Bunu aşan değişkenler koşullandırma şemasına dahil edilir (Hothorn ve Zeileis, 2023, s. 54).

Koşullu Çıkarım Ağaçlarına Dayalı Gradyan Artırma

Bu algorithmada temel öğrenciler olarak koşullu çıkarım ağaçlarının kullanıldığı gradyan artırma uygulanmaktadır (Hothorn ve ark., 2023, s.22). R yazılımındaki mboost (Hothorn ve ark., 2024) paketiyle bu hesaplamalar yapılabilir. Model tabanlı artırma (Model-based Boosting) ifadesinin kısaltmasına göre isimlendirilmiş bu pakette esasında yüksek boyutlu veriler için genelleştirilmiş doğrusal, eklemeli (yordayıcıların etkisinin toplanabildiği modeller) ve etkileşim içeren modellerin uydurulması (fitting) için temel öğrenciler olarak bileşen bazlı (cezalandırılmış) en küçük kareler tahminleri veya regresyon ağaçları kullanılmaktadır. Pakette genel kayıp fonksiyonlarını uygun hale getirmek için (optimize etmek) artırma algoritması uygulanır (Hothorn ve ark., 2010; Hothorn ve Bühlmann, 2006). Bu kayıp fonksiyonları kullanıcı tarafından değiştirilebilir. Bu şekilde farklı artırma algoritmalarının uygulanmasına izin veren paketin oldukça esnek olduğu söylenebilir (Bühlmann ve Hothorn, 2007). Hem bu anlamda hem de pek çok regresyona dayalı temel öğrencinin kullanılabilmesi açısından paket oldukça kapsamlıdır (Mayr ve ark., 2014). Burada kullanılan fonksiyonlar için bağımlı değişken kategorik, sıralı, sürekli, sansürlü ya da sayım (count) ölçek türünde olabilir (Hothorn ve ark., 2024, s. 2). Pakette artırma algoritması tek bir uygulamayla tüm olası modellere (doğrusal, eklemeli, ağaç tabanlı) ve tüm olası kayıp fonksiyonlarına göre oluşturulabilir (Hothorn ve ark., 2010). Bu şekilde bağımsız değişkenlerin bağımlı değişkenle arasındaki doğrusal veya daha yumuşak

(smooth) ilişkiye bağlı olarak ortaya çıkabilecek model seçim yanlılığı engellenmiş olur. Çünkü, bu iki tür ilişki söz konusu olduğunda gerçek etkiden bağımsız olarak artırma algoritmasının neredeyse her zaman, daha esnek olduğu için, yumuşak temel öğrenciyi tercih ettiği görülür (Hofner ve ark., 2011). Belirtildiği gibi eş zamanlı bir uygulamayla araştırmacıya modeller arasında karşılaştırma yapma imkanı sunulmuş olur.

Çalışmadaki hesaplamalar mboost paketindeki blackboost() fonksiyonuyla gerçekleştirilmiştir. Bu fonksiyonda esasında gbm (Ridgeway ve ark., 2024) paketindeki gradyan artırma algoritması uygulanır. İkisi arasındaki temel fark kayıp fonksiyonunun mboost paketinde keyfi seçilmesiyle gbm paketinde sabit belirlenmiş bir fonksiyonun olmasıdır (Hothorn ve ark., 2024, s. 22). Bununla birlikte hem gbm paketindeki bu fonksiyon hem de blackboost fonksiyonundaki varsayılan kayıp fonksiyonu en küçük kareleri kullanan “gaussian” ifadesine karşılık gelir (Hothorn ve ark., 2024, s.22; Ridgeway ve ark., 2024, s.6). Diğer bir fark ise blackboost fonksiyonunda daha esnek olan koşullu çıkarım ağaçlarının temel öğrenci olarak kullanılmasıdır (Hothorn ve ark., 2024, s.22).

Değişken Önemi

Koşullu çıkarım ağaçlarına dayalı gradyan artırma için DALEX (Biecek, 2018) paketiyle değişken önem grafikleri elde edilmiştir. Paketteki hesaplamaların gerçekleşme biçimi “Koşullu Çıkarım Ağaçları” başlığı altında verilen “Değişken Önemi” bölümünde yazılanlarla aynıdır.

Koşullu Çıkarım Ağaçlarına Dayalı Algoritmaların Kullanıldığı Araştırmalar

Hothorn ve arkadaşlarının (2006) çalışmasında koşullu çıkarım ağaçları önerilmiş ve ampirik örnekler üzerinden rpart, QUEST ve GUIDE algoritmalarıyla performansı karşılaştırılmıştır. Bu kapsamda hem simülasyon çalışması hem de sınıflandırma ve regresyon ağaçlarını içeren gerçek veri setlerinde hesaplamalar yapılmıştır. Yöntemlerin performansları güç, tahmin doğruluğu, terminal düğüm sayısı, ağaçların yapısı gibi ölçütlere göre kıyaslanmıştır. Simülasyon çalışması için 3 değişken sürekli, 1 değişken rastgele %25

eksik değere sahip, 1 değişkenin değerleri bir basamağa yuvarlanmış, 1 değişken de kategorik (0-1) olacak şekilde üretilmiştir. Buna göre rpart ile elde edilen ağaçların birçok kesme noktasına sahip değişkenler lehine yanlılık gösterdiği bulunmuştur. ctree (koşullu çıkarım ağaçları) ile yapılan hesaplamalarda bu soruna rastlanmadığı vurgulanmıştır. Öte yandan tüm algoritmaların karşılaştırılmasına göre koşullu çıkarım ağaçlarının tahmin doğruluğunun diğer üç algoritmayla rekabet edebildiği (neredeyse eşdeğer olduğu) ifade edilmiştir. Buna karşılık koşullu çıkarım ağaçlarının farklı değişkenleri bölünmelerde kullanabilmesi sebebiyle, farklı algoritmalarından elde edilen ağaçların tahmin doğruluğunun benzer olmasına rağmen yapısal olarak farklı olabileceği ve yorumlamanın ihtiyatlı yapılması gerektiğine dikkat çekilmiştir. Sonuç olarak koşullu çıkarım ağaçlarının hem tahmin hem de ilişkileri açıklama için çok uygun olduğu ifade edilmiştir.

Strobl ve arkadaşları (2007) çalışmalarında farklı ölçek türlerindeki yordayıcıların varlığında rastgele orman algoritmasındaki değişken önem ölçütlerinin yanlılıklarını incelemişler ve bu duruma bir çözüm önerisi sunmuşlardır. Bu amaçla daha fazla kesme noktasına sahip değişkenlerin lehine olan seçim yanlılığı incelenmiştir. Sınıflandırma ağaçlarının kullanıldığı araştırmada bağımlı değişken iki kategorili olacak şekilde üretilmiştir. Öte yandan birbirleriyle ilişkisiz olacak şekilde beş farklı yordayıcı üretilmiştir. Bu değişkenlerden bir tanesi sürekli diğerleri sırasıyla 2,4,10 ve 20 kategorilidir. Örneklem büyüklüğünün 120 olarak seçildiği çalışmada tüm simülasyonlar 1000 tekrar üzerinden yapılmıştır. randomForest paketi ve cforest paketiyle hesaplamalar gerçekleştirilmiştir. Karşılaştırmalarda kullanılan değişken önem ölçütleri ise seçim sıklığı, Gini önemi ve permütasyon önemidir. Gini önemi cforest paketinde uygulanmadığı için sadece randomForest paketiyle yapılan hesaplamalarda kullanılmıştır. Üç değişken önemi bootstrap örnekleme (replace=TRUE) ve değiştirmesiz alt örnekleme (replace=FALSE) koşulları altında incelenmiştir. Bu incelemeler iki farklı hipotez altında ayrı ayrı ele alınmıştır. Buna göre boş durum olarak adlandırılan ilk hipotez hiçbir yordayıcının bağımlı değişkenle ilgili olmadığını ifade eder. İkinci hipotez ise güç durumuna karşılık gelir. Burada sadece 2

kategorili yordayıcı bağımlı değişkenle ilgili bilgi vermektedir. Elde edilen sonuçlara göre, randomForest paketi için üç ölçüt de daha fazla kesme noktasına sahip değişkeni tercih ederek güvenilir sonuçlar vermiştir. En fazla yanlılığın Gini öneminde görüldüğü ifade edilmiştir. cforest için ise permütasyon öneminin değiştirmesiz alt örnekleme seçeneğiyle kullanılırsa yanlılık oluşturmadağı görülmüştür. Dolayısıyla yanlılık kaynağı olarak, rastgele ormanların oluşturulduğu sınıflandırma ağaçlarının değişken seçim mekanizması ve bootstrap örnekleme mekanizması gösterilmiştir. Burada bootstrap örneklemenin değişkenler arasında yapay bir ilişkiye neden olarak yanlılık oluşturduğu raporlanmıştır. Sonuç olarak koşullu çıkarım ağaçlarının değiştirmesiz alt örnekleme seçeneğiyle kullanılması önerilmiştir. Çalışmada ek olarak rpart (sınıflandırma ağacına dayalı) ve ctree paketleriyle yapılan hesaplamalar da boş durum için seçim sıklığı bakımından incelenmiştir. Elde edilen sonuca göre rpart ile oluşturulan ağacın 20 kategorili değişkeni seçerek yüksek düzeyde yanlılık gösterdiği bulunmuştur.

Strobl ve arkadaşları (2008) rastgele orman algoritması için korelasyonlu yordayıcı değişkenlerin varlığında oluşan değişken seçim yanlılığına çözüm önerisinde bulunmayı amaçlamışlardır. Bu sebeple yeni bir koşullu permütasyon önemi hesaplaması önerilmiş ve cforest paketinde koşullu çıkarım ağaçlarına dayalı olarak gösterilmiştir. Çalışmada hem simülatif veri seti hem de gerçek veri seti üzerinde önerilen yöntemin performansı araştırılmıştır. Koşullu ve koşulsuz önem değerleri ayrı ayrı ele alınmıştır. Simülasyon çalışması üç farklı mtry (1,3,8) değerine göre tekrarlanmıştır. Sınıflandırma ağaçları temelinde gerçekleştirilen hesaplamalarda simülasyon çalışmasında 12 yordayıcı doğrusal bir modele göre veriler üretilmiştir. Bu yordayıcılar ilişkili, ilişkisiz, önemli ve önemsiz olacak şekilde elde edilmiştir. Öte yandan gerçek veri seti olarak peptit bağlama verilerinde amino asit dizisinin fenotiple ilişkisini incelemek amacıyla n=310 amino asit ve p=104 yordayıcı içeren bir veri seti kullanılmıştır. Elde edilen sonuçlara göre, önerilen koşullu yöntemin koşulsuz yöntemle göre daha güvenilir olduğu bulunmuştur. Koşulsuz önem ölçütünün ilişkili yordayıcıların önem puanlarını yapay olarak artırdığı görülmüştür.

Koşullu permütasyon önemi verideki gerçek örüntüyü daha iyi yansıtmıştır. Ayrıca koşullu önem farklı mtry değerlerinde, koşulsuz öneme göre daha düşük değişkenlik göstermiştir. Sonuç olarak ilişkili yordayıcı değişkenlerin önem değerlerinin yapay şekilde fazla kestirilmesinin rastgele ormanlarda kullanılan permütasyon önemi hesaplamasının doğasından yani, ağaçların ilk bölünmelerde ilişkili yordayıcıları seçmesinden kaynaklı olduğu vurgulanmıştır.

Nicodemus ve arkadaşları (2010) araştırmalarında Strobl ve arkadaşlarının (2008) çalışmasını genişletmişlerdir. Buna göre rastgele orman algoritmasında kullanılan koşulsuz ve koşullu permütasyon önemi ölçütlerinin korelasyonlu yordayıcılar olduğu durumda karşılaştırılmasını amaçlamışlardır. Yapılan simülasyon çalışmasında 12 yordayıcı içeren basit doğrusal bir model üretilmiştir. Model dört adet güçlü korelasyonlu ($r=0.90$) ve sekiz adet korelasyonsuz ($r=0$) yordayıcı içermektedir. Korelasyonlu ve korelasyonsuz yordayıcılardan üçer tanesi sıfırdan farklı katsayıya sahiptir. Önem değerleri her durumda farklı mtry değerlerine göre incelenmiştir. Sonuçta, yordayıcıların bağımlı değişkenle ilişkili olduğu ve kendi aralarında da korelasyon olduğu durumda koşulsuz değişken öneminin ilişkili yordayıcıların önem değerlerini daha yüksek kestirdiği görülmüştür. Bununla birlikte yordayıcıların bağımlı değişkenle ilişkili olmadığı durumda ise koşulsuz değişken öneminin yansız olduğu görülmüştür. Koşullu önem ölçütünün ise ilişkili yordayıcılar varlığında daha düşük önem değeri kestirerek yansız olduğu bulunmuştur. Öte yandan koşullu ve koşulsuz durumda da ölçeklemenin yanlılığa sebep olduğu ifade edilmiştir. Koşulsuz ölçeklenmemiş önem ölçütünün büyük veri setleri için hesaplama süresi açısından kolaylaştırıcı bir katkısı olduğu vurgulanmıştır. Buna bağlı olarak korelasyonlu yordayıcıların sorun olup olmadığının araştırmanın kapsamına bağlı olarak değerlendirilebileceği ifade edilmiştir. Bu bağlamda korelasyonlu değişkenler arasından etkili olanların belirlenmesini içeren küçük ölçekli çalışmalarda koşullu önemin, büyük ölçekli tarama çalışmalarında (çalışmanın kapsamına bağlı olarak) ise koşulsuz önemin de tercih edilebileceği raporlanmıştır.

Salis ve Kliem (2014) saldırganlık uygulayan ve uygulamayan bireylerin sınıflandırılmasını amaçlamışlardır. Bunun için 299 kadın ve 268 erkekten oluşan örneklem kullanılmıştır. Hesaplamalar koşullu çıkarım ağaçlarıyla gerçekleştirilmiştir. Çalışmada çiftlerin evliliklerinin başlangıcında ve 12 ay sonra fiziksel ve psikolojik saldırganlık öz bildirim ölçeğine verdiği yanıtlar kullanılmıştır. Sonuçta saldırganlık açısından erkekler iki (düşük-yüksek) kadınlar ise üç (hafif-orta-yüksek) gruba ayrılmışlardır. Elde edilen sonuçlar psikolojik açıdan yorumlanmıştır.

Valdiviezo ve Van Aelst (2015) tarafından yapılan araştırmada çeşitli eksik veri tamamlama yöntemlerinin performansının bir dizi ağaç tabanlı yöntemle göre incelemesi gerçekleştirilmiştir. Bunun için hem simülatif veri seti hem de 4 farklı gerçek veri seti kullanılmıştır. Simülatif veriler tamamen rastgele kayıp (MCAR), rastgele kayıp (MAR) ve rastgele olmayan kayıp (MNAR) mekanizmalarına göre üretilmiştir. İncelemeler hem sınıflandırma hem de regresyon ağaçlarında gerçekleştirilmiştir. Performanslar yanlış sınıflandırma oranı ve hata kareler ortalamasına göre kıyaslanmıştır. Ağaç tabanlı yöntemlerden sınıflandırma ve regresyon ağaçları, koşullu çıkarım ağaçları, rastgele orman, koşullu çıkarım ağaçlarına dayalı rastgele orman, torbalama ve koşullu çıkarım ağaçlarına dayalı torbalama kullanılmıştır. Kayıp veri tamamlama yöntemlerinden ise vekil bölünme, tekli imputasyon (medyan/mod atama, yakınlık matrisi, k-en yakın komşu), çoklu imputasyon (MICE (Zincirleme denklemlerle çok değişkenli imputasyon), MIST (Sıralı regresyon ağaçlarıyla çok değişkenli imputasyon) kullanılmıştır. Kayıp veri oranı yüzde 10,20,30,40 olarak ayarlanırken performans farklı örneklem büyüklükleri altında incelenmiştir. Elde edilen sonuçlara göre küçük miktarda kayıp veri varsa vekil bölünme veya tekli imputasyonla herhangi bir topluluk öğrenme yöntemini uygulamak yeterlidir. Orta veya büyük miktarda kayıp veri olduğunda MICE veya MIST ile koşullu çıkarım ağaçlarına dayalı rastgele orman güvenilirdir. Yüksek boyutlu veri setleri varsa vekil bölünmenin kullanıldığı koşullu çıkarım ağaçlarına dayalı torbalama kullanılabilir. Çalışmada ayrıca

çoklu imputasyon yöntemlerinin kullanıldığı topluluk öğrenme yöntemlerinin daha iyi performans gösterdiği de vurgulanmıştır.

Ferré ve arkadaşları (2019) çalışmalarında Chiari malformasyonu (anormal oluşum) görülen hastalarda uyku ile ilişkili solunum bozukluğu yüksek olanların tespit edilmesi amacıyla lojistik regresyon ve koşullu çıkarım ağacı yöntemleri karşılaştırılmıştır. Çeşitli klinik ve nöroradyolojik verilerin bağımsız değişken olarak kullanıldığı araştırmada her iki analizin de solunum bozukluğu tanısında faydalı olduğu görülmüştür. Buna ek olarak koşullu çıkarım ağaçlarının tıp alanındaki çalışmaların karmaşıklığı açısından uygulanması ve yorumlanmasının daha kolay olduğu da vurgulanmıştır.

Gholami ve arkadaşları (2020) araştırmalarında İran'ın güneydoğusundaki bir havzanın toz emisyonuna karşı duyarlılığını tespit etmek için farklı makine öğrenme algoritmalarını karşılaştırmışlardır. Çalışmada toprak, insan faaliyetleri, litoloji, meteoroloji alanlarından 14 farklı yordayıcı belirlenmiştir. Makine öğrenme algoritmalarından ise XGBoost, Cubist, BMARS, ANFIS, koşullu çıkarım ağaçlarına dayalı rastgele orman ve Elasticnet kullanılmıştır. Yöntemlerin performansları RMSE, beklenen ve gözlenen değerler arası korelasyon değeri, Nash Sutcliffe katsayısı ve MAE metriklerine göre karşılaştırılmıştır. Elde edilen sonuca göre tüm modeller oldukça iyi performans göstermekle birlikte bunların içinden koşullu çıkarım ağaçlarına dayalı rastgele orman algoritması en iyi sonuçları vermiştir.

Gomes ve arkadaşları (2020) çalışmalarında regresyon ağacı ve koşullu çıkarım ağacının performansını karşılaştırmışlardır. Bunun için Brezilya'da uygulanan 2011 ulusal lise eğitimi sınavı verileri kullanılmıştır. Bağımlı değişken olarak kompozisyon yazma performansının kullanıldığı araştırmada 53 farklı bağımsız değişken bulunmaktadır. 3.670.089 kişinin verilerinin analiz edildiği çalışmada algoritmalar açıklanan varyans değerlerine göre karşılaştırılmıştır. Elde edilen sonuçlara göre regresyon ağacının açıklanan varyans değeri %11.51, koşullu çıkarım ağacının ise %3.31 olarak bulunmuştur. Öte yandan algoritmalar kullandıkları değişkenlerin sayısı ve ölçek türleri bakımından da

farklılık göstermiştir. Araştırmacılar büyük veri setlerinde koşullu çıkarım ağaçlarının değişken seçim yanlılığından kaçınırken tahmini kötüleştirebileceğine ve regresyon ağaçlarının daha iyi sonuç verebileceğine dair yorumlarda bulunmuşlardır.

Schivinski (2021) araştırmasında tüketicilerin sosyal medya kullanımlarına ilişkin davranış kalıplarının belirlenmesini amaçlamıştır. Bunun için 1356 kişinin verileri kullanılmıştır. Tüketicilerin marka değeri algılarına ve sosyal medyada markalarla olan etkileşimlerine ilişkin ölçeklerle veriler toplanmıştır. Koşullu çıkarım ağaçlarının kullanıldığı araştırmada markalarla ilgili sosyal medya davranışını belirleyen beş farklı bilişsel örüntü ortaya çıkarılmıştır.

Cho ve arkadaşları (2022) çalışmalarında çeşitli biyofiziksel özellikleri kullanarak cilt yaşlanma düzeyini tahmin etmeyi amaçlamışlardır. Bunun için 20 ila 69 yaşları arasında sağlıklı 300 Koreli kadın katılımcıyla araştırma gerçekleştirilmiştir. Çalışmada doğrusal regresyon, koşullu çıkarım ağaçlarına dayalı rastgele orman, destek vektör makinesi, gauss süreçleri, koşullu çıkarım ağaçlarına dayalı gradyan artırma ve tekli regresyon ağacıyla analizler gerçekleştirilmiştir. Algoritmalar açıklanan varyans ve RMSE değerlerine göre kıyaslanmıştır. Elde edilen sonuca göre en iyi tahmin performansını koşullu çıkarım ağaçlarına dayalı rastgele orman algoritması göstermiştir.

Cavalli ve arkadaşları (2024) tarafından yapılan araştırmada disleksisi olan ve olmayanları daha iyi ayırt etmeyi sağlayacak standartlaştırılmış bir test prosedürü geliştirilmesi amaçlanmıştır. Fransa'daki bir üniversiteye yeni başlayan disleksisi olan 60 kişi ve olmayan 65 kişi ile çalışma yürütülmüştür. Katılımcılara yedi farklı test uygulanmıştır. Koşullu çıkarım ağaçlarıyla bağımlı değişkenle en ilgili testler, en uygun kesme puanları ve en uygun test sıralaması belirlenmeye çalışılmıştır. Sonuç olarak yedi yerine dört testle ve daha uygun kesme puanlarıyla verimli tarama yapılabilir standartlaştırılmış bir test prosedürü geliştirilmiştir.

Koşullu çıkarım ağaçlarına dayalı yöntemlerin farklı disiplinlerde kullanıldığı görülmektedir. Buna göre değişken önem ölçütleriyle ilgili yöntem geliştirme ve

karşılaştırma çalışmaları, eksik değer mekanizmaları ve ağaç tabanlı yöntemlerin birlikte incelenmesi kapsamında araştırmalar olduğu gibi psikoloji, eğitim, tıp, ekoloji, sosyal medya gibi alanlarda da bahsedilen algoritmaların uygulandığı görülmüştür. Öte yandan koşullu çıkarım ağaçlarının geleneksel regresyon yöntemleriyle, tekli ağaçlarla ve topluluk öğrenme yöntemleriyle karşılaştırmalı performanslarının incelendiği de söylenebilir. Bununla birlikte bu algoritmaların eğitim alanında kullanımlarının oldukça sınırlı olduğu da dikkat çekmiştir.



Bölüm 3

Yöntem

Bu bölümde araştırmanın türü, evren ve örneklem, verilerin elde edilmesi ve analizi ile ilgili açıklamalar yer almaktadır.

Araştırmanın Türü

Bu çalışmada PISA 2012 problem çözme testinde yer alan klima kontrolü ünitesi 1. maddesine ait günlük dosyasından elde edilen süreç verileri ve problem çözmeye açıklık, azim, matematik, fen bilimleri, okuduğunu anlama becerisine ilişkin puanları içeren sonuç verilerinin ve bu iki veri türünü birlikte içeren veri setinin sekiz farklı makine öğrenme algoritmasıyla (regresyon ağaçları, torbalama, rastgele orman, gradyan artırma, koşullu çıkarım ağaçları, koşullu çıkarım ağaçlarına dayalı torbalama, rastgele orman ve artırma) karmaşık problem çözme becerisini yordama düzeylerinin incelenmesi amaçlanmıştır. Çalışmanın türü, değişkenleri etkilemeye yönelik herhangi bir işlem yapılmadan bu değişkenler arasındaki ilişkilerin incelenmesi sebebiyle ilişkisel araştırmadır (Fraenkel ve ark., 2012, s. 331).

Araştırmanın Evreni ve Örneklemi

PISA 2012 döngüsüne Türkiye’den katılmış öğrenciler çalışmanın evrenini oluşturmaktadır. Örneklem ise bu döngüdeki KPÇ değerlendirmesine katılanları içermektedir. Çalışmada PISA 2012 KPÇ becerisine yönelik yordama yapılması hedeflendiği için bu testteki klima kontrolü ünitesinin 1. maddesini yanıtlayan bireyler analize dahil edilmiştir. Buna göre seçilen örneklemde 994 kişi bulunmaktadır. Süreç verisi değişkenlerinden biri olan müdahale sayısının elde edilmesi sırasında “uygula (apply)” tuşuna hiç tıklamamış bireyler analizden çıkarılmıştır. Çünkü sorunun çözümünde değişkenler arasındaki ilişkilerin keşfedilebilmesi için bu tuşa tıklanması gerekir. Dolayısıyla bu işlemi yapmayanlar çözüme dair anlamlı bir girişimde bulunmadıkları için son durumda 915 bireyden oluşan örneklem üzerinden analizler gerçekleştirilmiştir.

Verilerin Elde Edilmesi

Çalışmada kullanılan PISA 2012 veri setlerine ve günlük dosyalarına <https://www.oecd.org/pisa/pisaproducts/database-cbapisa2012.htm> adresinden

25.12.2022 tarihinde ulaşılmıştır.

Bağımlı Değişkenin Elde Edilmesi

Çalışmanın bağımlı değişkeni olan KPÇ becerisi bilgisayar tabanlı bir testle ölçülmüştür. PISA 2012 değerlendirmesinde bilgisayar ortamında yalnızca problem çözme değerlendirmesine katılan ülkeler için 31'den 38'e kadar sekiz farklı form bulunmaktadır. Problem çözme değerlendirmesinde 15 tanesi statik, 27 tanesi etkileşimli olan ve dört farklı problem çözme sürecini temsil eden (keşfetme ve anlama, temsil etme ve formüle etme, planlama ve uygulama, izleme ve yansıtma) 42 maddeyle problem çözme becerisi değerlendirilmiştir (OECD, 2014a, s. 29, 79). Dolayısıyla PISA 2012'de rapor edilen problem çözme puanı belirtilen iki problem kategorisinden elde edilen yanıtlara dayalı hesaplanmıştır. Ancak bu çalışmada KPÇ becerisine odaklanıldığı için, yalnızca karmaşık problemlerden elde edilen ham puanlardan yararlanarak, madde tepki kuramına dayalı bir yaklaşımla KPÇ puanı hesaplanmıştır. Buna göre hesaplamada kullanılan maddeler ve kodları Tablo 1'de verilmiştir.

Tablo 1

Bağımlı Değişkenin Hesaplanmasında Kullanılan Maddeler

Madde Kodu	Ünite Adı	Kategori Aralığı	Madde Kodu	Ünite Adı	Kategori Aralığı
CP014Q01	Lost	0-2	CP034Q01T	Clock	0-1
CP014Q02		0-1	CP034Q02		0-1
CP014Q06		0-1	CP034Q05		0-1
CP025Q01	Climate	0-2	CP036Q01	Lights	0-1
CP025Q02		0-2	CP036Q03		0-1
CP027Q01T	Aquarium	0-2	CP037Q01	Vitamins	0-1
CP027Q02		0-2	CP037Q02		0-1
CP028Q01	Advertising	0-1	CP037Q03	Tickets	0-1
CP028Q02		0-2	CP038Q01		0-2
	Remote Controls	0-2			0-1
CP029Q01			CP038Q02		
CP029Q02		0-2	CP038Q03		0-1
CP032Q01	Drink Machine	0-1	CP041Q02	Smiley	0-2

CP032Q02T	0-2	CP041Q03	0-2
CP032Q04	0-1		

Tablo 1’de verildiği üzere veri seti iki kategorili ve çok kategorili puanlanan maddelerden oluşmaktadır. Yetenek kestirimi yapılması için KPÇ maddelerini içeren tüm kitapçıklar (31’den 38’e kadar) dikkate alınarak 2022 kişinin verileri kullanılmıştır. Genelleştirilmiş kısmi puan modeli (GKPM) ve kısmi puan modeli (KPM) ile kestirimler R yazılımındaki (R Core Team, 2023) TAM (Robitzsch ve ark., 2024) paketinde beklenti maksimizasyonu (EM) algoritması (Bock ve Aitkin, 1981) ile tam bilgi maksimum olabilirlik yaklaşımı kullanılarak gerçekleştirilmiştir. Aşamalı tepki modeline (ATM) dayalı kestirimler için ise ltm (Rizopoulos, 2006) paketi kullanılmıştır. Tüm kestirimlerde öğrenci ağırlıkları dahil edilmiştir. Üç modele ilişkin model-veri uyumu sonuçları Tablo 2’de verilmiştir.

Tablo 2

Model-Veri Uyumu Sonuçları

	-2LL	AIC	BIC
GKPM	33751.30	33883.00	34253.00
KPM	34110.58	34191.00	34415.00
ATM	33790.96	33922.95	34293.33

Tablo 2 incelendiğinde en iyi uyumun GKPM’de elde edildiği görülmektedir. İki kategorili puanlanan maddelere göre elde edilen bir madde tepki modeli, kategori sayısının iki olduğu çok kategorili madde tepki kuramı modelinin özel bir durumu olarak görülebilir (Muraki, 1992). Çalışmada kullanılan ham puanlar hem iki hem de çok kategorili maddeleri içermektedir. Bu durumda GKPM’nin kestirim için uygun olduğu söylenebilir. Madde tepki kuramına dayalı modellerle kestirim yapılabilmesi için bazı varsayımların sağlanması gerekir. Buna göre tek boyutluluk varsayımının test edilmesi için GKPM’nin sırasıyla 1,2 ve 3 boyutlu durumları için kestirimler yapılarak model-veri uyumu sonuçları incelenmiştir. Elde edilen sonuçlar Tablo 3’te verilmektedir.

Tablo 3*GKPM için Model-Veri Uyumu Sonuçları*

	Boyut Sayısı	-2LL	AIC	BIC
GKPM	1	33751.30	33883.00	34253.00
	2	35076.14	35190.00	35509.00
	3	35919.40	36023.00	36315.00

Tablo 3'e göre en iyi uyum 1 boyutlu model için elde edilmiştir. Dolayısıyla modelin tek boyutluluk varsayımını karşıladığı görülmüştür. Yerel bağımsızlık incelemesi için tek boyutlu modele Yen's Q3 testi yapılmıştır. Artık korelasyon matrisinde 0.2 değerini aşan durumlar olmadığına yerel bağımsızlık sağlanmış olur (Chen ve Thissen, 1997). Çalışmada bu değerlerin -0.22 ile 0.13 arasında kaldığı ve yerel bağımsızlık varsayımının karşılandığı görülmüştür. GKPM'ye göre yapılan madde parametre kestirimlerine ilişkin değerler EK-C'de sunulmuştur. TAM paketinde yetenek kestirimleri ağırlıklı olabilirlik (WLE), maksimum olabilirlik (MLE) ve varsayılan seçenek olan beklenen sonsal (EAP) yaklaşımlarına göre yapılabilmektedir. Tablo 4'te üç farklı yaklaşımla elde edilen yetenek kestirimlerine ilişkin güvenilirlik değerleri verilmektedir.

Tablo 4*Yetenek Kestirimlerine İlişkin Güvenirlik Değerleri*

	Güvenirlik
EAP	0.747
MLE	0.727
WLE	0.711

Tablo 4'e göre en yüksek güvenilirlik EAP seçeneğinde oluşmuştur. Bu sebeple yetenek puanı olarak EAP kestirimine göre elde edilen değerler kullanılmıştır. Bu değerler -2.135 ile 2.638 arasında değişmektedir. Puan yorumlamalarının anlaşılabilirliğinin artırılması ve çalışmanın yaygın etkisi göz önüne alındığında, PISA 2012'de de kullanıldığı üzere, yetenek puanları ortalaması 500 standart sapması 100 olacak şekilde ölçeklenmiştir. Son olarak çalışmada dikkate alınan klima kontrolü ünitesinin bulunduğu CP2 kümesini içeren 31, 32, 35 ve 36 kodlu formları alan bireylerin yetenek puanları makine öğrenme algoritmalarına ilişkin analizlerde kullanılmıştır.

Süreç ve Sonuç Verisi Değişkenleri

Bu bölümde günlük dosyasından elde edilen süreç verisi değişkenleri ve sonuç verisi olarak kullanılan değişkenlerden bahsedilmektedir.

Süreç Verisi Değişkenleri

Çalışmada süreç verisine ilişkin değişkenlerin oluşturulmasında klima kontrolü ünitesinde yer alan 1. maddeye ait günlük dosyasından yararlanılmıştır. Bilgisayar ortamında gerçekleştirilen problem çözme değerlendirmesi, her biri 20 dakikada tamamlanabilecek dört küme halinde olmak üzere toplam 16 üniteden oluşmaktadır. Her öğrenciye, bilgisayar tabanlı okuduğunu anlama ve matematik testine katılıp katılamama durumuna göre bir veya iki küme verilmiştir. Bilgisayar ortamındaki testlerden yalnızca problem çözme değerlendirmesine katılan ülkelerde⁸ dört farklı küme (CP1-CP4) kullanılmıştır (OECD, 2014a, s. 150). Bu değerlendirme boyunca öğrencinin bilgisayarla olan her türlü etkileşimi (sürükle-bırak, tıklama vb.) günlük dosyalarında saklanmıştır. Bu dosyalarda eylemlerin sırası, etkileşim sıklığı, herhangi bir anda sistemin durumu, eylemlerin gerçekleştikleri ana ilişkin zaman damgası gibi bilgiler yer almaktadır (OECD, 2014a, s. 33).

Günlük dosyasından örnek bir kesit Şekil 6'da verilmiştir.

⁸ Türkiye, bilgisayar ortamındaki testlerden yalnızca problem çözme değerlendirmesine katılmıştır (OECD, 2014b, s. 221).

Şekil 6

Günlük Dosyasından Örnek Kesit

	cnt	schoolid	StIDStd	event	time	event_number	event_type	top_setting	central_setting	bottom_setting	temp_value	humid_value	diag_state
889394	TUR	0000060	01555	START_ITEM	1554,1000	1	NULL	NULL	NULL	NULL	NULL	NULL	NULL
889395	TUR	0000060	01555	ACER_EVENT	1589,5000	2	apply	1	0	0	27	25	NULL
889396	TUR	0000060	01555	ACER_EVENT	1593,1000	3	Diagram	NULL	NULL	NULL	NULL	NULL	'000000
889397	TUR	0000060	01555	ACER_EVENT	1593,6000	4	Diagram	NULL	NULL	NULL	NULL	NULL	'000000
889398	TUR	0000060	01555	ACER_EVENT	1594,6000	5	Diagram	NULL	NULL	NULL	NULL	NULL	'100000
889399	TUR	0000060	01555	ACER_EVENT	1597,5000	6	apply	1	-1	0	29	24	NULL
889400	TUR	0000060	01555	ACER_EVENT	1600,3000	7	Diagram	NULL	NULL	NULL	NULL	NULL	'100000
889401	TUR	0000060	01555	ACER_EVENT	1601,7000	8	Diagram	NULL	NULL	NULL	NULL	NULL	'100000
889402	TUR	0000060	01555	ACER_EVENT	1602,4000	9	Diagram	NULL	NULL	NULL	NULL	NULL	'100100
889403	TUR	0000060	01555	ACER_EVENT	1605,0000	10	apply	1	-1	-2	31	19	NULL
889404	TUR	0000060	01555	ACER_EVENT	1605,8000	11	apply	1	-1	-2	33	14	NULL
889405	TUR	0000060	01555	ACER_EVENT	1607,5000	12	apply	1	-1	-2	35	9	NULL
889406	TUR	0000060	01555	ACER_EVENT	1608,8000	13	reset	0	0	0	25	25	NULL
889407	TUR	0000060	01555	ACER_EVENT	1611,5000	14	apply	0	0	-2	25	21	NULL
889408	TUR	0000060	01555	ACER_EVENT	1615,8000	15	Diagram	NULL	NULL	NULL	NULL	NULL	'100100
889409	TUR	0000060	01555	ACER_EVENT	1612,5000	16	Diagram	NULL	NULL	NULL	NULL	NULL	'100101
889410	TUR	0000060	01555	END_ITEM	1618,5000	17	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Şekil 6'da verilen günlük dosyası örneği Türkiye'den katılmış 01555 numaralı öğrenciye aittir. Buna göre dosyadaki "cnt" sütunu ülke kodunu, "schoolid" sütunu okul kimlik kodunu, "StIDStd" sütunu öğrenci kimlik kodunu, "event" sütunu olayları ifade eder. Bu sütunda sistem tarafından oluşturulan ve eylemin başlama bitiş durumlarını bildiren START_ITEM, END_ITEM veya öğrenci tarafından oluşturulmuş ACER_EVENT etiketli olaylar kaydedilir. Eylemlere ilişkin zaman damgalarının yer aldığı "time" isimli sütunda ise değerlendirmenin başlangıcından itibaren saniye cinsinden ilgili olayın gerçekleştiği zaman yer almaktadır. Olayların sıra numarası ise "event_number" sütununda kayıtlıdır. Uygulama esnasında gerçekleştirilen olayların türü "event_type" sütununda bulunmaktadır. "apply", "reset" ve "diagram" olayları ACER_EVENT türünde kaydedilirken sistemin başlama ve bitiş anları NULL türünde kaydedilir. Öğrenci "apply" butonuna tıkladığında üst, orta ve alt kontrol çubuklarına bağlı olarak sıcaklık ve nem durumuna ilişkin sistemin aldığı genel durum kaydedilmektedir. Kontrol çubukları kaydırıldığında -2,-1,0,1,2 değerlerinden birini alırlar. Örneğin Şekil 6'daki 2 numaralı olay incelenecek olursa, öğrenci "apply" butonuna tıkladığında üst kontrol çubuğu 1, orta kontrol çubuğu 0, alt kontrol çubuğu 0 konumunda olup sıcaklık 27, nem 25 değerlerindedir. Bu aşamada öğrenci diyagram çizimi yapmadığı için "diag_state" olayı NULL olarak kaydedilmiştir. Öte yandan "reset" butonuna

tıklandığında değişkenlerin tümü başlangıç ayarlarına döner. Buna 13 numaralı olay örnek olarak verilebilir. Diyagram çizilmesi için ise öğrencinin iki defa tıklama yapması gerekir. İlk tıklama soldaki bir değişkenden çizim yapmaya başlamak, ikinci tıklama ise sağdaki bir değişkenle çizgiyi birleştirmek içindir. Eğer öğrenci çizgiyi çizmede başarısız olursa iki tıklama arasında diyagramın durumu değişmez. Öğrenci soldaki değişkene bir kez tıklar, ancak sağdaki değişkenle çizgiyi birleştirmeyerek olayı tamamlamazsa tek sayıda “diagram” olayı kaydedilir (OECD, t.y.). Soruda altı adet çizim yapma olasılığı vardır. Bunlar sırasıyla üst kontrol-sıcaklık, üst kontrol-nem, orta kontrol-sıcaklık, orta kontrol-nem, alt kontrol-sıcaklık, alt kontrol-nem olarak ifade edilebilir. Her kontrol çubuğunun durumu iki basamakla kodlanır; (0,0), (1,0), (0,1), (1,1). Bu kodlamalar sırasıyla kontrol çubuğunun “hiçbiri”, “sadece sıcaklık”, “sadece nem” ve “her ikisi” ile bağlantılı olduğunu temsil eder (Xu ve diğerleri, 2018). Şekil 6’daki 16 numaralı olay incelenecek olursa diag_state sütunundaki ilgili hücrede 100101 şeklinde bir kodlama yapıldığı görülmektedir. Buna göre öğrencinin üst kontrol-sıcaklık, orta kontrol-nem ve alt kontrol-nem arasında çizim yaparak soruyu doğru yanıtladığı görülmektedir.

Günlük dosyası ile ilgili verilen bilgilere göre çeşitli süreç verisi değişkenleri elde edilmiştir. İlgili değişkenler R yazılımında (R Core Team, 2024) araştırmacı tarafından yazılan kodlarla oluşturulmuştur. Bu değişkenlere ilişkin bilgiler Tablo 5’te yer almaktadır.

Tablo 5

Oluşturulan Süreç Verisi Değişkenleri

Görevde geçirilen toplam süre	Görevin başlamasından tamamlanmasına kadar geçen süreyi ifade eder. END_ITEM olayına ait zaman damgasından START_ITEM olayına ait zaman damgasının çıkarılmasıyla elde edilmiştir.
Strateji (VOTAT) puanı	Üç girdi değişkeninden hiçbiri için doğru strateji uygulanmamışsa 0, yalnızca bir tanesi için uygulanmışsa 1, iki tanesi için uygulanmışsa 2, üç tanesi için de uygulanmışsa 3 puan verilmiştir.
Göreve başlamadan önce geçirilen süre (Bekleme süresi)	Gerçekleştirilen ilk ACER_EVENT olayı ile görevin başladığı anı ifade eden START_ITEM olayı arasındaki farktır.
Eylem sayısı	ACER_EVENT olaylarının sayısıdır.
Sıfırlama	“reset” türündeki olayların sayısıdır.
olaylarının sayısı	

Müdahale sayısı	“apply” türündeki olayların sayısıdır.
Müdahalesiz gözlem puanı	Tüm kontrol çubuklarının değeri “0” iken “apply” türünde olay olup olmamasını ifade eder. 0-1 olarak kodlanmıştır.
Müdahalesiz gözlem sayısı	Tüm kontrol çubuklarının değeri “0” iken “apply” türündeki olayların sayısıdır.
Strateji (VOTAT) uygulama sayısı	VOTAT stratejisinin toplam uygulanma sayısıdır.
Her bir kontrol çubuğu için VOTAT sayısı	Her bir kontrol çubuğu (üst-orta-alt) için ayrı ayrı VOTAT stratejisinin kullanıldığı durumların sayısıdır.
İlk müdahaleye kadar geçen süre	İlk “apply” türündeki olayın gerçekleşme anı ile START_ITEM olayına ait zaman arasındaki farktır.
Diyagram olaylarının sayısı	“Diagram” türündeki olayların sayısıdır.
Son kontrol süresi	END_ITEM olayına ait zaman damgası ile son ACER_EVENT olayına ait zaman damgası arasındaki farktır.
Hız	Toplam eylem sayısının, görevde geçirilen toplam süreye oranlanmasıyla elde edilmiştir.

*Süre ile ilgili tüm değişkenler dakika cinsinden elde edilmiştir.

Sonuç Verisi Değişkenleri

Çalışmada sonuç verisi değişkenleri olarak PISA 2012 değerlendirmesindeki matematik, okuduğunu anlama, fen bilimleri puanlarına ait birinci makul değerler ile problem çözmeye açıklık ve azim ölçeklerinden elde edilen puanlar kullanılmıştır. Buna göre PISA 2012’de bireylerin değişen derecelerde sahip olduğu ifade edilen matematik okuryazarlığı değerlendirmesinde matematiksel bilgilerin yeniden üretilmesine ve bu bilgilerin çıkarım, açıklama, tahmin gibi yeni durumlarda uygulanabilmesine yönelik maddeler yer almaktadır (OECD, 2014c, s. 37). PISA 2009’da temel alan olan okuduğunu anlama becerisi *bireyin hedeflerine ulaşması, bilgi ve potansiyelini geliştirmesi ve topluma katılması için yazılı metinleri anlaması, kullanması, üzerinde düşünmesi ve bunlarla etkileşim kuması* olarak tanımlanmıştır (OECD, 2009, s. 23). PISA 2006’da temel alan olan fen bilimleri becerisi ise *bireyin bilim alanındaki bilgisi ve bu bilgiyi soruları tanımlamak, yeni bilgi edinmek, bilimsel olguları açıklamak ve bilimle ilgili konularda kanıta dayalı sonuçlar çıkarmak için kullanması* olarak tanımlanmıştır (OECD, 2007, s. 34). Kağıt tabanlı matematik, okuduğunu anlama ve fen bilimleri maddeleri 13 madde kümesi haline (7 matematik, 3 okuduğunu anlama, 3 fen bilimleri) 30 dakika sürecek şekilde oluşturulmuştur. Maddeler 13 kitapçık halinde ve her kitapçık 4 kümeden oluşacak şekilde iki saatlik test oturumlarında uygulanmıştır. Buna göre

PM1, PM2, PM3, PM4, PM5, PM6A ve PM7A kümeleri matematik; PR1, PR2 ve PR3 okuduğunu anlama; PS1, PS2, PS3 fen bilimleri kümelerini temsil etmektedir (OECD, 2014b, s. 30). Her bir değerlendirmede öğrenciler için 5 farklı makul değer elde edilmiştir (OECD, 2014b, s. 146). Çalışmada birinci makul değerler kullanılmıştır. Diğer sonuç verisi değişkenleri olan azim ve problem çözmeye açıklık öğrenci anketleriyle ölçülmüştür. Beşer madde ile ölçülen bu değişkenler, ifadelerin katılımcıya çok uyduğu ve hiç uymadığına yönelik etiketlenmiş beş kategorili ölçeklerdir. Buna göre ilgili anket soruları Tablo 6'da verilmiştir (OECD, 2014b, s. 337).

Tablo 6

Azim ve Problem Çözmeye Açıklık ile İlgili Anket Soruları

	<i>Madde Kodu</i>	<i>Q36: Aşağıdaki ifadelerin her biri sizi ne kadar iyi tanımlıyor?</i>
Azim	ST93Q01	Bir sorunla karşılaştığımda kolayca pes ederim.
	ST93Q03	Zor sorunları ertelerim.
	ST93Q04	Başladığım görevlerle ilgilenmeye devam ederim.
	ST93Q06	Her şey mükemmel olana kadar görevler üzerinde çalışmaya devam ederim.
	ST93Q07	Bir sorunla karşılaştığımda benden beklenenden fazlasını yaparım.
Problem çözmeye açıklık		<i>Q37: Aşağıdaki ifadelerin her biri sizi ne kadar iyi tanımlıyor?</i>
	ST94Q05	Çok fazla bilgiyle başa çıkabilirim
	ST94Q06	Bir şeyleri anlamakta hızlıyım.
	ST94Q09	Olayların açıklamalarını ararım.
	ST94Q10	Gerçekleri kolayca birbirine bağlayabilirim.
	ST94Q14	Karmaşık problemleri çözmeyi severim

Tablo 6'da verilen azim maddelerinden 04, 06 ve 07 numaralı olanlar (5=0), (4=1), (2=3), (1=4) olacak şekilde ters kodlanmıştır. 01 ve 03 numaralı olanlar ise (5=4), (4=3), (3=2), (2=1), (1=0) olarak kodlanmıştır. Problem çözmeye açıklık maddelerinin tümü (5=0), (4=1), (2=3), (1=4) olacak şekilde ters kodlanmıştır. Her iki beceri için toplam puan, tüm soruların toplamının geçerli yanıtların en yüksek puanına oranlanmasıyla hesaplanmıştır (OECD, 2014b, s. 425).

Verilerin Analizi

Çalışmada sadece süreç verisi değişkenlerini içeren, sadece sonuç verisi değişkenlerini içeren ve her iki türdeki değişkenleri içeren üç farklı veri seti için sekiz farklı makine öğrenme algoritmasıyla analizler gerçekleştirilmiştir. Günlük dosyasından özellik

elde edilmesi ve makine öğrenme algoritmalarıyla gerçekleştirilen analizler R yazılımındaki (R Core Team, 2024) Tablo 7’de verilen paketlerle yapılmıştır.

Tablo 7

Kullanılan Paketler

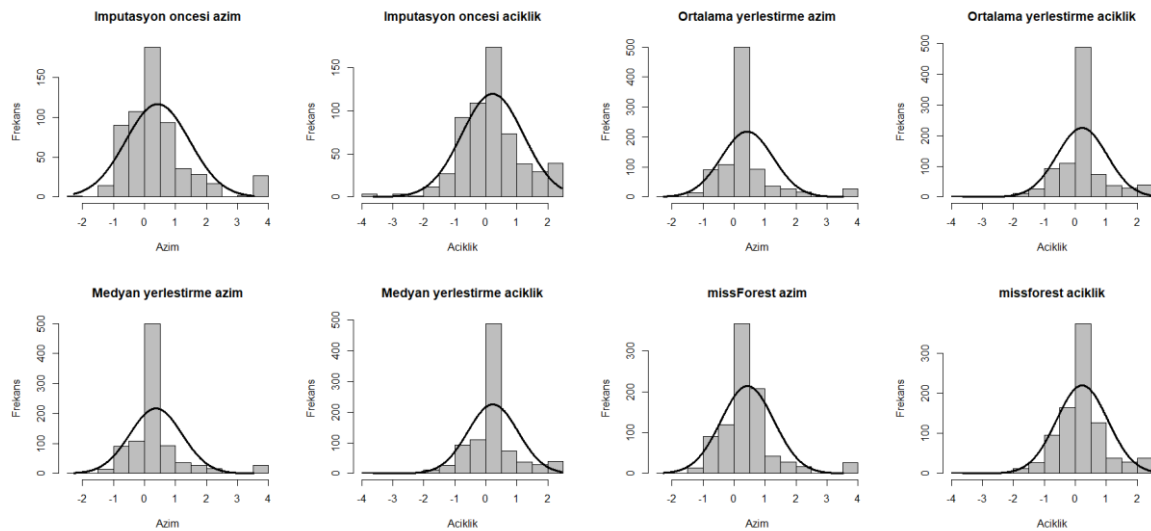
Paket Adı	Kullanım Amacı
haven (Wickham, Miller ve ark., 2023) stringr (Wickham, 2023) tidyr (Wickham ve ark., 2024) dplyr (Wickham, Francois ve ark., 2023) reshape2 (Wickham, 2007)	Günlük dosyasından süreç verisi değişkenlerinin elde edilmesinde kullanılmıştır.
psych (Revelle, 2024)	Betimsel istatistiklerin hesaplanmasında kullanılmışlardır.
GGally (Schloerke ve ark., 2024)	Korelasyon değerlerinin ve değişken dağılımlarının elde edilmesinde kullanılmıştır.
TAM (Robitzsch ve ark.,2024) lrm (Rizopoulos, 2006)	KPÇ puanının kestirilmesinde kullanılmıştır.
naniar (Tierney ve Cook, 2023)	Little’s MCAR testi için kullanılmıştır.
missForest (Stekhoven, 2022)	Eksik değer tamamlama işleminde kullanılmıştır.
MVN (Korkmaz ve ark., 2014)	Azim, problem çözmeye açıklık ve KPÇ değerlerine ilişkin çok değişkenli normallik incelemesinde kullanılmıştır.
Metrics (Hamner ve Frasco, 2018)	Metriklerin hesaplanmasında kullanılmıştır.
caret (Kuhn, 2008)	Makine öğrenme algoritmaları için en uygun hiper-parametre değerlerinin ve açıklanan varyansın hesaplanmasında kullanılmıştır.
rpart (Therneau ve Atkinson, 2023b) rpart.plot (Milborrow, 2024)	Regresyon ağacı analizleri için kullanılmıştır.
randomForest (Liaw ve Wiener, 2002)	Torbalama ve rastgele orman analizleri için kullanılmıştır.
partykit (Hothorn ve Zeileis, 2015)	Koşullu çıkarım ağacı, koşullu çıkarım ağaçlarına dayalı torbalama ve rastgele orman analizleri için kullanılmıştır.
gbm (Ridgeway ve ark.,2024) DALEX (Biecek, 2018)	Gradyan artırma analizinde kullanılmıştır. Koşullu çıkarım ağaçları ve koşullu çıkarım ağaçlarına dayalı gradyan artırma analizlerine ilişkin değişken önem grafiklerinin elde edilmesinde kullanılmıştır.
mboost (Hothorn ve ark., 2024)	Koşullu çıkarım ağaçlarına dayalı gradyan artırma analizinde kullanılmıştır.
tictoc (Izrailev, 2024)	Analiz sürelerinin hesaplanmasında kullanılmıştır.

Çalışmada azim değişkeninde 312 kişinin (%34), problem çözmeye açıklık değişkeninde 315 kişinin (%34) verileri eksiktir. İlgili değişkenler için Little’s MCAR testi yapılmış ve manidar sonuç elde edilmemiştir ($p=0.289$). Buna göre eksik değerlerin tamamen rastgele olduğu sonucuna varılmıştır. Azim, problem çözmeye açıklık ve KPÇ becerisine ilişkin değerler için Mardia’nın çok değişkenli çarpıklık ve basıklık testi sonucu

manidar sonuçlar elde edilmiş ve çok değişkenli normallik varsayımı karşılanmamıştır. Bu sebeple eksik veri tamamlamada parametrik teknikler kullanılmamıştır. Buna göre iki değişkene ilişkin eksik değerler ortalama yerleştirme, medyan yerleştirme ve Stekhoven ve Bühlmann (2012) tarafından geliştirilen parametrik olmayan bir teknik olan rastgele ormanlara dayalı yaklaşımla eksik değer kestirimi yapan missForest yöntemiyle doldurulmuştur. Bu yöntemlerden elde edilen veri setlerine göre tüm analizler tekrarlanmış ve değişken önemi sıralaması ya da metrik değerleri arasında belirgin farklar oluşmamıştır. Eksik veri tamamlama yöntemlerine göre elde edilen histogramlar Şekil 7’de verilmiştir.

Şekil 7

Eksik Veri Tamamlama İşlemlerine Göre Elde Edilen Histogramlar



Şekil 7’ye göre missForest yöntemiyle doldurulan verilerin çarpıklık ve basıklık anlamında en uygun olduğuna karar verilmiştir. Bu sebeple tüm analizler bu yöntemle doldurulmuş veri setleri ile gerçekleştirilmiştir.

Kurulan modellerin uygunluğunu değerlendirmek için Metrics (Hamner ve Frasco, 2018) paketiyle hesaplanan ortalama mutlak hata (Mean Average Error/MAE), ortalama mutlak hata yüzdesi (Mean Absolute Percentage Error/MAPE), yanlışlık, hata kareler ortalaması (Mean Squared Error/MSE), hata kareler ortalamasının karekökü (Root Mean Squared Error/RMSE), gerçek ve beklenen değerler arası korelasyon ve caret (Kuhn, 2008)

paketiyle hesaplanan açıklanan varyans değerleri kullanılmıştır. Metriklerle ilişkin açıklamalar aşağıda yer almaktadır. Her bir eşitlikte y_i gözlenen değeri, $\hat{y}_i$ beklenen değeri ifade etmektedir:

MAE (Ortalama Mutlak Hata): Beklenen ve gerçek değerler arasındaki ortalama mutlak farkları ifade eder. Bu değer ne kadar düşükse model veri uyumu o kadar yüksektir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

MAPE (Ortalama Mutlak Hata Yüzdesi): Tahmin edilen değerler ile gerçek değerler arasındaki yüzde farklarının mutlağının ortalamasını ifade eder. Bu değer ne kadar düşükse model veri uyumu o kadar yüksektir.

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Yanlılık: Tahmin edilen değerlerin ortalamasının gerçek değerlerden ne kadar uzaklaştığını ifade eder. Pozitif yanlılık modelin gerçek değerlerin üzerinde tahmin yaptığı, negatif yanlılık ise modelin gerçek değerlerin altında tahmin yaptığı anlamına gelir. Sıfıra yakın yanlılık ise modelde genel olarak bir yanlılık bulunmadığını ve tahminlerin ortalamada gerçek değerlere yakın olduğunu gösterir.

$$Yanlılık = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)$$

MSE (Hata Kareler Ortalaması): Gerçek ve beklenen değerler arasındaki farkın kareler toplamının ortalamasını temsil eder. Bu değer ne kadar düşükse model veri uyumu o kadar yüksektir.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

RMSE (Hata Kareler Ortalamasının Karekökü): MSE değerinin kareköküdür. Hatanın standart sapmasını ölçer. Bu değer ne kadar düşükse model veri uyumu o kadar yüksektir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

R (Beklenen ve gözlenen değerler arası korelasyon): Bu değer ne kadar yüksekse model veri uyumu o kadar yüksektir.

R^2 (Açıklanan varyans): Bu değer için kullanılan pakette iki hesaplama mevcuttur. Buna göre formula= “corr” seçeneği beklenen ve gözlenen değerler arasındaki korelasyonun karesini hesaplarken, çalışmada kullanılan “traditional” seçeneğinde aşağıdaki gibi hesaplama yapılmaktadır. Bu değer ne kadar yüksekse model veri uyumu o kadar yüksektir (Kuhn, 2023, s.33):

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Çalışmada kullanılan makine öğrenme algoritmalarından daha iyi performans gösterenler değerlendirilirken MAE, MAPE, MSE, RMSE için daha düşük, yanlışlık için sıfıra daha yakın, R ve R^2 için daha yüksek değerlere sahip olan modeller dikkate alınmıştır. Bu metriklerden daha fazlası için uygun değere sahip olan algoritmanın daha iyi performans gösterdiğine dair yorumlar yapılmıştır.

Etkili değişkenlerin yorumlanmasında değişkenlerin önemlerine göre sıralamaları dikkate alınmıştır. Strobl (2013, s. 697) tarafından ifade edildiği üzere, permütasyon önemi hesaplanırken farklı algoritmalarda etkili değişkenlerin sıralaması benzer olsa da önem puanlarının mutlak değerleri aynı olmayabilir. Çünkü bunlar ayarlanan hiper-parametrelerin değerlerine ve veri setinin özelliklerine bağlıdır. Bu nedenle önem puanlarının mutlak değerleri karşılaştırılmamalı veya yorumlanmamalıdır. Bunun yerine yalnızca değişkenlerin sıralaması dikkate alınmalıdır.

Denetimli öğrenme tekniklerinin çoğu rastgele örnekleme içerir (Sinharay, 2016). Dahası bunlardan rastgele orman analizlerinde rastgele örnekleme ek olarak yordayıcıların da rastgele seçimi söz konusudur (Strobl, 2013, s. 696). Ayrıca hiper-parametre ayarlama işleminde de rastgelelik vardır (Kuhn, 2019). Dolayısıyla bu rastgele süreçlerin yol açabileceği tekrarlanabilirlik sorununu çözerek aynı modelin farklı

hesaplamalarda aynı sonuçları üretmesi için kodlar set.seed() fonksiyonuyla çalıştırılmıştır (Strobl, 2013, s.696; Sinharay, 2016; Kuhn, 2019).

Regresyon Ağacı Analizleri

Regresyon ağacı analizlerinde en uygun eğitim/test verisi örneklem büyüklüğünü tespit etmek için %50-%50, %60-%40, %70-%30, %80-%20 oranlarında farklı veri setleriyle hesaplamalar gerçekleştirilmiştir. Elde edilen sonuçlara göre üç farklı veri setinde (süreç, sonuç, süreç+sonuç) en iyi sonuçları veren durumlar için, topluluk öğrenme algoritmaları ve koşullu çıkarım ağaçlarına dayalı algoritmalarla analizler gerçekleştirilmiştir. Bu sebeple Tablo 8’de dört farklı koşula göre regresyon ağacı analizi sonuçları verilmiştir.

Tablo 8

Farklı Örneklem Büyüklüğü Koşullarına Göre Elde Edilen Regresyon Ağaçlarına İlişkin Metrik Değerleri

	Eğitim-Test	MAE	MAPE	BIAS	MSE	RMSE	R	R ²
Süreç verisi	%50-%50	55.038	0.113	1.036	4625.817	68.013	0.630	0.394
	%60-%40	52.842	0.108	0.418	4208.566	64.873	0.685	0.464
	%70-%30	49.807	0.102	-1.147	3710.082	60.910	0.724	0.515
	%80-%20	48.317	0.099	-0.311	3487.953	59.059	0.701	0.491
Sonuç verisi	%50-%50	49.084	0.101	4.566	3717.664	60.973	0.720	0.513
	%60-%40	47.082	0.098	0.850	3410.129	58.396	0.752	0.566
	%70-%30	47.809	0.099	-1.142	3348.036	57.862	0.750	0.562
	%80-%20	48.440	0.099	1.914	3293.241	57.387	0.725	0.520
Süreç+ sonuç verisi	%50-%50	44.255	0.091	2.491	3172.439	56.324	0.770	0.584
	%60-%40	43.084	0.090	-2.604	2879.833	53.664	0.799	0.633
	%70-%30	42.268	0.087	-0.947	2621.323	51.199	0.811	0.657
	%80-%20	43.160	0.088	4.440	2759.967	52.535	0.781	0.597

Tablo 8 incelendiğinde üç veri seti için sırasıyla en iyi eğitim-test verisi büyüklüklerinin %80-%20, %60-%40, %70-%30 olduğu görülmüştür. Bundan sonraki makine öğrenme yöntemlerine dayalı analizler bu eğitim-test verilerinde gerçekleştirilmiştir. Çalışmada en uygun eğitim-test büyüklüğü koşulları için elde edilen regresyon ağaçları bulgular kısmında verilmiştir. Diğer koşullar için elde edilen regresyon ağaçları tüm veri setleri için EK-D, EK-E ve EK-F’de sunulmuştur.

Eğitim verisinden regresyon ağacı kurulurken herhangi bir ön budama işlemi yapılmayarak karmaşıklık parametresi (cp) değeri için 0 kullanılmıştır. Yani ilk aşamada maksimum boyutta ağaçlar elde edilmiştir. SRA’ların düzgün bir şekilde çalışması için

mümkün olan en büyük ağacı büyütmek ve ardından budama yapmak esastır. Öte yandan regresyon ağaçlarında optimum ağaç genellikle mümkün olan en büyük ağacın $1/10$ 'u kadardır ve bu nedenle bilgisayarın işlem süresinden tasarruf sağlamak için maksimum boyutta bir ağaç büyütmek gerekli değildir (Steinberg ve Colla, 1995, s. 176). Kullanılan rpart paketinde varsayılan cp değeri 0.01'dir. Bu değerle ağaçta bir ön budama yapılmış olur. Bu işlem başarılı olmakla birlikte, özellikle büyük veri kümelerinde bazen aşırı budama yapılmasına yol açabilir (Therneau ve Atkinson, 2023a, s. 24). Çalışmadaki veri seti çok büyük olmamakla birlikte, cp=0 değeriyle gerçekleştirilen hesaplamaların oldukça kısa sürmesi ve buna bağlı olarak olası aşırı budama riskinin azaltılması için budamadan önceki cp değeri 0 olarak kullanılmıştır. Daha sonra çapraz doğrulama hatasının minimuma düşüren cp değeri ile budamalar yapılmıştır. En uygun cp değeri, kullanılan pakette varsayılan olarak 10 kat çapraz doğrulamayla seçilir. Son olarak budanmış ağaç ile test verisinde kestirimler gerçekleştirilmiştir.

Topluluk Öğrenme Yöntemlerine Dayalı Analizler

Bu aşamada regresyon ağacı analizlerinden elde edilen en uygun eğitim/test verisi örneklem büyüklüklerine göre torbalama, rastgele orman ve gradyan artırma analizleri gerçekleştirilmiştir. Buna göre süreç verisi için %80-%20, sonuç verisi için %60-%40, süreç+sonuç verisi için %70-%30 eğitim/test örneklem büyüklüğü kullanılmıştır. Torbalama ve rastgele orman analizleri randomForest (Liaw ve Wiener, 2002) paketiyle, gradyan artırma ise gbm (Ridgeway ve ark., 2024) paketiyle gerçekleştirilmiştir. Modeller kurulmadan önce en uygun hiper-parametre değerleri caret (Kuhn, 2008) paketiyle araştırılmıştır. Makine öğrenme algoritmalarının performansı önemli ölçüde hiper-parametrelerin değerine bağlıdır ve farklı değerlere göre algoritma çok farklı davranabilir (Hutter ve ark., 2014, Yang ve Shami, 2020). Bunların en uygun değerleriyle analizin gerçekleştirilmesi makine öğrenme modelinin performansını büyük ölçüde artırır (Yang ve Shami, 2020). Tüm ayarlamalar 10 kat çapraz doğrulamayla yapılmıştır. Çalışmada ayarlamalar caret paketinin içeriğinde sunulanlarla sınırlıdır. Buna göre torbalama için

herhangi bir ayarlama yapılmamış ve varsayılan değerler kullanılmıştır. Rastgele orman algoritması için, her bölünmede dikkate alınacak rastgele değişken sayısını temsil eden “mtry” değeri zaman tasarrufu sağlaması açısından ızgara yöntemiyle (varsayılan) ayarlanmıştır. Bu yöntemde kullanıcı tarafından ayarlanacak parametrelere ilişkin olası değerleri içeren bir veri çerçevesi oluşturulur (Kuhn, 2019). Öte yandan randomForest paketindeki importance() fonksiyonunda varsayılan olarak değişken önem değeri, Breiman ve Cutler (2003) tarafından önerildiği şekliyle, ağaç sayısından elde edilen bir standart hataya göre standartlaştırılır. Ancak bu ölçekleme işlemiyle, testin gücünün örneklem büyüklüğü ile artmadığı, bunun yerine ağaç sayısına kıyasla geniş örneklem büyüklükleri için sıfır olduğu gösterilmiştir (Strobl ve Zeileis, 2008). Bu sebeple rastgele orman analizinde ilgili fonksiyonun scale argümanı FALSE ayarında da kullanılmıştır. Gradyan artırma için ise ızgara yöntemiyle ağaç sayısı (n.trees), her ağacın maksimum derinliği (interaction.depth), küçültme parametresi (shrinkage) ve ağaçların terminal düğümlerindeki minimum gözlem sayısı (n.minobsinnode) değerleri ayarlanmıştır. Ayarlamalar sonucu elde edilen en uygun değerler Tablo 9’da verilmiştir.

Tablo 9

Topluluk Öğrenme Algoritmaları İçin En Uygun Hiper-parametre Değerleri

		Ayarlama yapılmadı	mtry	n.trees	interaction.depth	shrinkage	n.minobsinnode
Süreç Verisi	Torbalama	X					
	Rastgele orman		2				
	Gradyan Artırma			25	4	0.1	5
Sonuç Verisi	Torbalama	X					
	Rastgele orman		3				
	Gradyan Artırma			250	4	0.01	10
Süreç+ Sonuç Verisi	Torbalama	X					
	Rastgele orman		5				
	Gradyan Artırma			50	2	0.1	5

Koşullu Çıkarım Ağaçlarına Dayalı Analizler

Bu aşamada partykit (Hothorn ve Zeileis, 2015; Hothorn ve ark., 2006) paketindeki ctree() fonksiyonuyla koşullu çıkarım ağacı, cforest() fonksiyonuyla koşullu çıkarım ağaçlarına dayalı torbalama ve rastgele orman analizleri yapılmıştır. ctree() fonksiyonu için “mincriterion” argümanının en uygun değeri rastgele arama yöntemiyle hesaplanmıştır. Bu değer sürekli bir ölçekte olduğu için ızgara arama yöntemi yerine rastgele arama yöntemi kullanılmıştır. Bu yöntemde olası ayar parametresi kombinasyonları rastgele bir şekilde araştırılır (Kuhn, 2019). Koşullu çıkarım ağaçlarına dayalı torbalama analizi için ayarlama yapılmamıştır. Koşullu çıkarım ağaçlarına dayalı rastgele orman için ise ızgara yöntemiyle “mtry” değeri ayarlanmıştır. cforest() fonksiyonunda Strobl ve arkadaşları (2007) tarafından önerilen değiştirmesiz alt örnekleme seçeneği varsayılan değerdir. Öte yandan değişken öneminin hesaplanmasında kullanılan varimp() fonksiyonu hem koşulsuz (randomForest paketinde hesaplanan) hem de koşullu permütasyon önemini hesaplar. Dolayısıyla bu fonksiyon conditional=TRUE ayarında da kullanılmıştır. Son olarak, koşullu çıkarım ağaçlarına dayalı gradyan artırma algoritması ile analizler mboost (Hothorn ve ark., 2024) paketindeki blackboost() fonksiyonuyla yapılmıştır. Bu analiz için keyfi kayıp fonksiyonu varsayılan değeri olan “Gaussian” seçeneğinde bırakılırken, iterasyon sayısını belirten “mstop” değeri ve ağacın maksimum derinliğini ifade eden “maxdepth” değeri ızgara yöntemiyle ayarlanmıştır. Ayarlamalar sonucu elde edilen en uygun değerler Tablo 10’da verilmiştir.

Tablo 10

Koşullu Çıkarım Ağaçlarına Dayalı Algoritmalar İçin En Uygun Hiper-parametre Değerleri

		Ayarlama yapılmadı	mincriterion	mtry	mstop	maxdepth
Süreç Verisi	Koşullu çıkarım ağacı		0.8830174			
	K- torbalama	X				
	K- rastgele orman			10		
	K- gradyan artırma				25	3
Sonuç Verisi	Koşullu çıkarım ağacı		0.8830174			
	K- torbalama	X				
	K- rastgele orman			3		
	K- gradyan artırma				250	4
	Koşullu çıkarım ağacı		0.8830174			

Süreç+	K- torbalama	X			
Sonuç	K- rastgele orman		8		
Verisi	K- gradyan artırma			150	5



Bölüm 4

Bulgular, Yorumlar ve Tartışma

Bu bölümde araştırmanın alt problemlerine ait bulgulara yer verilmiştir. Elde edilen bulgular yorumlanmış ve alan yazındaki diğer çalışmalar da dikkate alınarak tartışılmıştır.

1. Karmaşık problem çözme becerisine ilişkin elde edilen süreç verilerinin, tekli regresyon ağacı, torbalama, rastgele orman, gradyan artırma, koşullu çıkarım ağaçları, koşullu çıkarım ağaçlarına dayalı torbalama, rastgele orman, gradyan artırma algoritmaları kullanılarak ilgili beceriyi yordama performansı nedir ve hangi yöntem daha iyi performans sergilemiştir?

Tablo 11

Süreç Verisi-Makine Öğrenme Algoritmaları Metrik Değerleri

	MAE	MAPE	BIAS	MSE	RMSE	R	R ²
Regresyon ağacı	48.317	0.099	-0.311	3487.953	59.059	0.701	0.491
Torbalama	49.837	0.102	0.303	3777.750	61.463	0.673	0.449
Rastgele Orman	48.477	0.099	0.513	3503.500	59.190	0.699	0.489
Gradyan Artırma	48.390	0.100	1.266	3438.504	58.639	0.710	0.498
Koşullu çıkarım ağacı	47.865	0.098	1.761	3432.953	58.591	0.707	0.499
K-torbalama	48.136	0.099	1.756	3420.574	58.486	0.708	0.501
K-rastgele orman	48.288	0.099	2.377	3428.096	58.550	0.708	0.500
K-gradyan artırma	48.455	0.100	1.661	3420.492	58.485	0.719	0.501

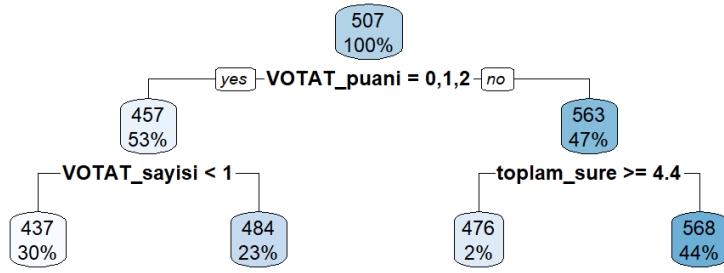
Tablo 11 incelendiğinde süreç verisi için tüm yöntemlerin orta düzeyde tahmin performansı sergilediği söylenebilir. Buna ek olarak, metrik değerleri arasında marjinal farklılıklar olmamakla birlikte, daha iyi tahmin performansının koşullu çıkarım ağaçlarına dayalı gradyan artırma algoritmasında görüldüğü söylenebilir.

Çalışmada kullanılan topluluk öğrenme yöntemleri ve bunların koşullu versiyonları için bir ağaç şeması elde edilememektedir. Yalnızca regresyon ağaçları ve bunların koşullu

karşılığı için birer ağaç şeması elde edilebilir. Süreç verisinden elde edilen bu ağaç şemaları Şekil 8 ve Şekil 9'da verilmiştir.

Şekil 8

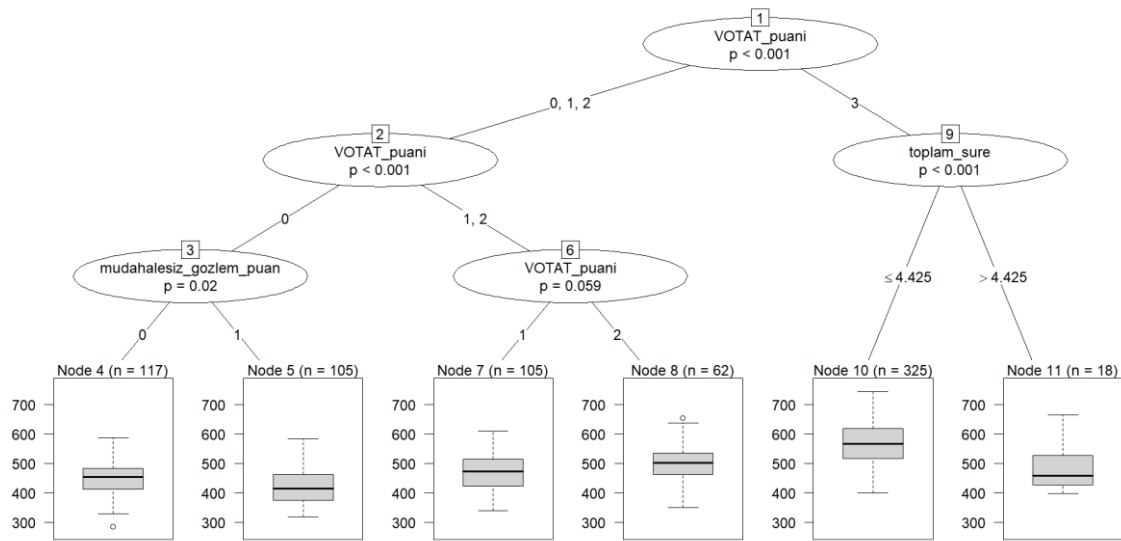
Süreç Verisinden Elde Edilen Regresyon Ağacı



Şekil 8'e göre ilk bölünmenin VOTAT puanına göre gerçekleştiği görülmektedir. Şekildeki mavi kutucukların içinde üst sırada yazılı olan değer, bu bölünmedeki bireylerin KPÇ puanı ortalamasıdır. Altta yazan değer ise ilgili bölünmede örneklemin ne kadarının kullanıldığını belirtir. Buna göre örneğin en soldaki bölünme incelenecek olursa, VOTAT puanı 0,1,2 olanların ve VOTAT uygulama sayısı 1'den küçük olanların KPÇ puan ortalaması 437 olup, bu bireyler grubun %30'unu oluşturmaktadır. En sağdaki bölünmeye göre ise VOTAT puanı 3 olanların ve toplam süresi 4.4 dakikadan kısa olanların KPÇ puan ortalaması 568 olup, bu bireyler grubun %44'ünü oluşturmaktadır. Dolayısıyla süreç verisine göre KPÇ'den daha yüksek puan almak için VOTAT puanının yüksek, toplam sürenin daha kısa olması gerektiği söylenebilir.

Şekil 9

Süreç Verisinden Elde Edilen Koşullu Çıkarım Ağacı



Şekil 9'da verilen koşullu çıkarım ağacında, tekli regresyon ağaçlarından farklı olarak her bölünme için anlamlılık değeri raporlandığı görülmektedir. Buna göre ilk bölünme (tekli regresyon ağacıyla benzer şekilde) VOTAT puanına göre gerçekleşmiştir. Örneğin en soldaki bölünmeler yorumlanırsa, VOTAT puanı 0,1 veya 2 olan ve müdahalesiz gözlem puanı 0 olanların KPÇ puanının 300-600 arasında kaldığı görülmektedir. Bu bireyler 117 kişiden oluşmaktadır. En sağdaki bölünmeye göre ise, VOTAT puanı 3 olanların ve toplam süresi 4.425 dakikadan uzun olanların yaklaşık 400-700 puan aralığında değerlere sahip oldukları görülmektedir. Bu bireyler 18 kişiden oluşmaktadır. Bu bölünmenin sonucunun hem bölünmede seçilen değişkenler hem de kesme noktaları bakımından tekli regresyon ağacıyla, en sağ ve soldaki bölünmeler bağlamında büyük ölçüde uyumlu olduğu dikkat çekmektedir. Bununla birlikte her iki ağaç kullandıkları bazı değişkenler bakımından belli ölçüde farklılıklara da sahiptir.

2. Çalışmada ele alınan sonuç verilerinin, tekli regresyon ağacı, torbalama, rastgele orman, gradyan artırma, koşullu çıkarım ağaçları, koşullu çıkarım ağaçlarına dayalı torbalama, rastgele orman, gradyan artırma algoritmaları kullanılarak ilgili beceriyi yordama performansı nedir ve hangi yöntem daha iyi performans sergilemiştir?

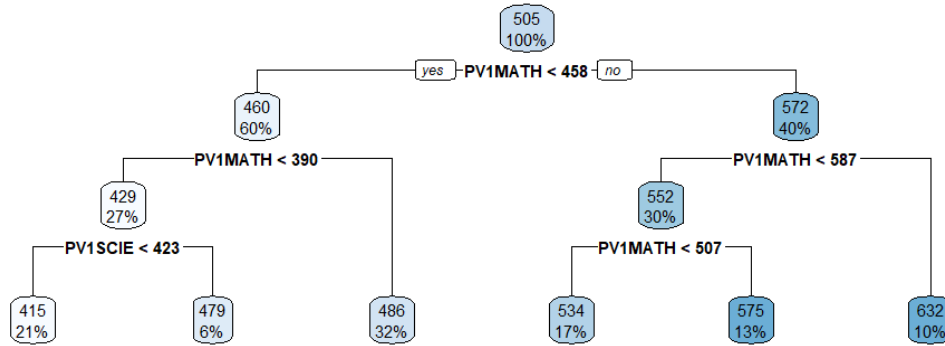
Tablo 12*Sonuç Verisi-Makine Öğrenme Algoritmaları Metrik Değerleri*

	MAE	MAPE	BIAS	MSE	RMSE	R	R ²
Regresyon ağacı	47.082	0.098	0.850	3410.129	58.396	0.752	0.566
Torbalama	46.044	0.096	-1.843	3250.181	57.010	0.766	0.586
Rastgele Orman	45.673	0.095	-1.445	3194.820	56.523	0.770	0.593
Gradyan Artırma	45.422	0.094	0.774	3142.674	56.060	0.778	0.600
Koşullu çıkarım ağacı	46.785	0.097	0.763	3286.262	57.326	0.763	0.581
K-torbalama	45.859	0.095	-0.478	3192.481	56.502	0.770	0.593
K-rastgele orman	45.526	0.095	-0.619	3152.678	56.149	0.774	0.598
K-gradyan artırma	45.650	0.095	0.294	3137.749	56.016	0.778	0.600

Tablo 12 incelendiğinde sonuç verisi için tüm yöntemlerin orta-iyi düzeyde tahmin performansı sergilediği söylenebilir. Buna ek olarak, metrik değerleri arasında marjinal farklılıklar yoktur. Bununla birlikte nispeten koşullu çıkarım ağaçlarına dayalı gradyan artırma algoritmasının daha iyi performans gösterdiği söylenebilir. Ancak bu algoritmanın metrik değerlerinin bazılarının gradyan artırmanın metrik değerleriyle aynı olduğu da göze çarpmaktadır. Bu durumda iki algoritma arasında net bir ayırım yapmanın mümkün olmadığı düşünülmektedir.

Şekil 10

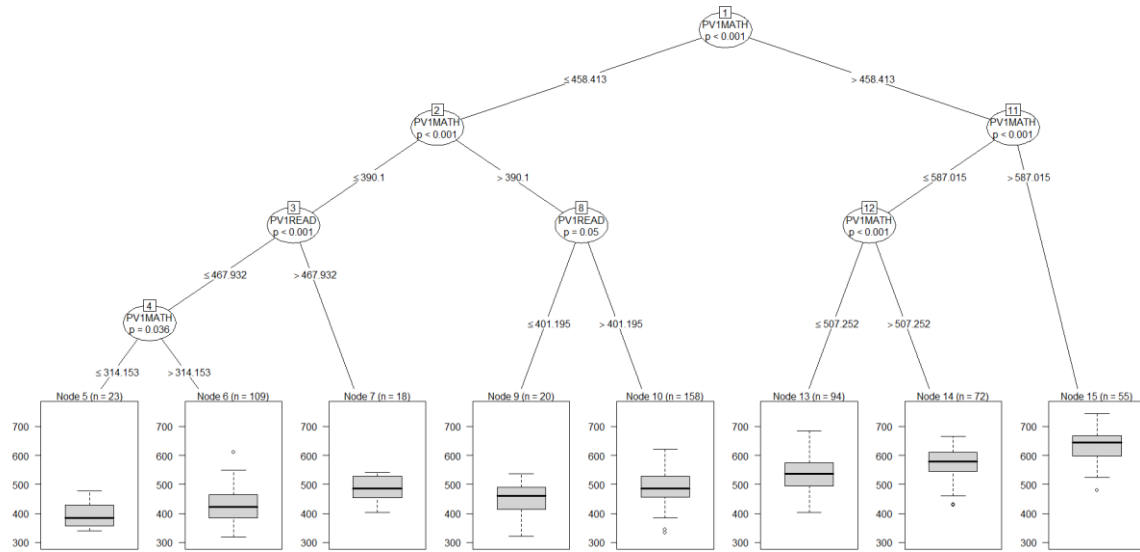
Sonuç Verisinden Elde Edilen Regresyon Ağacı



Şekil 10'a göre ilk bölünmenin matematik puanına göre gerçekleştiği görülmektedir. Regresyon ağaçlarında safsızlığı en çok azaltan değişkenlerin seçilmesine yönelik bir yaklaşım söz konusu olduğu için aynı değişkenin ağacın farklı bölünmelerinde görülmesi doğaldır. Buna göre örneğin en soldaki bölünme incelenecek olursa, matematik puanı sırasıyla 458'den ve 390'dan küçük olan, fen puanı 423'ten küçük olanların KPÇ puan ortalaması 415'tir. Bu bireyler grubun %21'ini oluşturmaktadır. Öte yandan en sağdaki bölünmeye göre matematik puanı 458'den ve 587'den büyük ve eşit olanların KPÇ puan ortalaması 632 olup, bu bireyler grubun %10'unu oluşturmaktadır. Dolayısıyla sonuç verisine göre KPÇ'den daha yüksek puan almak için matematik başarısının baskın şekilde etkili olduğu görülmektedir.

Şekil 11

Sonuç Verisinden Elde Edilen Koşullu Çıkarım Ağacı



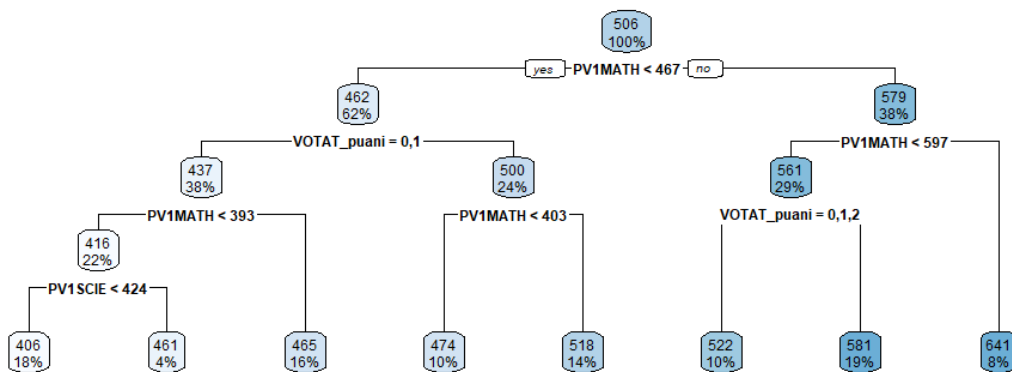
Şekil 11’de verilen koşullu çıkarım ağacına göre, ilk bölünmenin (tekli regresyon ağacıyla benzer şekilde) matematik puanına göre gerçekleştiği söylenebilir. Örneğin en soldaki bölünmeler yorumlanırsa, matematik puanı 458.413’ten ve 390.1’den küçük ve eşit olan, okuduğunu anlama puanı 467.932’den küçük ve eşit olan, matematik puanı 314.153’ten küçük ve eşit olanların KPÇ puanının 300-500 arasında kaldığı görülmektedir. Bu bireyler 23 kişiden oluşmaktadır. En sağdaki bölünmeye göre ise, matematik puanının 458.413 ve 587.015’ten büyük olanların 500-800 puan aralığında değerlere sahip oldukları görülmektedir. Bu bireyler 55 kişidir. Aynı zamanda tekli regresyon ağacı ve koşullu çıkarım ağaçlarında kesme noktalarının, en sağ ve soldaki bölünmeler bakımından, büyük ölçüde örtüştüğü söylenmelidir. Bununla birlikte her iki ağaç kullandıkları bazı değişkenler bakımından belli ölçüde farklılıklara da sahiptir.

3. Süreç ve sonuç verilerinin birlikte ele alındığı durumda oluşan veri setindeki değişkenlerin, tekli regresyon ağacı, torbalama, rastgele orman, gradyan artırma, koşullu çıkarım ağaçları, koşullu çıkarım ağaçlarına dayalı torbalama, rastgele orman, gradyan artırma algoritmaları kullanılarak ilgili beceriyi yordama performansı nedir ve hangi yöntem daha iyi performans sergilemiştir?

Tablo 13*Süreç+Sonuç Verisi-Makine Öğrenme Algoritmaları Metrik Değerleri*

	MAE	MAPE	BIAS	MSE	RMSE	R	R ²
Regresyon ağacı	42.268	0.087	-0.947	2621.323	51.199	0.811	0.657
Torbalama	39.512	0.082	-1.110	2362.952	48.610	0.832	0.691
Rastgele Orman	38.470	0.080	-0.667	2251.254	47.447	0.840	0.706
Gradyan Artırma	38.717	0.080	-1.835	2225.891	47.179	0.842	0.709
Koşullu çıkarım ağacı	41.779	0.086	-1.747	2639.376	51.375	0.811	0.655
K-torbalama	39.393	0.081	-0.589	2299.317	47.951	0.836	0.699
K-rastgele orman	39.191	0.081	-0.941	2276.171	47.709	0.838	0.702
K-gradyan artırma	40.352	0.083	-0.947	2344.930	48.424	0.834	0.693

Tablo 13 incelendiğinde süreç ve sonuç verisi için tüm yöntemlerin iyi düzeyde tahmin performansı sergilediği söylenebilir. Buna ek olarak, metrik değerleri arasında marjinal farklılıklar olmamakla birlikte, daha iyi tahmin performansı gradyan artırma algoritmasından elde edilmiştir.

Şekil 12*Süreç+Sonuç Verisinden Elde Edilen Regresyon Ağacı*

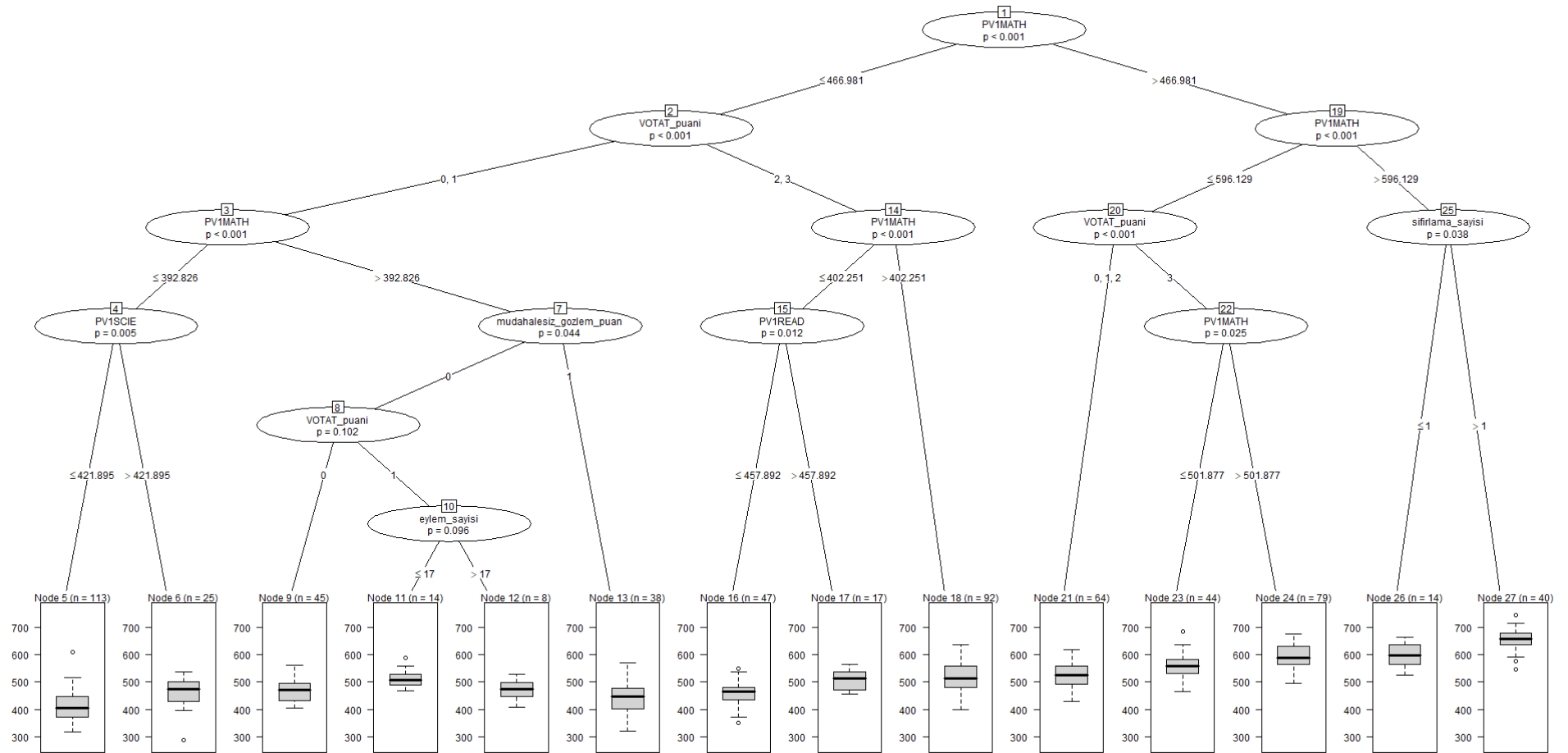
Şekil 12'ye göre ilk bölünmenin sonuç verisiyle benzer şekilde matematik puanına göre gerçekleştiği görülmektedir. Buna göre örneğin en soldaki bölünme yorumlanacak

olursa, matematik puanı 467'den küçük, VOTAT puanı 0 veya 1 olan, matematik puanı 393'ten küçük ve fen puanı 424'ten küçük olanların KPÇ puan ortalaması 406'dır ve bu bireyler grubun %18'ini oluşturmaktadır. En sağdaki bölünmeye göre ise, matematik puanı 467'den ve 597'den büyük ve eşit olanların bu grup için en yüksek KPÇ ortalaması olan 641 puana sahip olma eğiliminde oldukları ve bu bireylerin grubun %8'ini oluşturduğu söylenebilir. Süreç ve sonuç verilerinin birlikte ele alındığı durum için bölünmede en baskın değişkenin matematik puanı olduğu dikkat çekmektedir. Süreç verisi değişkenlerinden ise VOTAT puanının bölünmelerde etkili olduğu ifade edilebilir.



Şekil 13

Süreç+Sonuç Verisinden Elde Edilen Koşullu Çıkarım Ağacı



Şekil 13'te verilen koşullu çıkarım ağacına göre, ilk bölünmenin (tekli regresyon ağacıyla benzer şekilde) matematik puanına göre gerçekleştiği söylenebilir. Örneğin en soldaki bölünmeler yorumlanırsa, matematik puanı 466.981'den küçük ve eşit, VOTAT puanı 0 veya 1 olan, matematik puanı 392.826'dan küçük ve eşit olan, fen puanı 421.895'ten küçük ve eşit olanların KPÇ puanının 300-500 arasında kaldığı görülmektedir. Bu bireyler 113 kişiden oluşmaktadır. En sağdaki bölünmeye göre ise, matematik puanının 466.981 ve 596.129'dan büyük, sıfırlama sayısı 1'den büyük olanların 600-700 puan aralığında değerlere sahip oldukları görülmektedir. Bu bireyler 40 kişiden oluşmaktadır. Verilen ağaçtan elde edilen sonuçlar, en sağ ve soldaki bölünmeler bakımından, tekli regresyon ağacıyla bölünmede seçilen değişkenler ve kesme noktaları bağlamında büyük ölçüde uyum göstermektedir. Bununla birlikte koşullu çıkarım ağacında bölünme için tekli regresyon ağacında olmayan bazı değişkenler de kullanılmıştır.

İlk üç araştırma problemine ilişkin sonuçlar birlikte değerlendirildiğinde süreç verisinin orta, sonuç verisinin orta-iyi, süreç+sonuç verisinin iyi düzeyde tahmin performansı sergilediği ifade edilebilir. Bu noktada tüm veri setlerinin kabul edilebilir sonuçlar verdiği görülmüştür. KPÇ bağlamında yalnızca süreç verilerinin dikkate alındığı ve makine öğrenme algoritmaları bağlamında yapılan çalışmalarda süreç verilerinin iyi düzeyde tahmin performansına sahip olduğu raporlanmıştır (Pejić ve Molcer, 2012; Qiao ve Jiao, 2018). Bununla birlikte yine KPÇ bağlamında süreç ve sonuç verilerinin bir arada incelendiği Zhan ve Qiao (2022) tarafından yapılan bir başka çalışmada süreç verileriyle yapılan kestirimin sadece sonuç verisine dayalı bir kestirime göre bireyin yetenek düzeyini daha iyi yansıttığı ifade edilmiştir. Öte yandan süreç verilerinin farklı bağlamlarda kullanıldığı çalışmalarda da bu veri türünün ölçme-değerlendirme uygulamalarına katkısı vurgulanmaktadır. Buna göre; süreç verileriyle yapılan yetenek kestiriminin, sadece sonuç verilerinden yararlanılarak yapılan kestirime göre bireyin yetenek düzeyini daha iyi yansıttığına, yetenek tahmininin kesinliğini, güvenilirliğini ve performansını artırdığına dair çeşitli bulgular söz konusudur (Reis Costa ve ark., 2021; Scalise ve Clarke Midura, 2018; Tang ve ark., 2020; Zhang ve

ark., 2023). Bununla birlikte yazma değerlendirmesi kapsamında süreç ve sonuç verilerini birlikte ele alan bazı çalışmalarda ise, sonuç verilerinin daha yüksek regresyon katsayısına sahip olduğu veya bunların süreç verilerine göre daha iyi tahmin performans sergilediği gibi bulgular elde edilmiştir (Deane, 2014; Sinharay ve ark., 2019; Zhang ve Deane, 2015). Ancak bu durum yorumlanırken ilgili çalışmalarda kullanılan bağımlı değişken puanlarının süreç özelliklerinden ziyade sonuç özelliklerine göre oluşturulduğunun da göz önünde bulundurulması gerektiğine dikkat çekilmiştir (Zhang ve Deane, 2015). Ayrıca Deane (2014) çalışmasında süreç verilerinin kullanımının avantajına ve bu verilerin yalnızca sonuç özelliklerinin kullanıldığı durumda doğrudan yakalanamayacak yönleri ortaya çıkardığına vurgu yapmıştır. Aynı zamanda araştırmacı süreç ve sonuç özelliklerinin birlikte kullanımının, incelenen örtük özellikteki performansı daha iyi ortaya koyabileceğini ifade etmiştir. He ve arkadaşları da (2019) süreç verilerinin özellikle bireyler soruya yanlış yanıt verdiklerinde ek müdahalelerin belirlenmesinde bilgilendirici olabileceğini vurgulamışlardır. Bununla birlikte Zhang ve arkadaşları (2023) araştırmalarının sonucunda, süreç verilerinin yetenek kestirimine dahil edilmesiyle daha kısa testlerde de güvenilir kestirimler yapılabileceğini belirtmişlerdir. Böylelikle, süreç verilerinin değerlendirmeye dahil edilmesinin test geliştirme süreçlerine katkı sağlayabileceği söylenebilir. Sonuç olarak her türdeki verinin sağladığı avantajların ve sınırlılıkların olduğu da unutulmamalıdır. Dolayısıyla kullanılacak veri türüne karar verilirken araştırmacının amacı, testin uygulama koşulları ve biçimi gibi özelliklerin göz önünde bulundurularak tercih yapılması önerilmektedir.

Yine ilk üç alt problemde yapılan yöntem karşılaştırmasına göre (tüm yöntemler birlikte ele alındığında) süreç verisinde koşullu çıkarım ağaçlarına dayalı gradyan artırma, sonuç verisinde gradyan artırma ve koşullu çıkarım ağaçlarına dayalı gradyan artırma, süreç+sonuç verisinde gradyan artırma daha iyi performans göstermiştir. Burada tüm yöntemler için metrik değerleri arasında marjinal farklılıklar bulunmadığı vurgulanmalıdır. Bu durum yapılan çeşitli araştırmalarda da doğrulanmıştır. Örneğin Bühlmann ve Hothorn

(2007) araştırmasında koşullu çıkarım ağaçlarına dayalı gradyan artırma ve gradyan artırma yöntemlerinin tahmin edilen değerler bağlamında önemli ölçüde örtüşen sonuçlar verdiğini raporlamıştır.

İlk üç araştırma probleminden elde edilen bulgulara bağlı olarak yöntemler arası performans kıyaslaması da yapılmıştır. Buna göre elde edilen sonuçlar aşağıda sıralanmıştır:

a) Tekli regresyon ağacı ile topluluk öğrenme yöntemlerinin (torbalama, rastgele orman, gradyan artırma) çalışmada kullanılan veri setlerindeki (süreç/sonuç/süreç+sonuç) yordama performansları incelendiğinde daha iyi sonuç veren algoritma hangisidir?

Tablo 11, 12 ve 13'e göre tekli regresyon ağacı ve bunun bazı sınırlılıklarına çözüm sunabileceği için kullanımı önerilen topluluk öğrenme yöntemlerinin üç veri seti için performansı karşılaştırıldığında tüm veri setlerinde gradyan artırmanın daha iyi performans sergilediği söylenebilir. Ancak metrik değerleri arasında önemli farklılıklar olmadığı da göz önünde bulundurulmalıdır. Örneğin, Sinharay (2016) çalışmasında, kullandığı yöntemler arasında önemli farklılıklar olmamakla birlikte, rastgele orman ve ardından gradyan artırma algoritmasının iyi performans sergilediğini bulmuştur.

b) Tekli regresyon ağacı ile koşullu çıkarım ağacı yöntemlerinin çalışmada kullanılan veri setlerindeki (süreç/sonuç/süreç+sonuç) yordama performansları incelendiğinde daha iyi sonuç veren algoritma hangisidir?

Tablo 11, 12 ve 13'e göre tekli regresyon ağacı ve koşullu çıkarım ağacı yöntemlerinin üç veri seti için performansı karşılaştırıldığında süreç verisi ve sonuç verisi için koşullu çıkarım ağacı, süreç+sonuç verisi için regresyon ağacı daha iyi performans göstermiştir. Ancak metrik değerleri arasında önemli farklılıklar yoktur. Alan yazında bahsedilen yöntemlerin karşılaştırıldığı bazı çalışmalar bulunmaktadır. Örneğin, Hothorn ve arkadaşları (2006) tekli regresyon ağacı ve koşullu çıkarım ağaçlarını karşılaştırdıklarında,

tekli regresyon ağaçlarının fazla sayıda kesme noktasına sahip değişkenler lehine seçim yanlılığı gösterdiğini ifade etmişlerdir. Aynı zamanda araştırmacılar bu iki tür ağacın tahmin performansı benzerlik gösterse de ağaçların yapısal olarak farklı olabileceğine dikkat çekmişlerdir. Bununla birlikte Gomes ve arkadaşları (2020) çalışmalarında regresyon ağacı ve koşullu çıkarım ağacının bölünmelerde kullandıkları değişkenler bakımından farklılık gösterdiğini ifade etmişlerdir. Buna benzer bir durum bu çalışmada da elde edilmiştir. Şekil 8,9,10,11,12 ve 13 incelendiğinde regresyon ağaçlarıyla, koşullu çıkarım ağaçlarının yapısal olarak bazı farklılıklar gösterdiği ancak metrik değerleri incelendiğinde büyük ölçüde benzer sonuçlar elde edildiği görülmüştür. Dolayısıyla amacın yalnızca tahmin olduğu durumlarda iki yöntemin de belli ölçüde birbirlerinin yerine kullanılabileceği düşünülebilir. Ancak, farklı ölçek türlerindeki değişkenler içeren bir kapsamda önemli değişkenlerin tespiti gibi bir hedef söz konusuysa regresyon ağaçlarında tespit edilen değişken seçim yanlılığı riski göz önünde bulundurularak ihtiyatlı yorumlar yapılması önerilmektedir. Bununla birlikte alan yazında farklı bakış açıları da söz konusudur. Örneğin Gomes ve arkadaşları (2020) 3.670.089 kişinin verilerini içeren bir durumda regresyon ağacı ve koşullu çıkarım ağaçlarını karşılaştırmışlardır. Daha düşük performans gösterdiği bulunan koşullu çıkarım ağaçlarının, büyük veri setlerinde değişken seçim yanlılığından kaçınırken tahmin performansını kötüleştirebileceğine dair çıkarımda bulunmuşlardır. Bahsedilen değişken seçim yanlılığının ortaya koyulduğu çalışmada Hothorn ve arkadaşlarının (2006) simülasyon çalışmasına ek olarak kullandıkları gerçek veri setlerinin büyüklüğü 167 ile 990 arasında değişmektedir. Dolayısıyla tespit edilen yanlılığın örneklem büyüklüğüyle (varsa) ilişkisinin ortaya koyulması için daha fazla araştırma yapılması gerektiği düşünülmektedir.

c) *Topluluk öğrenme yöntemleri (torbalama, rastgele orman, gradyan artırma) ve koşullu çıkarım ağaçlarına dayalı topluluk öğrenme yöntemlerinin (koşullu çıkarım ağaçlarına dayalı torbalama, rastgele orman, gradyan artırma) çalışmada kullanılan veri setlerindeki (süreç/sonuç/süreç+sonuç) yordama performansları incelendiğinde daha iyi sonuç veren algoritma hangisidir?*

Tablo 11, 12 ve 13'e göre bu iki yöntem grubu içinde, aralarında ciddi farklılıklar olmamakla birlikte, süreç verisinde koşullu çıkarım ağacına dayalı gradyan artırma, sonuç verisinde gradyan artırma ve koşullu çıkarım ağacına dayalı gradyan artırma daha iyi performans gösterirken, süreç+sonuç verisinde gradyan artırma daha iyi performans göstermiştir.

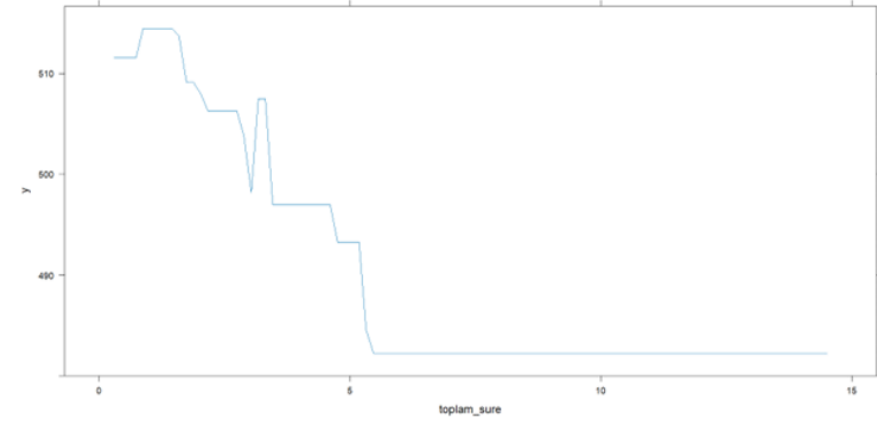
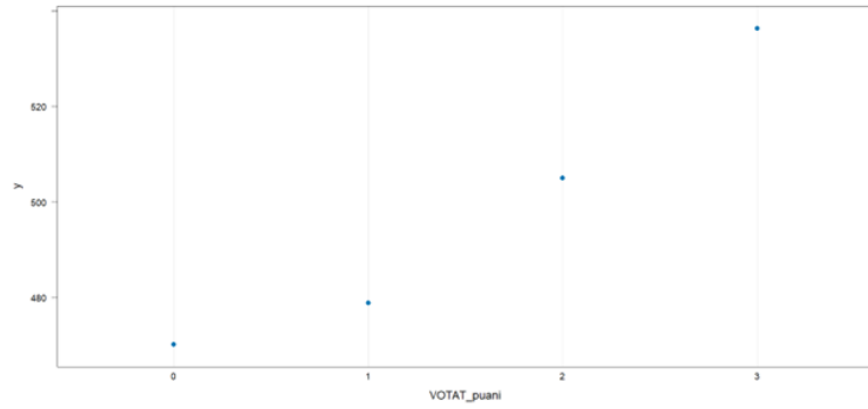
4a. Süreç verisinde karmaşık problem çözme becerisini yordamada değişkenlerin önem düzeyi sıralamaları tüm algoritmalar için nasıldır?

Tüm algoritmalara göre elde edilen değişken önem değerlerine ilişkin sıralamalar Tablo 14'te verilmiştir. Bu sıralamalara ait grafikler ise EK-G'de yer almaktadır. Bununla birlikte 4a, 4b ve 4c araştırma sorularına göre genel olarak en etkili olduğu görülen ilk iki değişken için tüm veri setlerinde gradyan artırma algoritmasından elde edilen kısmi bağımlılık grafikleri de aşağıda yer almaktadır.

Tablo 14*Süreç Verisi-Değişken Önem Sıralamaları*

	Regresyon ağacı	Torbalama	Rastgele orman	Gradyan artırma	Koşullu çıkarım ağacı	K-torbalama	K-rastgele orman	K-gradyan artırma
VOTAT puanı	1.	1.	1.	1.	1.	1.	1.	1.
Toplam süre	7.	2.	4.	3.	2.	2.	3.	2.
Bekleme süresi	-	9.	13.	8.	14.	16.	9.	15.
Eylem sayısı	-	12.	11.	15.	12.	14.	14.	13.
Sıfırlama olaylarının sayısı	6.	6.	8.	4.	6.	4.	5.	3.
Müdahale sayısı	-	7.	9.	11.	9.	6.	8.	6.
Müdahalesiz gözlem puanı	-	15.	15.	16.	3.	3.	7.	4.
Müdahalesiz gözlem sayısı	-	11.	16.	14.	8.	8.	13.	10.
VOTAT uygulama sayısı	2.	3.	2.	6.	16.	10.	4.	16.
Üst kontrol çubuğu için VOTAT sayısı	5.	13.	7.	13.	4.	11.	10.	7.
Orta kontrol çubuğu için VOTAT sayısı	4.	14.	5.	2.	7.	9.	2.	9.
Alt kontrol çubuğu için VOTAT sayısı	3.	16.	3.	12.	15.	12.	6.	5.
İlk müdahaleye kadar geçen süre	-	5.	10.	9.	10.	13.	16.	11.
Diyafram olaylarının sayısı	-	10.	14.	10.	13.	7.	11.	14.
Son kontrol süresi	-	8.	12.	7.	5.	5.	12.	8.
Hız	-	4.	6	5	11.	15	15	12.

* Diğer algoritmalarda değişken önemi permütasyon tabanlı yöntemle göre hesaplanırken, regresyon ağacı analizleri için kullanılan rpart paketinde değişken önemi farklı biçimde hesaplanır ve önem değeri %1'den az olan değişkenler gösterilmez (regresyon ağaçlarında değişken önemi hesaplaması için ayrıntılı bilgi Bölüm 2'de sunulmuştur).

Şekil 14*Süreç Verisi-Kısmi Bağımlılık Grafikleri*

Tablo 14'te verilen süreç verisine ilişkin değişken önemi sıralamalarına göre en önemli ilk iki değişkenin genel olarak VOTAT puanı ve toplam süre olduğu söylenebilir. KPÇ değerlendirmelerinde VOTAT kullanımının önemine ve performansla pozitif ilişkide olduğuna dair pek çok araştırma bulunmaktadır (Chen ve ark., 2019; Greiff ve ark., 2016; Greiff ve ark., 2018; Greiff, Wüstenberg ve ark., 2015; Han ve ark., 2019; Stadler ve ark., 2019; Teig ve ark., 2020; Wüstenberg ve ark., 2012). Bu stratejinin kullanımı arttıkça KPÇ performansının arttığı Şekil 8'de verilen regresyon ağacı, Şekil 9'da verilen koşullu çıkarım ağacı ve Şekil 14'te verilen kısmi bağımlılık grafiğinde görülmektedir. Öte yandan etkili bulunan toplam süre ile KPÇ arasında Şekil 14'teki kısmi bağımlılık grafiğine göre negatif yönde ve doğrusal olmayan ilişki olduğu söylenebilir. Buna göre daha kısa sürelerin daha yüksek KPÇ puanıyla ilişkilendirilebileceği söylenebilir. Süre 5 dakikaya yaklaşıırken KPÇ puanı azalma eğilimindedir. 5 dakikadan sonra ise KPÇ puanının anlamlı biçimde etkilenmemekle birlikte yüne düşük seviyelerde seyrettiği görülmektedir. Aynı zamanda bu grafikten elde edilen bulgu regresyon ağacı ve koşullu çıkarım ağacından elde edilen kesme değerleriyle uyumludur. Regresyon ağacında 4.4 dakikadan az, koşullu çıkarım ağacında 4.425 dakikadan fazla sürenin daha yüksek puanlara yol açtığı yönünde bir sonuç elde edilmiştir. Dikkat çeken bir değişken olan süre, süreç verisi bağlamında yapılan pek çok araştırmaya konu olmuştur (örn: Bolsinova ve Molenaar, 2018; Goldhammer ve ark., 2014; Guo ve ark., 2020; He ve ark., 2019; Klotzke ve Fox, 2019; Lee ve ark., 2019; Reis Costa ve ark., 2021; Teig ve ark., 2020; Toton ve Maynes, 2019; Wang ve ark., 2019; Yavuz, 2019; Zhan ve ark., 2021). Bununla birlikte alan yazında süre ile KPÇ performansı arasındaki ilişkiye dair bulgular çeşitlidir. Örneğin Scherer ve arkadaşları (2015) ikisi arasında orta düzeyde pozitif ilişki elde etmişlerdir. Öte yandan Csanyi ve Molnar (2023) ise süre ile KPÇ performansı arasında ilişki bulamamıştır. Bununla birlikte Greiff ve arkadaşları (2016) süre ile KPÇ yeteneğinin iki boyutu arasında ters U şeklinde bir ilişki raporlamışlardır. Dolayısıyla bu doğrusal olmayan ilişkinin sonuçlarının dikkatli biçimde yorumlanması gerekmektedir. Çok az zaman harcanması bireyin çok fazla çaba göstermediği şeklinde yorumlanabilecekken, çok fazla zaman harcanması da düşük bir

beceri ya da teknik sorunların varlığına işaret edebilir (OECD, 2024, s. 214). Sonuç olarak sürenin yetenek kestiriminde kullanılmasının daha güvenilir sonuçlara ve öğrencilerin performans farklılıklarının daha iyi anlamlandırılmasına katkı sağlayacağı ifade edilebilir.

Çalışmada belirgin biçimde etkili bulunmayan pek çok değişken de bulunmaktadır. Ancak alan yazında bu konuda da çeşitli bulgular söz konusudur. Örneğin, Greiff, Wüstenberg ve arkadaşları (2015) etkileşim sayısı, ilk etkileşime kadar geçen süre, diyagram sayısı, sıfırlama sayısı gibi değişkenlerin etkilerini oldukça küçük bulmuşlardır. Öte yandan Greiff ve arkadaşları (2016) düşük sıklıkta müdahalede bulunanların ve müdahalesiz gözlemde bulunanların daha yüksek KPÇ puanına sahip olduğunu ifade etmişlerdir. Bu çalışmada müdahale sayısı etkili değişkenler arasında bulunmamıştır. Müdahalesiz gözlem puanı ise koşullu çıkarım ağacı ve koşullu çıkarım ağaçlarına dayalı torbalama yönteminde üçüncü sırada yer almıştır ancak etkisi oldukça düşüktür. Chen ve arkadaşları (2019) çalışmasında yönergeyi daha hızlı okuyanların görevi daha düşük doğrulukta ve daha kısa sürede tamamladıklarını bulmuşlardır. Ancak bu çalışmada bekleme süresi önem sıralamasında tüm yöntemlerde oldukça alt sıralarda yer almıştır. Sıfırlama tuşunun kullanımı ile ilgili de bazı sonuçlar bulunmaktadır. Buna göre sıfırlama tuşunun performansla pozitif ilişkisi olduğunu gösteren çalışmalar olduğu gibi (Chen ve ark., 2019; Eichmann ve ark., 2020) bu tuşun kullanımının anlamlı bir etkisi olmadığını ifade eden araştırmalar da bulunmaktadır (Stadler ve ark., 2019). Bu çalışmada sıfırlama sayısı süreç+sonuç verisindeki koşullu çıkarım ağacında bölünmede kullanılmış ve süreç verisindeki koşullu çıkarım ağaçlarına dayalı gradyan artırma algoritmasında üçüncü sırada yer almıştır.

Tüm algoritmalara göre değişken önem sıralamaları birlikte incelendiğinde koşullu yöntemlerde etkilerin daha belirgin biçimde ortaya çıktığı söylenebilir. Örneğin, süreç verisindeki tekli regresyon ağacından elde edilen değişken önem sıralamasına göre süre son sırada yer alırken en etkili yordayıcılar arasında birbirleriyle yüksek ilişkide olan değişkenler yer almıştır. Ancak süre diğer algoritmalarda en etkili değişkenler arasındadır.

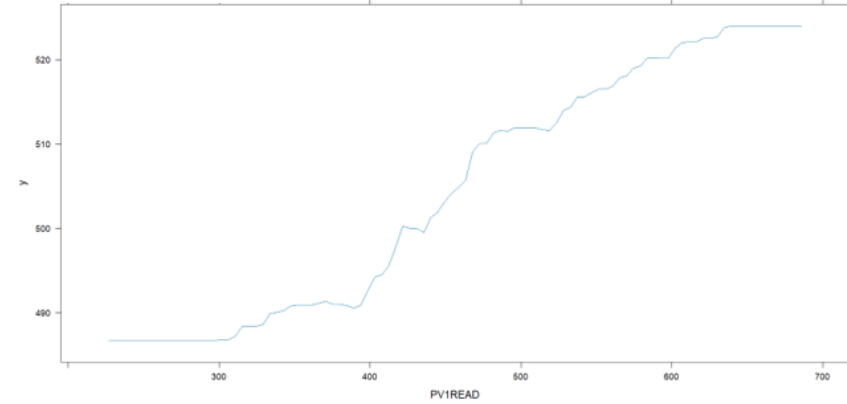
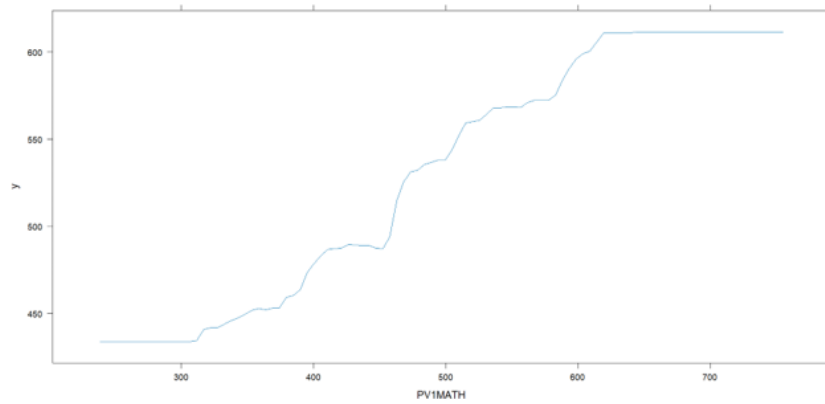
Dolayısıyla tekli regresyon ağacında ilişkili yordayıcıların varlığında meydana gelen değişken seçim yanlılığının ortaya çıktığına dair yorumda bulunulabilir. Bu tür yanlılık alan yazında rastgele orman algoritması kapsamında incelenmiştir (Nicodemus ve ark., 2010; Strobl ve ark., 2008). Bununla birlikte rastgele orman algoritmasında temel öğrenici olarak regresyon ağaçlar kullanılmaktadır. Dolayısıyla tekli regresyon ağaçlarında da bu tür bir yanlılık oluşmuş olabileceği düşünülmektedir.

4b. Sonuç verisinde karmaşık problem çözme becerisini yordamada değişkenlerin önem düzeyi sıralamaları tüm algoritmalar için nasıldır?

Tüm algoritmalara göre elde edilen değişken önem değerlerine ilişkin sıralamalar Tablo 15'te verilmiştir. Bu sıralamalara ait grafikler ise EK-H'de yer almaktadır.

Tablo 15*Sonuç Verisi-Değişken Önem Sıralamaları*

	Regresyon ağacı	Torbalama	Rastgele orman	Gradyan artırma	Koşullu çıkarım ağacı	K-torbalama	K-rastgele orman	K-gradyan artırma
Matematik	1.	1.	1.	1.	1.	1.	1.	1.
Okuduğunu anlama	3.	2.	2.	2.	2.	2.	2.	2.
Fen	2.	3.	3.	3.	3.	3.	3.	3.
Azim	4.	4.	4.	5	4.	4.	4.	4.
Problem çözme	5.	5.	5.	4.	5.	5.	5.	5.
açıklık								

Şekil 15*Sonuç Verisi-Kısmi Bağımlılık Grafikleri*

Tablo 15'te verilen sonuç verisine ilişkin değişken önemi sıralamalarına göre en önemli ilk iki değişkenin genel olarak matematik puanı ve okuduğunu anlama puanı olduğu söylenebilir. Şekil 15'te verilen kısmi bağımlılık grafiklerindeki güçlü doğrusal ilişki dikkate alındığında bu iki beceriyle KPÇ arasında pozitif ilişki olduğu söylenebilir. Bu bulgu regresyon ağacı ve koşullu çıkarım ağacından elde edilen kesme değerleriyle belli ölçüde uyumludur. Örneğin hem regresyon ağacı hem de koşullu çıkarım ağacının en sağında (daha yüksek puan alma eğiliminin olduğu bölünmeler) art arda ve daha yüksek matematik kesme puanlarına göre bölünmelerin gerçekleştiği görülmektedir. KPÇ'nin matematik ve okuduğunu anlama ile arasında bulunan bu ilişkiye benzer sonuçların elde edildiği araştırmalar bulunmaktadır (Greiff, Fischer ve ark., 2013; Molnar ve ark., 2013; Sonnleitner ve ark., 2013). Aynı zamanda dikkat çekici bir bulgu olarak KPÇ'nin okul notlarını yordadığını ve genel zihinsel yetenekle ilişkili olduğunu ifade eden çeşitli çalışmalar vardır (Greiff, Wüstenberg ve ark., 2013; Scherer ve ark., 2015; Wüstenberg ve ark., 2012). Dolayısıyla bu bulgular ışığında KPÇ becerisinin öğretiminin okul müfredatlarında yer alması gerektiği ve bu sayede bireylerin zihinsel ve akademik gelişimlerine katkıda bulunulabileceği sonucuna ulaşılabılır. Buna ek olarak tüm algoritmalarla göre değişken önem sıralamaları birlikte incelendiğinde süreç verisindeki benzer biçimde koşullu yöntemlerde etkilerin daha belirgin biçimde ortaya çıktığı söylenebilir.

4c. Süreç ve sonuç verisinde karmaşık problem çözme becerisini yordamada değişkenlerin önem düzeyi sıralamaları tüm algoritmalar için nasıldır?

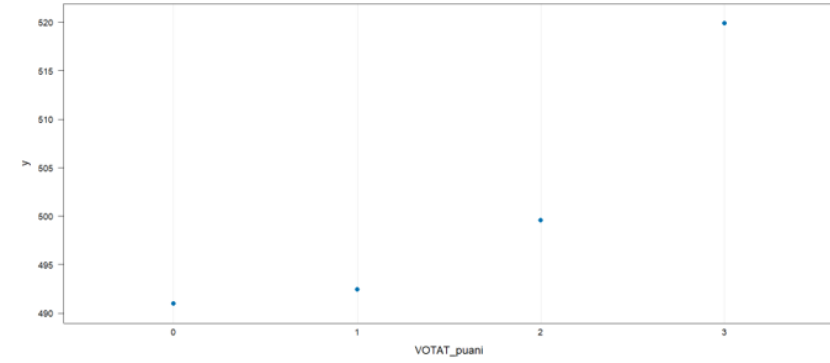
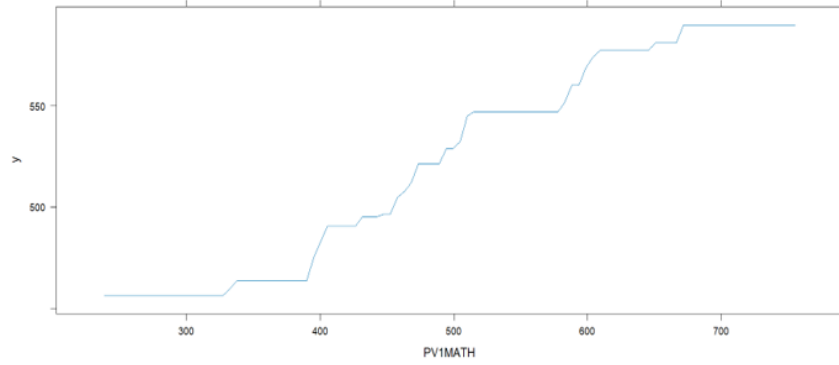
Tüm algoritmalarla göre elde edilen değişken önem değerlerine ilişkin sıralamalar Tablo 16'da verilmiştir. Bu sıralamalara ait grafikler ise EK-I'da yer almaktadır.

Tablo 16*Süreç+Sonuç Verisi-Değişken Önem Sıralamaları*

	Regresyon ağacı	Torbalama	Rastgele orman	Gradyan artırma	Koşullu çıkarım ağacı	K-torbalama	K-rastgele orman	K-gradyan artırma
VOTAT puanı	4.	2.	3.	2.	2.	2.	2.	2.
Toplam süre	-	9.	11.	9.	9.	11.	10.	6.
Bekleme süresi	-	19.	17.	13.	17.	14.	20.	17.
Eylem sayısı	-	17.	15.	20.	7.	20.	19.	15.
Sıfırlama olaylarının sayısı	9.	12.	10.	14.	4.	5.	8.	4.
Müdahale sayısı	-	18.	12.	12.	13.	15.	14.	12.
Müdahalesiz gözlem puanı	-	14.	18.	18.	6.	6.	11.	7.
Müdahalesiz gözlem sayısı	-	13.	14.	10.	12.	7.	12.	9.
VOTAT uygulama sayısı	6.	6.	5.	5.	19.	8.	6.	19.
Üst kontrol çubuğu için VOTAT sayısı	8.	4.	8.	6.	8.	10.	9.	11.
Orta kontrol çubuğu için VOTAT sayısı	5.	11.	6.	7.	11.	9.	5.	10.
Alt kontrol çubuğu için VOTAT sayısı	7.	5.	7.	11.	18.	13.	7.	18.
İlk müdahaleye kadar geçen süre	-	15.	16.	21.	14.	18.	16.	13.
Diyafram olaylarının sayısı	-	16.	20.	15.	16.	12.	18.	16.
Son kontrol süresi	-	10.	13.	17.	10.	17.	17.	8.
Hız	-	8.	9.	8.	15.	16.	13.	14.
Matematik	1.	1.	1.	1.	1.	1.	1.	1.
Okuduğunu anlama	3.	7.	4.	3.	5.	3.	4.	3.
Fen	2.	3.	2.	4.	3.	4.	3.	5.
Azim	-	20.	19.	16.	20.	19.	15.	20.
Problem çözmeye açıklık	-	21.	21.	19.	21.	21.	21.	21.

Şekil 16

Süreç+Sonuç Verisi-Kısmi Bağımlılık Grafikleri



Tablo 16’da verilen süreç+sonuç verisine ilişkin tüm algoritmalarından elde edilen değişken önemi sıralamalarına göre en önemli ilk iki değişkenin genel olarak matematik puanı ve VOTAT puanı olduğu söylenebilir. Şekil 16’da verilen kısmi bağımlılık grafikleri ve ağaç şemaları göz önüne alındığında bu iki özellik ve KPÇ performansı arasında pozitif ilişki olduğu söylenebilir. Öte yandan baskın biçimde matematik puanının önemli değişken olmasının bazı sebepleri olduğu düşünülmektedir. Örneğin Greiff ve arkadaşları (2018) KPÇ’de yüksek bir yeterliğin farklı alanlardaki problemlerin çözülmesinde de etkili olacağını ifade etmişlerdir. Aynı zamanda Molnar ve arkadaşları (2013) özellikle matematikle ilgili problemleri içeren alana özgü problem çözme ile KPÇ arasında 0.35 korelasyon olduğunu bulmuşlardır. Bununla birlikte çalışmada ele alınan klima kontrolü ünitesinin birinci maddesi KPÇ’nin bilgi edinme boyutuna karşılık gelir (OECD, 2014a, s. 38). Bilgi edinme ile ilgili görevler ise PISA kapsamında “keşfetme ve anlama” veya “temsil etme ve formüle etme” süreçlerini içerirler (OECD, 2014a, s.84). Bu süreçler PISA kapsamındaki genel problem çözme süreçlerindeki ilk iki basamağa karşılık gelmektedir. Öte yandan bu iki süreç Polya (1945) tarafından kavramsallaştırılan problem çözme basamaklarıyla örtüşmektedir. Polya’ya (1945) göre problem çözme; problemin anlaşılması, çözüm için plan yapılması, planın uygulanması ve geriye dönüp çözümün gözden geçirilmesi süreçlerinden oluşur. Okullardaki matematik öğretiminde kullanılan ders kitapları incelendiğinde problem çözme ile ilgili bölümlerin bu dört basamağa uygun biçimde yapılandırıldığı görülür (bkz. Korkmaz ve ark., 2023; Özçelik, 2023). Yani bu problem çözme süreçleri matematik müfredatında yerleşik bir konudur. Bununla birlikte çalışmada kullanılan KPÇ sorusu dinamik yapısı bakımından matematik derslerinde yaygın olarak bilinen sorularından ayrışsa da aynı problem çözme süreçlerini gerektirmektedir. Bu bağlamda matematik performansı ile KPÇ performansı arasındaki bu belirgin ilişkinin bahsedilen benzerlikten kaynaklı olabileceği düşünülmektedir.

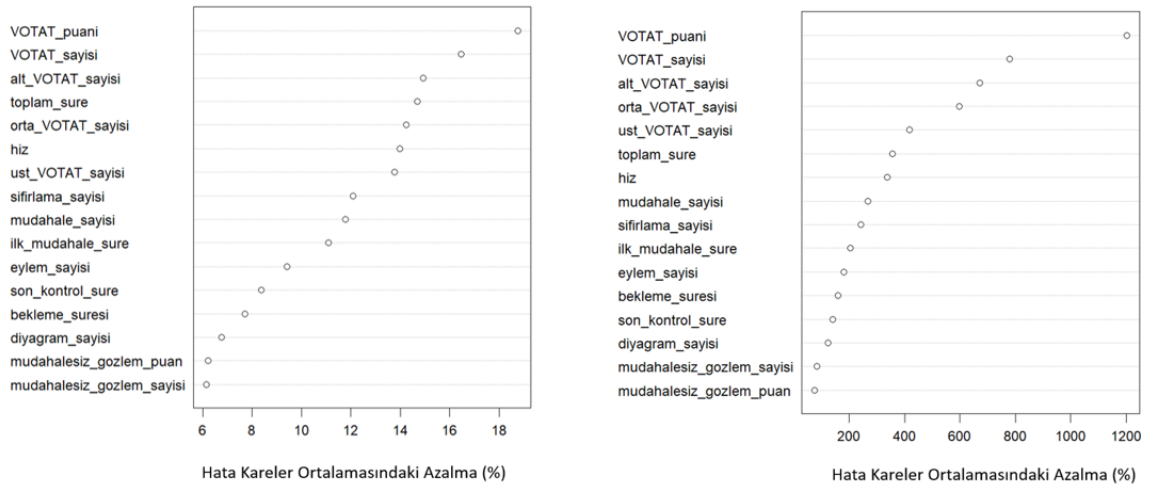
Süreç+sonuç verisindeki tüm değişken önem grafikleri incelendiğinde değişken seçim yanlılığını önlediği belirtilen koşullu çıkarım ağaçlarına dayalı yöntemle değişkenlerin

etkilerinin daha net biçimde ortaya çıktığı düşünülmektedir. Koşullu olmayan yöntemlerde birbiriyle ilişkili değişkenlerin daha önemli göründükleri dikkat çekmektedir. Örneğin, diğer algoritmalarda en etkili değişkenler arasında yer alan ve kategorik olan VOTAT puanı tekli regresyon ağacında 4. Sırada yer almıştır. İlk üç değişken sürekli ölçek türünde olan değişkenlerden oluşmuştur. Dolayısıyla bu durumun Hothorn ve arkadaşları (2006) tarafından gösterilen değişken seçim yanlılığıyla ilişkili olduğu düşünülmektedir.

5a. Süreç verisi için rastgele orman algoritmasında ölçeklenmiş ve ölçeklenmemiş versiyonlarına göre değişken önem düzeyi sıralamaları nasıldır?

Şekil 17

Süreç Verisi-Rastgele Orman Değişken Önemi (Sol: Ölçeklenmiş. Sağ: Ölçeklenmemiş)

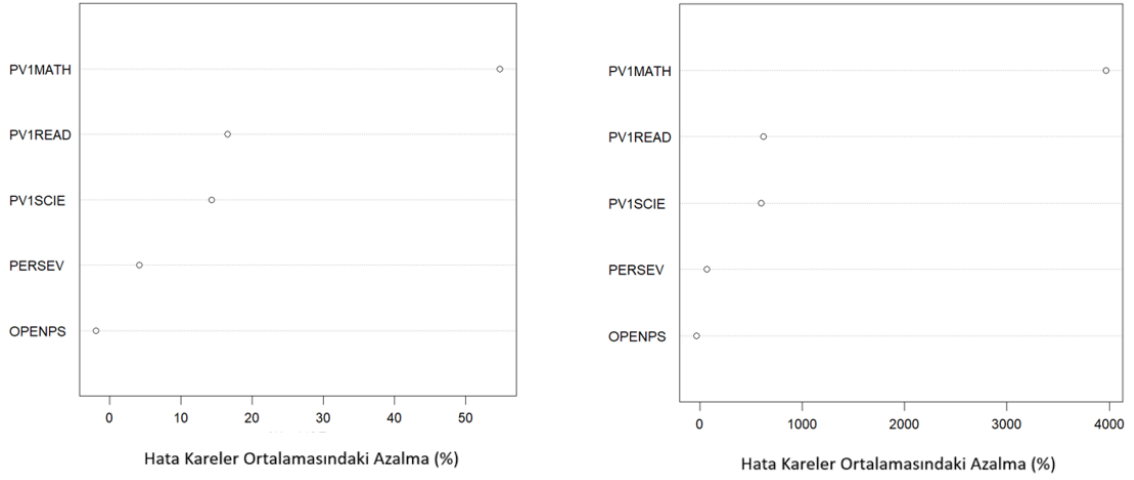


Şekil 17'ye göre değişken önemi sıralaması rastgele ormanların her iki versiyonunda da büyük ölçüde benzerlik göstermektedir. En önemli değişkenlerin VOTAT puanı, VOTAT sayısı, alt VOTAT sayısı şeklinde devam ettiği görülmektedir. Bununla birlikte VOTAT puanının etkisi ölçeklenmemiş durumda daha belirgindir.

5b. Sonuç verisi için rastgele orman algoritmasında ölçeklenmiş ve ölçeklenmemiş versiyonlarına göre değişken önem düzeyi sıralamaları nasıldır?

Şekil 18

Sonuç Verisi-Rastgele Orman Değişken Önemi (Sol: Ölçeklenmiş. Sağ: Ölçeklenmemiş)

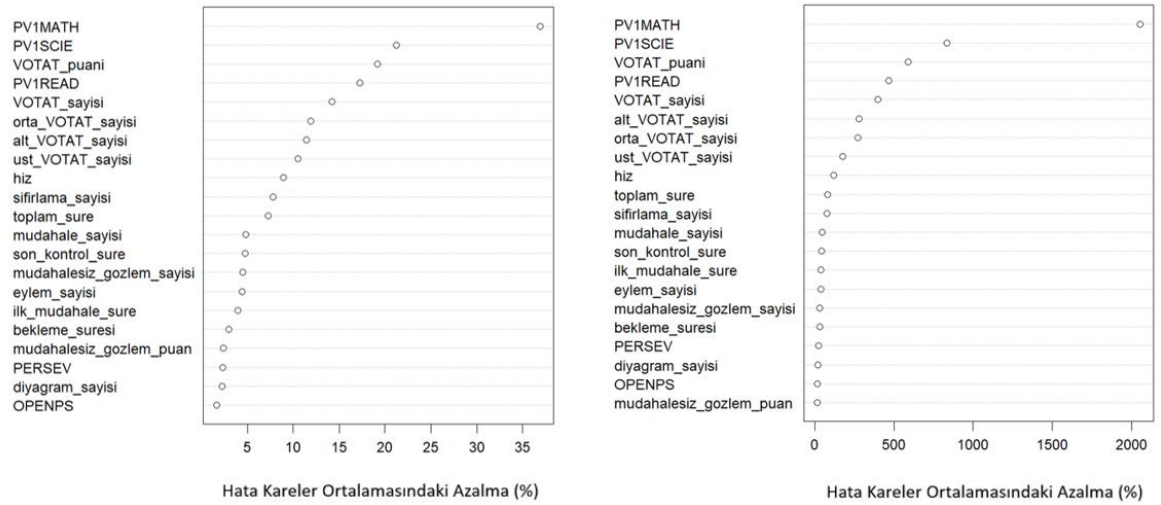


Şekil 18'e göre değişken önemi sıralamaları rastgele ormanların her iki versiyonunda aynıdır. En önemli değişken belirgin olarak matematik puanıdır. Bununla birlikte matematik puanının etkisi ölçeklenmemiş durumda daha belirgindir.

5c. Süreç+sonuç verisi için rastgele orman algoritmasında ölçeklenmiş ve ölçeklenmemiş versiyonlarına göre değişken önem düzeyi sıralamaları nasıldır?

Şekil 19

Süreç+Sonuç Verisi-Rastgele Orman Değişken Önemi (Sol: Ölçeklenmiş. Sağ: Ölçeklenmemiş)



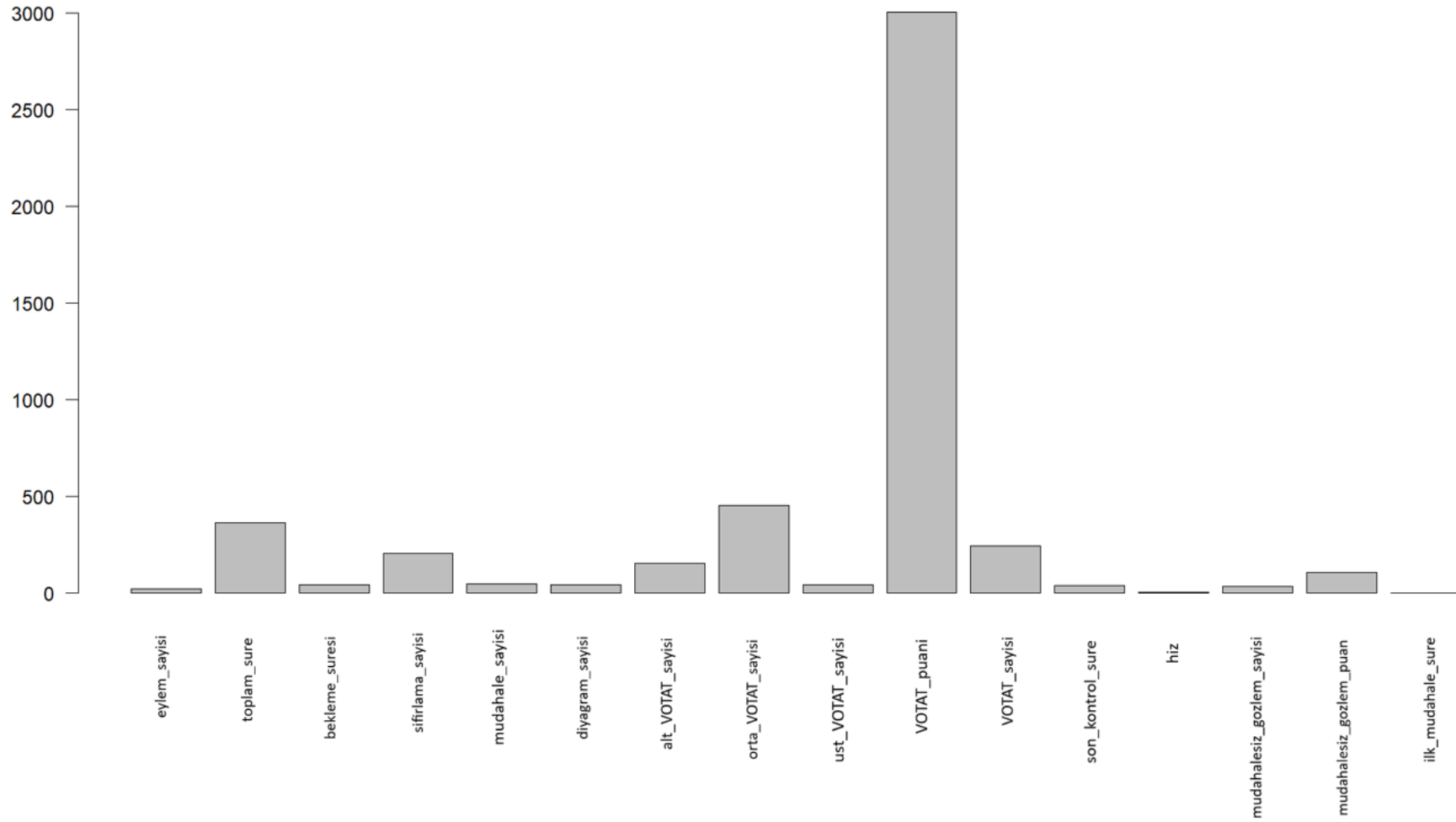
Şekil 19'a göre değişken önemi sıralamaları rastgele ormanların her iki versiyonunda büyük ölçüde benzerlik göstermektedir. En önemli değişken belirgin şekilde matematik puanıdır. Bununla birlikte matematik puanının etkisi ölçeklenmemiş durumda daha belirgindir.

Rastgele orman analizinin ölçeklenmemiş ve ölçeklenmiş versiyonlarına göre üç veri setinde elde edilmiş değişken önem sıralamalarının büyük ölçüde benzer kaldığı dikkat çekmiştir. Öte yandan üç veri setinde de tüm algoritmalar tarafından en önemli değişkenler olarak görünen matematik puanı ve VOTAT puanının etkisinin, ölçeklenmemiş versiyonlarda daha belirgin biçimde ortaya çıkmış olmasının ölçeklemenin bir dezavantajı olarak görülebileceği de düşünülmektedir. Strobl ve Zeileis (2008) çalışmasında, ölçeklemenin negatif etkilerini testin gücü bağlamında incelemişlerdir. Değişken önemi sıralamasına etkisi ile ilgili bir inceleme yapmamışlardır. Bununla birlikte Nicodemus ve arkadaşları (2010) koşullu ve koşulsuz permütasyon öneminde ölçeklemenin yanlılığa sebep olduğunu ifade etmişlerdir. Sonuç olarak çalışmada, ölçeklenmiş ve ölçeklenmemiş durumlarda değişken önem sıralamasına yönelik elde edilen bulguların alan yazına ek bir bilgi sağladığı düşünülmektedir.

6a. Süreç verisi için koşullu çıkarım ağaçlarına dayalı rastgele orman algoritmasında koşullu ve koşullu olmayan versiyonlarına göre değişken önem düzeyi sıralamaları nasıldır?

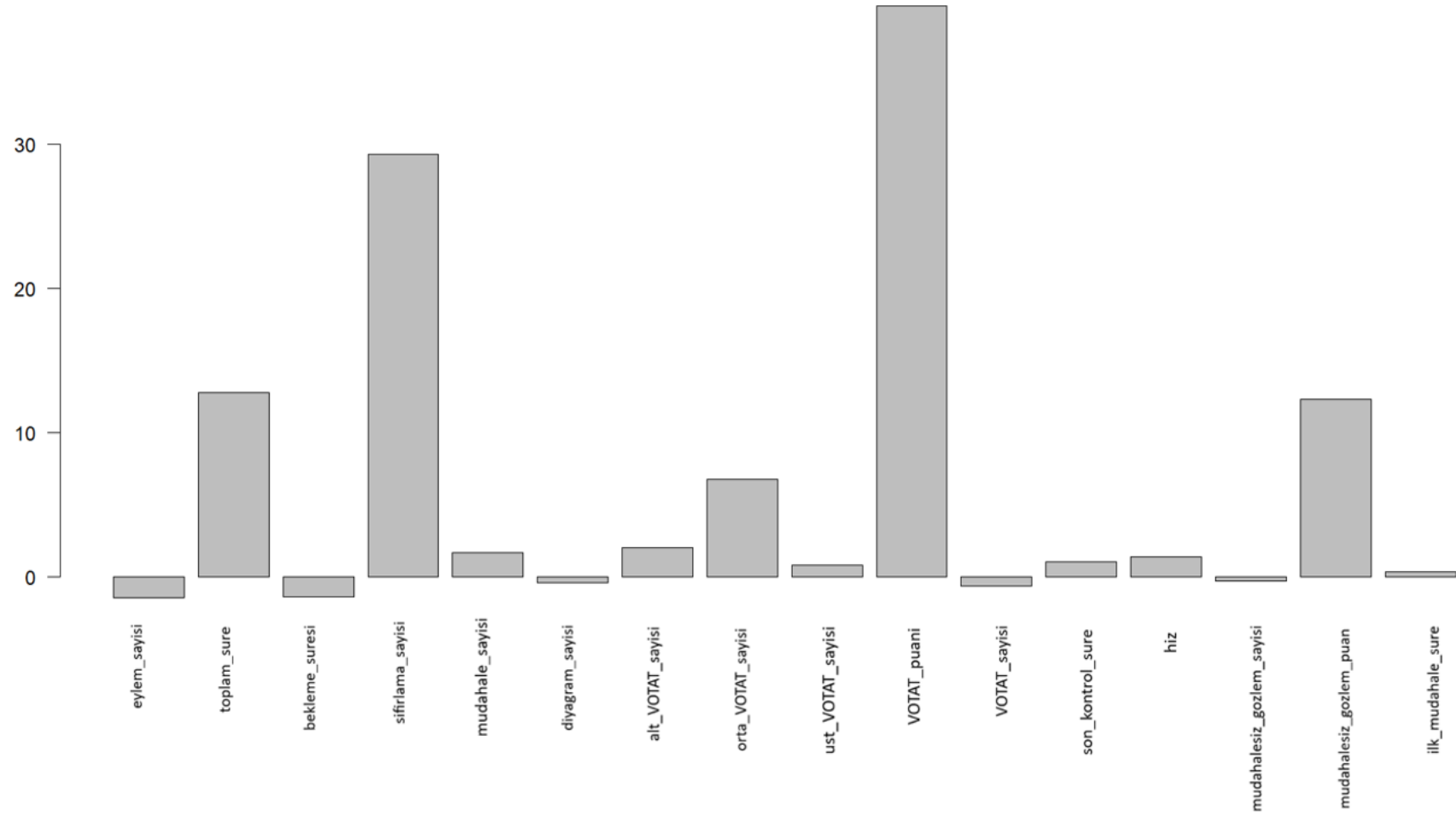
Şekil 20

Süreç Verisi-Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman Değişken Önem Grafiği (Koşullu Olmayan Versiyon)



Şekil 21

Süreç Verisi-Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman Değişken Önem Grafiği (Koşullu Versiyon)

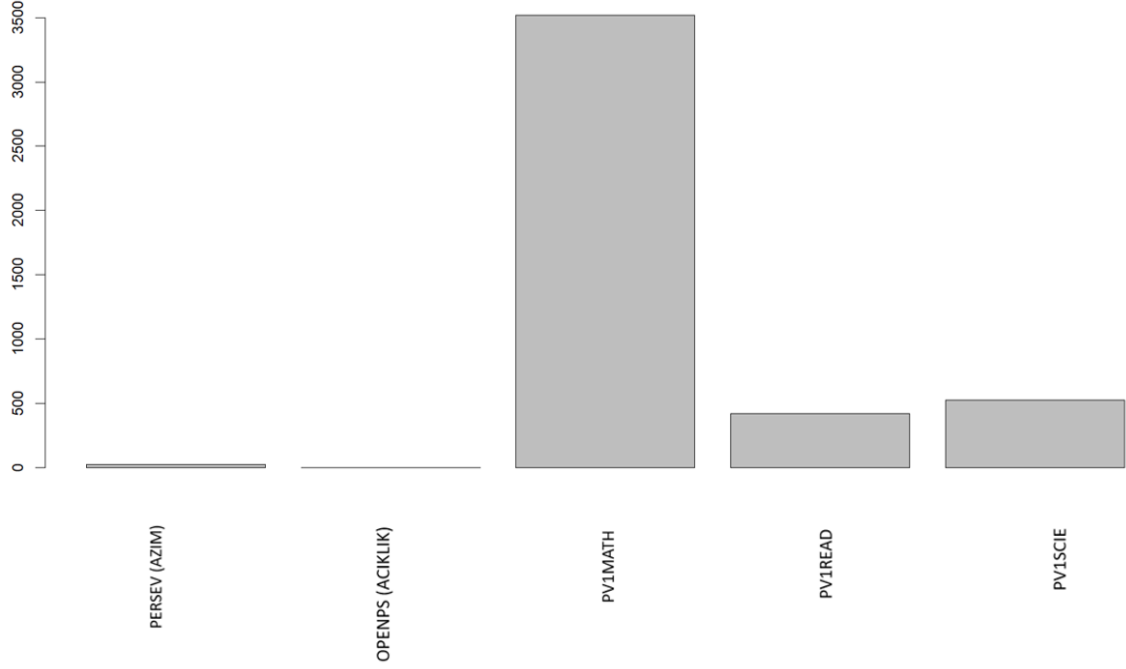


Şekil 20 ve 21'e göre her iki durumda VOTAT puanının en önemli değişken olduğu söylenebilir. Bununla birlikte koşullu durumu temsil eden Şekil 21'deki grafikte bazı değişkenlerin etkilerinin arttığı bazılarının ise azaldığı dikkat çekmektedir. Koşullu permütasyon önemi, ilişkili yordayıcılar söz konusu olduğunda değişken öneminin yapay bir şekilde artması problemine çözüm olarak geliştirilmiştir (Strobl ve ark., 2008). Bu sebeple ortaya çıkan sonuç değişkenler arası korelasyon değerleri bağlamında yorumlanmıştır. Buna göre süreç verisi söz konusu olduğunda değişkenler genelde birbirlerinden türetildikleri için ya da soru çözümünde gerekli işlemlerin birbirine bağlı olmasından dolayı ilişkili veri olması beklenen bir durumdur. Şekil 21'de sıfırlama sayısı, toplam süre ve müdahalesiz gözlem puanının etkisinin belirgin derecede arttığı göze çarpmaktadır. Bununla birlikte bazı değişkenlerin ise negatif değere düştüğü söylenebilir. Bu durumda değişkenler arasındaki ilişkiler incelendiğinde örneğin, sıfırlama sayısı ile orta düzeyde ilişkili olan değişkenler (Spearman korelasyon katsayısına göre) eylem sayısı (0.457), alt VOTAT sayısı (0.555), orta VOTAT sayısı (0.555), üst VOTAT sayısı (0.491), VOTAT puanı (0.567) ve VOTAT sayısı (0.563) olarak sıralanabilir. Koşullu durumda dikkat edilecek olursa, eylem sayısı ve VOTAT sayısının etkilerinin azaldığı görülmektedir. Yine toplam süre ile orta düzeyde ilişkili olan değişkenler eylem sayısı (0.468), bekleme süresi (0.637), ilk müdahale süresi (0.697) olarak ifade edilebilir. Yine ikinci grafikte eylem sayısı ve bekleme süresinin etkisinin azaldığı, ilk müdahale süresinin de önemsiz miktarda arttığı söylenebilir. Öte yandan müdahalesiz gözlem puanıyla, müdahalesiz gözlem sayısı arasında 0.954 korelasyon bulunmaktadır. İkinci grafikte artan müdahalesiz gözlem puanı etkisinin bu ve benzeri ilişkilerden kaynaklanabileceği düşünülmüştür.

6b. Sonuç verisi için koşullu çıkarım ağaçlarına dayalı rastgele orman algoritmasında koşullu ve koşullu olmayan versiyonlarına göre değişken önem düzeyi sıralamaları nasıldır?

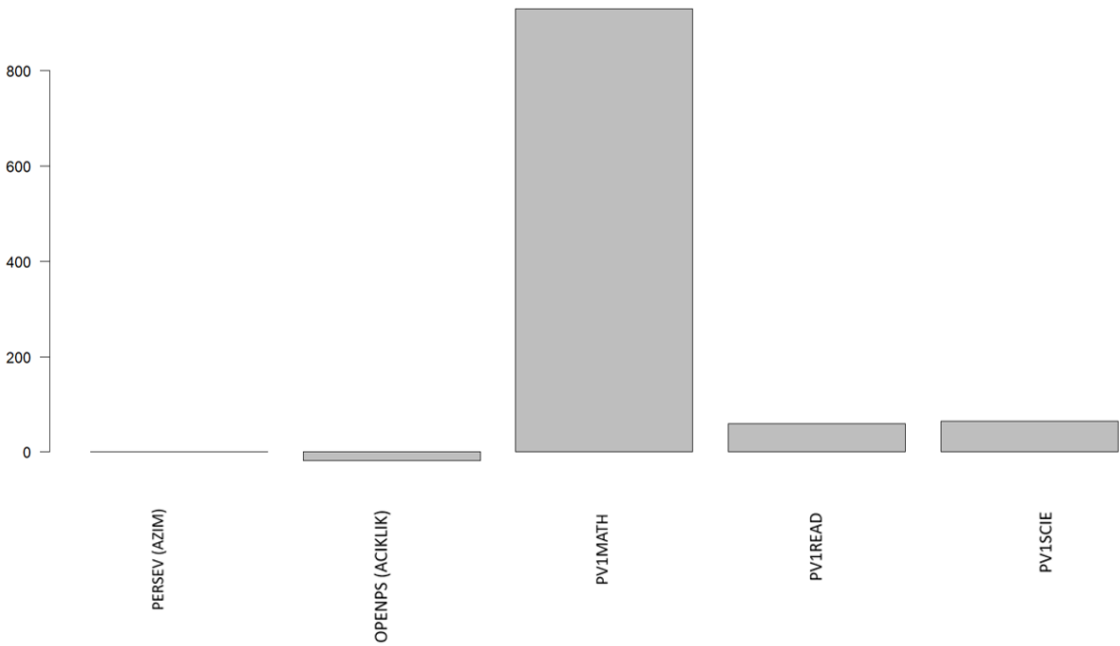
Şekil 22

*Sonuç Verisi-Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman Değişken Önem Grafiği
(Koşullu Olmayan Versiyon)*



Şekil 23

*Sonuç Verisi-Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman Değişken Önem Grafiği
(Koşullu Versiyon)*



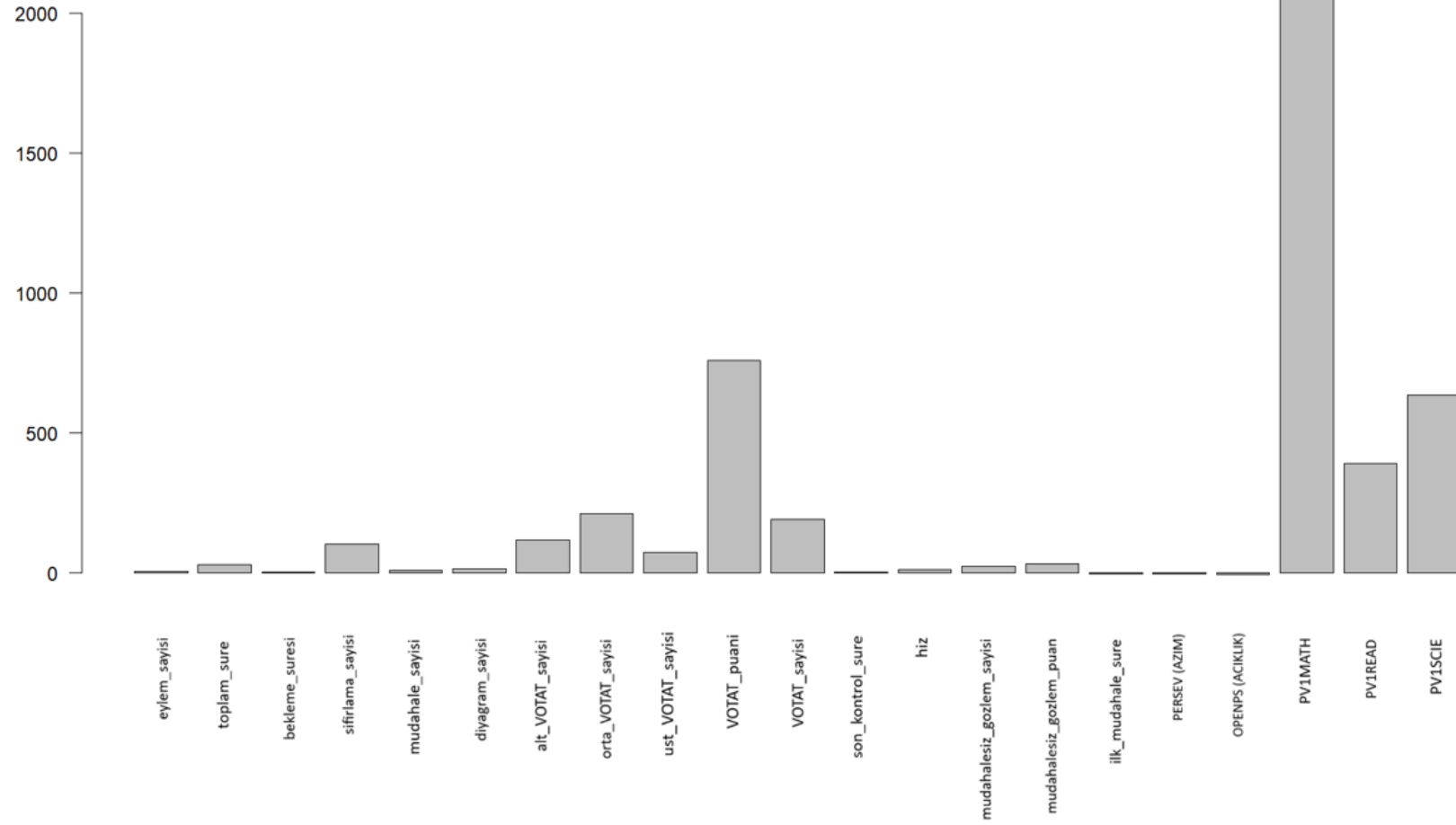
Şekil 22 ve 23'e göre her iki durumda matematik puanının en önemli değişken olduğu söylenebilir. Bununla birlikte koşullu durum için azalan okuduğunu anlama ve fen puanı etkisinin, bu değişkenlerin matematik puanıyla ilişkilerinin sırasıyla (Pearson korelasyon katsayısı) 0.797 ve 0.855 olmasıyla ilişkili olabileceği düşünülmektedir.

6c. Süreç+sonuç verisi için koşullu çıkarım ağaçlarına dayalı rastgele orman algoritmasında koşullu ve koşullu olmayan versiyonlarına göre değişken önem düzeyi sıralamaları nasıldır?



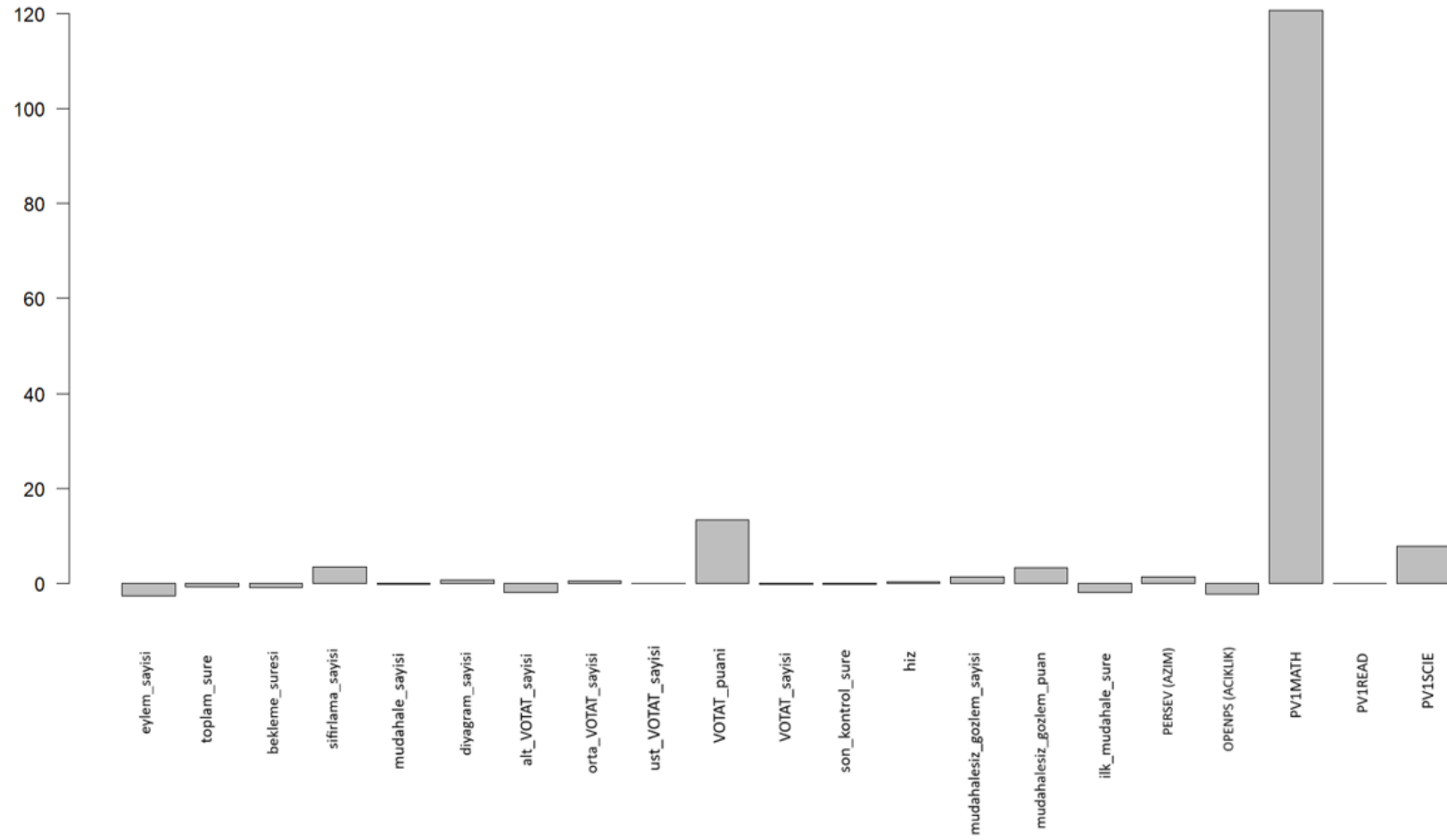
Şekil 24

Süreç+Sonuç Verisi-Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman Değişken Önem Grafiği (Koşullu Olmayan Versiyon)



Şekil 25

Süreç+Sonuç Verisi-Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman Değişken Önem Grafiği (Koşullu Versiyon)



Şekil 24 ve 25'e göre her iki durumda matematik, VOTAT ve fen puanının daha etkili değişkenler olduğu görülmektedir. Bununla birlikte koşullu versiyonda genel olarak değişkenlerin etkilerinin dikkat çekici biçimde azaldığı görülmüştür. Örneğin, matematik puanı ile VOTAT puanı arasında bulunan orta düzey (0.522), matematik ile fen arasındaki yüksek düzey ilişkinin (0.855) ve/veya matematik ile okuduğunu anlama arasındaki yüksek düzey ilişkinin (0.797) bu duruma etki ettiği yorumu yapılabilir.

Koşullu çıkarım ağaçlarına dayalı rastgele orman analizlerinde koşullu olan ve olmayan durumlar için elde edilen değişken önem sıralamalarına göre sonuç verisinde ve süreç+sonuç verisinde koşullu durumda değişkenlerin etkilerinin daha net biçimde ortaya çıkmış olduğu düşünülmektedir. Ancak süreç verisinde koşullu versiyonda, diğer yöntemlerde etkili bulunmayan, ilişkili yordayıcılardan sıfırlama sayısı ve müdahalesiz gözlem puanı değişkenlerinin etkilerinin belirgin biçimde artış göstermesi dikkat çekmiştir. Tüm veri setlerinde ilişkili yordayıcılar bulunmaktadır. Sonuç verisi ve süreç+sonuç verisi için koşullu permütasyon öneminin sağladığı yansızlığa ilişkin elde edilen bulguların alan yazındaki diğer çalışmalar tarafından desteklendiği söylenebilir (Nicodemus ve ark., 2010; Strobl ve ark., 2008). Ancak süreç verisinde ortaya çıkan beklenmedik durumun etkileri üzerine düşünülmelidir. Örneğin, süreç verisinde bulunan ilişkili yordayıcıların sayısı ve/veya aralarındaki ilişki düzeyinin elde edilen bulguyla ilişkisi üzerine çalışmalar yapılabilir. Buna ek olarak koşullu permütasyon yöntemine ilişkin grafiklerin elde edilmesi için geçen sürenin üç veri setinde de diğer yöntemlere göre çok daha fazla olduğu da belirtilmelidir. Diğer analizler birkaç saniye ya da dakika aralığında gerçekleşmişken, koşullu permütasyon önemi hesaplaması süreç verisinde 12.07 dk, sonuç verisinde 19.17 dk, süreç+sonuç verisinde 11.71 dk sürmüştür.

Bölüm 5

Sonuç ve Öneriler

Bu çalışmada süreç ve sonuç verilerinin karmaşık problem çözme becerisini yordama performanslarının çeşitli makine öğrenme algoritmaları kullanılarak incelenmesi hedeflenmiştir. Buna göre elde edilen temel sonuçlar aşağıda sıralanmıştır:

1. Tüm algoritmalar için metrik değerlerine göre; süreç verisi orta, sonuç verisi orta-iyi, süreç+sonuç verisi iyi düzeyde tahmin performansı sergilemiştir.
2. Süreç verisinde koşullu çıkarım ağaçlarına dayalı gradyan artırma, sonuç verisinde hem gradyan artırma hem de koşullu çıkarım ağaçlarına dayalı gradyan artırma, süreç+sonuç verisinde gradyan artırma algoritmasında daha iyi metrik değerleri elde edilmiştir. Tüm veri setlerinde metrik değerleri arasında mütevazı düzeyde farklılıklar olduğu göz önünde bulundurulmalıdır.
3. Tekli regresyon ağacı ve topluluk öğrenme yöntemlerinin (torbalama, rastgele orman, gradyan artırma) karşılaştırılmasına göre marjinal farklılıklar olmamakla birlikte tüm veri setlerinde gradyan artırma daha iyi performans göstermiştir.
4. Tekli regresyon ağacı ve koşullu çıkarım ağacının karşılaştırılmasına göre; süreç verisinde ve sonuç verisinde koşullu çıkarım ağacı, süreç+sonuç verisinde tekli regresyon ağacı daha iyi performans göstermiştir. Bununla birlikte iki ağaç arasında tüm veri setleri için yapısal bazı farklılıklar olmakla birlikte tahmin performansları benzerlik göstermiştir.
5. Topluluk öğrenme yöntemlerinin ve koşullu çıkarım ağaçlarına dayalı topluluk öğrenme yöntemlerinin karşılaştırılmasına göre; süreç verisinde koşullu çıkarım ağaçlarına dayalı gradyan artırma, sonuç verisinde gradyan artırma ve koşullu çıkarım ağaçlarına dayalı gradyan artırma, süreç+sonuç verisinde gradyan artırma nispeten daha iyi performans göstermiştir.

6. Süreç verisi için KPÇ'yi yordamada en etkili değişkenler tüm algoritmalar için genel olarak VOTAT puanı ve toplam süre olmuştur. VOTAT puanı ile KPÇ arasında pozitif, toplam süre ile KPÇ arasında negatif yönde ve doğrusal olmayan bir ilişki olduğu görülmüştür.
7. Sonuç verisi için KPÇ'yi yordamada en etkili değişken tüm algoritmalar için baskın biçimde matematik puanı olmuştur. Bunu okuduğunu anlama puanı takip etmiştir. Her iki değişkenle KPÇ arasında güçlü doğrusal ve pozitif bir ilişki olduğu görülmüştür.
8. Süreç+sonuç verisi için KPÇ'yi yordamada en etkili değişkenler tüm algoritmalar için genel olarak matematik puanı ve VOTAT puanı olmuştur. Her iki değişkenle KPÇ arasında pozitif ilişki olduğu görülmüştür.
9. Rastgele orman algoritmasında değişken önemi hesaplanırken ölçeklemenin yapıldığı ve yapılmadığı koşullarda tüm veri setleri için değişken önem sıralamaları neredeyse aynı kalmıştır. Bununla birlikte ölçeklenmemiş versiyonda en etkili süreç verisi değişkeni olan VOTAT puanının ve en etkili sonuç verisi değişkeni olan matematik puanının etkileri daha belirgin biçimde ayrılmıştır.
10. Koşullu çıkarım ağaçlarına dayalı rastgele orman algoritmasında değişken önemi hesaplanırken koşullama işleminin yapıldığı ve yapılmadığı durumlar için kendi aralarında ilişkili olan bazı yordayıcıların önem sıralamalarında artış-azalışlar görülmüştür. Özellikle bu çalışmadaki gibi ilişkili yordayıcıları içeren veri setlerinde koşullama işlemiyle değişken önemlerinin daha net biçimde ortaya çıkarılabileceği düşünülmektedir.

Elde edilen sonuçlar yorumlanırken, algoritmaların metrik değerleri arasındaki farkların oldukça mütevazi düzeyde olduğu vurgulanmalıdır. Öte yandan tüm veri setlerinin kabul edilebilir düzeyde tahmin performansı sergilemiş olması ve KPÇ'yi etkileyen/etkilemeyen değişkenlerin belirlenmiş olması bağlamında alan yazına katkıda

bulunulacağı umulmaktadır. Tüm bulgular birlikte dikkate alındığında çalışmanın temel olarak aşağıdaki konulara ışık tuttuğu düşünülmektedir:

1. Her veri türünün kendi içinde avantajları ve sınırlılıkları bulunmaktadır. Elbette ki araştırmacılar amaçlarına, testin uygulama koşullarına göre tercihte bulunacaklardır. Ancak süreç verilerinin kullanılmasının, bir anlamda bilişsel sürecin haritalandırılmasıyla, bireyler hakkında derinlemesine bilgi edinebilmek için sunduğu esneklik test geliştirme süreçlerinde dikkate alınmalıdır. Örneğin, bu çalışmada etkili olduğu görülen strateji ve süre değişkenleri test puanının hesaplanmasına dahil edilebilir. Ya da problem çözme esnasında bu değişkenlerden elde edilen bilgilere göre test oturumu sırasında uyarı ve/veya geri bildirim veren test sistemleri geliştirilebilir. Bununla birlikte elbette ki bu tip değerlendirme araçlarının bilgisayar tabanlı uygulama gerektirmesi sebebiyle ulaşılabilirlik sorunu, teknik alt yapı eksikliği vb. gibi bazı sınırlılıklara sahip olduğu ve buna bağlı olarak da yüksek riskli sınavlarla doğrudan uygulanamayabileceği söylenmelidir. Ancak bu test sistemleri örneğin, EBA (Eğitim Bilişim Ağı) platformu aracılığıyla uygulanabilirse hem bu sınırlılıklar önemli ölçüde aşılabilecek hem de süreç verilerinden elde edilen bilgilere göre öğrencilere uygulanacak ek öğretimlerin içerikleri daha isabetli hale getirilebilecektir. Örneğin, bu verilerden elde edilen bilgilere göre öğrenciyi, ek öğretime ihtiyacı olan konu içeriğine sahip eğitici videolara yönlendiren bir test sistemi zaman ve kaynak tasarrufu da sağlayacaktır. Bu sayede ölçme-değerlendirme sürecinden elde edilen bilgiler doğrudan öğretim amaçlı kullanılmış olacaktır. Buna ek olarak süreç verilerinin kullanımı yalnızca KPÇ alanı ile sınırlı değildir. Bu veriler fen bilimleri (örn. Scalise ve Clarke Midura, 2018; Lee ve ark., 2019; Teig ve ark., 2020), yazma becerileri (örn. Deane, 2014; Zhang ve Deane, 2015), yetişkin yeterliklerinin değerlendirilmesi gibi farklı eğitim basamaklarındaki aritmetik, problem çözme, okuma becerileri (örn. Goldhammer ve ark., 2014; He ve ark., 2019; Klotzke ve Fox, 2019; Liao ve ark., 2019; Tang ve ark.,

2020; Zhang ve ark., 2023) alanlarında veya farklı madde tiplerinde de (örn. Arieli Attali ve ark., 2019; Han ve Wilson, 2022) kullanılabilir. Bununla birlikte süreç ve sonuç verilerinin birlikte kullanımının bireylerin yetenekleri hakkında elde edilebilecek bilginin kapsamını artıracak ve birbirlerini destekler nitelikte oldukları ifade edilebilir.

2. Önemli bir 21. yüzyıl becerisi olan KPÇ'nin matematik, okuduğunu anlama, fen bilimleri alanlarındaki başarıyla olan ilişkisi de dikkate alındığında okul müfredatlarına bu beceriye ilişkin hedef davranışların eklenmesi önerilebilir. Örneğin, KPÇ'de gerekli olan problem çözme stratejilerinin derslerde öğretilmesiyle bireylerin zihinsel ve akademik gelişimlerine destek olunacağı düşünülmektedir.
3. Kullanılan makine öğrenme algoritmalarından elde edilen kesme değerlerinin tanılama, modelleme, anlık uyarı ya da geri bildirim veren test sistemlerinin geliştirilmesi gibi çalışmalarda kullanılmak üzere bilgi verebileceği düşünülmektedir.
4. Bu çalışmadaki gibi ilişkili, farklı ölçek türlerinde yordayıcılar içeren, parametrik tekniklerin varsayımlarının karşılanmadığı karmaşık veri setlerinde makine öğrenme algoritmalarının kullanımının önemli bir alternatif olabileceği düşünülmektedir.
5. Çalışmada tüm veri setlerindeki değişken önem sıralamaları incelendiğinde etkiler genel olarak koşullu yöntemlerde daha net biçimde görülmüştür. Başka bir deyişle, bu yöntemlerde etkiler her bir değişkenin bağımlı değişkenle olan korelasyonunu daha iyi yansıtacak şekilde ortaya çıkmıştır. Özellikle tekli regresyon ağaçlarında ilişkili yordayıcılar ve farklı ölçek türlerindeki değişkenlerin varlığında söz konusu olan değişken seçim yanlılığının görüldüğü düşünülmektedir. Alan yazında bu kapsamdaki bulgular sınıflandırma ağaçlarına dayalıdır. Bu çalışmada elde edilen bulguların ise regresyon ağaçları temelinde elde edilmiş olmasının önemli olduğu düşünülmektedir. Bununla birlikte elde edilen sonuçlara ilişkin daha kapsamlı araştırmalar yapılması gerektiği de söylenmelidir. Öte yandan değişkenlerin önem düzeylerinin belirlenmesinin birincil amaç olduğu çalışmalarda, bu çalışmadaki gibi

ilişkili yordayıcıları ve farklı ölçek türlerindeki değişkenleri içeren veri setleri söz konusuysa koşullu yöntemlerin dikkate alınması ve yorumlamaların ihtiyatla yapılması önerilmektedir.

Sınırlılıklar ve Öneriler

Bu bölümde sınırlılıklar, uygulamaya ve ileride yapılacak araştırmalara yönelik önerilere yer verilmiştir.

Uygulamaya Yönelik Sınırlılıklar ve/veya Öneriler

1. Makine öğrenme algoritmaları rastgele süreçleri içerdiği için gerekli fonksiyonların aynı `set.seed()` değeri ile çalıştırılmasına dikkat edilmelidir.
2. Çalışmada regresyon ağacı, rastgele orman ve gradyan artırma algoritmaları için hem hiper-parametre ayarlamasında hem de model kurulmasında öğrenci ağırlıkları dahil edilebilmiştir. Ancak koşullu çıkarım ağaçlarına dayalı yöntemlerde öğrenci ağırlıklarının dahil edildiği modeller çözümsüz kalmıştır. Dolayısıyla makine öğrenme algoritmaları için kurulan modellere ağırlıklar dahil edilememiştir.
3. Farklı ölçek türlerinden kaynaklı değişken seçim yanlılığı riski oluşturacak veri setinin bulunduğu bir kapsamda, bu çalışmada olduğu gibi koşullu çıkarım ağaçlarına dayalı yöntemler kullanılmayacaksa, her biri benzer değer aralığına sahip olacak şekilde değişken dönüşümü yapılması düşünülebilir. Ancak bu durumda ortaya çıkması muhtemel veri kaybı da dikkate alınmalıdır. Örneğin, süre ile ilgili değişkenler için logaritmik dönüşüm tercih edildiğinde 1 dakikaya eşit değerler için 0 sonucu elde edilmesi söz konusu olacaktır.
4. Koşullu çıkarım ağacı algoritmasında değişken önem grafiği için `partykit` paketinde yerleşik bir fonksiyon bulunmaktadır. Ancak bu çalışmada görsel açıdan daha uygun grafikler sunması sebebiyle bazı algoritmalar için `DALEX` paketi tercih edilmiştir.

İleride Yapılacak Araştırmalara Yönelik Öneriler

1. Bu çalışmada anlık süreç verileri kullanılmıştır. Bu verilere ek olarak ardışık olay dizilerini içeren (n-gram yaklaşımı) değişkenlerle çalışmalar yapılabilir.
2. Çalışmada çözüm için VOTAT stratejisinin kullanılmasını gerektiren bir tane KPÇ maddesine ait günlük dosyası kullanılmıştır. İleride yapılacak araştırmalarda PISA 2012 problem çözme değerlendirmesindeki farklı stratejiler gerektiren maddeler kullanılabilir.
3. Yetenek kestirimine süreç verilerinin de dahil edildiği puanlama sistemine sahip, test oturumu sırasında elde edilen süreç verilerine göre bireylere anlık geri bildirim sağlayan, sosyal-duygusal becerileri ölçen bilgisayar tabanlı test sistemleri geliştirilebilir.
4. Çalışmada yalnızca makine öğrenme algoritmaları kullanılmıştır. Daha sonraki araştırmalarda bu yöntemlerin parametrik karşılıklarıyla kıyaslamalar yapılarak veri setinin yapısı bağlamında bulgulara ulaşılabilir.
5. Rastgele orman algoritmasındaki değişken önemi hesaplanırken yapılan koşullama işleminde sonuç ve süreç+sonuç verisinde etkiler daha belirgin biçimde ortaya çıkmıştır. Ancak süreç verisinde koşullu versiyonda bazı ilişkili yordayıcıların etkileri beklenmedik biçimde artış göstermiştir. Bu durumun olası sebeplerine ilişkin kapsamlı simülasyon çalışmaları yapılabilir.
6. Çalışmada kullanılan algoritmalar denetimli öğrenme yöntemlerindendir. Denetimsiz öğrenme yöntemleri kullanılarak bireylerin yetenek profilleriyle ilgili incelemeler yapılabilir.
7. Çalışmada kullanılan veri seti orta büyüklüktedir. Buradan elde edilen sonuçların, özellikle değişken öneminin hesaplanmasında incelenen farklı koşullar için, örneklem büyüklüğü ile ilişkisine yönelik araştırmalar yapılabilir.

Kaynaklar

- Ahn, H. (2014). Fifty years of classification and regression trees. *International Statistical Review*, 82(3), 357–359. <https://doi.org/10.1111/insr.12061>
- Arieli-Attali, M., Ou, L., & Simmering, V. R. (2019). Understanding test takers' choices in a self-adapted test: A hidden Markov modeling of process data. *Frontiers in Psychology*, 10, Article 83. <https://doi.org/10.3389/fpsyg.2019.00083>
- Aydođan, İ., & Gelbal, S. (2022). Fen başarısını yordayan özelliklerin sınıflandırma ve regresyon ağacı (SRA) yöntemiyle belirlenmesi: TIMSS 2015 Türkiye örneđi. *Eđitim ve Bilim*, 47(209), 239–259. <https://doi.org/10.15390/EB.2022.9368>
- Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, 50(1), 84–94. <https://doi.org/10.1080/00461520.2015.1004069>
- Biecek, P. (2018). DALEX: Explainers for complex predictive models in R. *Journal of Machine Learning Research*, 19(84), 1–5. <https://doi.org/10.48550/arXiv.1806.08915>
- Biecek, P., & Burzykowski, T. (2020). *Explanatory model analysis: Explore, explain, and examine predictive models. With examples in R and Python.* <https://ema.drwhy.ai/featureImportance.html>
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Blanch, A., & Aluja, A. (2013). A regression tree of the aptitudes, personality, and academic performance relationship. *Personality and Individual Differences*, 54(6), 703–708. <https://doi.org/10.1016/j.paid.2012.11.032>
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443–459. <https://doi.org/10.1007/BF02293801>

- Bolsinova, M., & Molenaar, D. (2018). Modeling nonlinear conditional dependence between response time and accuracy. *Frontiers in Psychology*, 9, Article 370012. <https://doi.org/10.3389/fpsyg.2018.01525>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140. <https://doi.org/10.1007/BF00058655>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., & Cutler, A. (2003). *Random forests – Classification manual*. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#varimp
- Breiman, L., & Cutler, A. (2008). *Random forests – Classification manual*. <http://www.math.usu.edu/~adele/forests/>
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *CART: Classification and regression trees*. CRC Press.
- Buchner, A. (1995). Basic topics and approaches to the study of complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 27–63). NJ: Erlbaum.
- Bühlmann, P., & Hothorn, T. (2007) Boosting Algorithms: Regularization, prediction and model fitting. *Statistical Science*, 22(4), 477–505. <https://doi.org/10.1214/07-STS242>
- Cavalli, E., Brèthes, H., Lefèvre, E., Ahmadi, A., Duncan, L. G., Bianco, M., Melmi, J. B., Denis-Noël, A., & Colé, P. (2024). Screening for dyslexia in university students: A standardized procedure based on conditional inference trees. *Archives of Clinical Neuropsychology*, 39(5), 557–574. <https://doi.org/10.1093/arclin/acad103>
- Chen, W.H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265–289. <https://doi.org/10.2307/1165285>

- Chen, Y., Li, X., Liu, J., & Ying, Z. (2019). Statistical analysis of complex problem-solving process data: An event history analysis approach. *Frontiers in Psychology*, 10, Article 486. <https://doi.org/10.3389/fpsyg.2019.00486>
- Cho, C., Lee, E., Park, G., Cho, E., Kim, N., Shin, J., Woo, S., Ha, J., & Hwang, J. (2022). Evaluation of facial skin age based on biophysical properties in vivo. *Journal of Cosmetic Dermatology*, 21(8), 3546–3554. <https://doi.org/10.1111/jocd.14653>
- Csányi, R., & Molnár, G. (2023). How do test-takers rate their effort? A comparative analysis of self-report and log file data. *Learning and Individual Differences*, 106, 102340. <https://doi.org/10.1016/j.lindif.2023.102340>
- Cutler, A., Cutler, D.R., & Stevens, J.R. (2012). Random forests. In C. Zhang & Y. Ma (Eds.), *Ensemble machine learning: Methods and applications* (pp. 157–176). Springer.
- De Boeck, P., & Jeon, M. (2019). An overview of models for response times and processes in cognitive tests. *Frontiers in Psychology*, 10, Article 422756. <https://doi.org/10.3389/fpsyg.2019.00102>
- de Klerk, S., Veldkamp, B. P., & Eggen, T. J. H. M. (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. *Computers & Education*, 85, 23–34. <https://doi.org/10.1016/j.compedu.2014.12.020>
- Deane, P. (2014). Using writing process and product features to assess writing quality and explore how those features relate to other literacy tasks. *ETS Research Report Series*, 2014(1), 1–23. <https://doi.org/10.1002/ets2.12002>
- Debeer, D., & Strobl, C. (2020). Conditional permutation importance revisited. *BMC Bioinformatics*, 21, 307. <https://doi.org/10.1186/s12859-020-03622-2>
- Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, 8, Article 266802. <https://doi.org/10.3389/fpsyg.2017.01153>

- Eichmann, B., Goldhammer, F., Greiff, S., Pucite, L., & Naumann, J. (2019). The role of planning in complex problem solving. *Computers & Education*, 128, 1–12. <https://doi.org/10.1016/j.compedu.2018.08.004>
- Eichmann, B., Greiff, S., Naumann, J., Brandhuber, L., & Goldhammer, F. (2020). Exploring behavioural patterns during complex problem-solving. *Journal of Computer Assisted Learning*, 36(6), 933–956. <https://doi.org/10.1111/jcal.12451>
- El Morr, C., Jammal, M., Ali-Hassan, H., & Walid, E. H. (2022). *Machine learning for practical decision making: A multidisciplinary perspective with applications from healthcare, engineering and business analytics*. Springer. <https://doi.org/10.1007/978-3-031-16990-8>
- Engelhardt, L., & Goldhammer, F. (2019). Validating test score interpretations using time information. *Frontiers in Psychology*, 10, Article 424879. <https://doi.org/10.3389/fpsyg.2019.01131>
- Ferré, Á., Poca, M. A., Moncho, D., Urbizu, A., Romero, O., Sampol, G., & Sahuquillo, J. A. (2019). Conditional inference tree model for predicting sleep-related breathing disorders in patients with Chiari malformation type 1: Description and external validation. *Journal of Clinical Sleep Medicine*, 15(1), 89–99. <https://doi.org/10.5664/jcsm.7578>
- Ferreira, A.J., & Figueiredo, M.A.T. (2012). Boosting algorithms: A review of methods, theory, and applications. In C. Zhang & Y. Ma (Eds.), *Ensemble machine learning: Methods and applications* (pp. 35–86). Springer.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education*. McGraw Hill.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>

- Friedman, J., Hastie, T., & Tibshirani, R. (2000). Special invited paper. Additive logistic regression: A statistical view of boosting. *Annals of Statistics*, 28(2), 337–374. <https://doi.org/10.1214/aos/1016218223>
- Friedman, J.H. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Fu, Y., Chen, Q., & Zhan, P. (2023). Binary modeling of action sequences in problem-solving tasks: One-and two-parameter action sequence model. *Acta Psychologica Sinica*, 55(8), 1383–1396. <https://doi.org/10.3724/SP.J.1041.2023.01383>
- Fu, Y., Zhan, P., Chen, Q., & Jiao, H. (2023). Joint modeling of action sequences and action time in computer-based interactive tasks. *Behavior Research Methods*, 1–18. <https://doi.org/10.3758/s13428-023-02178-2>
- Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking & Reasoning*, 7(1), 69–89. <https://doi.org/10.1080/13546780042000046>
- Funke, J. (2003). *Problemlöendes Denken*. Kohlhammer.
- Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing*, 11(2), 133–142. <https://doi.org/10.1007/s10339-009-0345-0>
- Funke, J. (2012). Complex problem solving. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 682–685). Springer.
- Genuer, R., & Poggi, J. M. (2020). *Random forests with R*. Springer International Publishing.
- Gholami, H., Mohamadifar, A., Sorooshian, A., & Jansen, J. D. (2020). Machine-learning algorithms for predicting land susceptibility to dust emissions: The case of the Jazmurian Basin, Iran. *Atmospheric Pollution Research*, 11(8), 1303–1315. <https://doi.org/10.1016/j.apr.2020.05.009>

- Gobert, J. D., Kim, Y. J., Sao Pedro, M. A., Kennedy, M., & Betts, C. G. (2015). Using educational data mining to assess students' skills at designing and conducting experiments within a complex systems microworld. *Thinking Skills and Creativity*, 18, 81–90. <https://doi.org/10.1016/j.tsc.2015.04.008>
- Goldhammer, F., & Zehner, F. (2017). What to make of and how to interpret process data. *Measurement: Interdisciplinary Research and Perspectives*, 15(3-4), 128–132. <https://doi.org/10.1080/15366367.2017.1411651>
- Goldhammer, F., Hahnel, C., & Kroehne, U. (2020). Analysing log file data from PIAAC. In D. B. Maehler & B. Rammstedt (Eds.), *Large-scale cognitive assessment: Methodology of educational measurement and assessment* (pp. 239–269). Springer. https://doi.org/10.1007/978-3-030-47515-4_10
- Goldhammer, F., Naumann, J., Rölke, H., Stelter, A., & Tóth, K. (2017). Relating product data to process data from computer-based competency assessment. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Competence assessment in education: Research, models and instruments* (pp. 407–425). Springer International Publishing. https://doi.org/10.1007/978-3-319-50030-0_24
- Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, 106(3), 608–626. <https://doi.org/10.1037/a0034716>
- Gomes, C. M. A., & Jelihovschi, E. (2020). Presenting the regression tree method and its application in a large-scale educational dataset. *International Journal of Research & Method in Education*, 43(2), 201–221. <https://doi.org/10.1080/1743727X.2019.1654992>

- Gomes, C. M. A., Lemos, G. C., & Jelihovschi, E. G. (2020). Comparing the predictive power of the CART and CTREE algorithms. *Avaliação Psicológica*, 19(1), 87–96. <https://doi.org/10.15689/ap.2020.1901.17737.10>
- Gomes, C. M. A., Lemos, G. C., & Jelihovschi, E. G. (2021). The reasons why the regression tree method is more suitable than general linear model to analyze complex educational datasets. *Revista Portuguesa de Educação*, 34(2), 42–63. <https://doi.org/10.21814/rpe.18044>
- Goode, N., & Beckmann, J. F. (2010). You need to know: There is a causal relationship between structural knowledge and control performance in complex problem solving tasks. *Intelligence*, 38(3), 345–352. <https://doi.org/10.1016/j.intell.2010.01.001>
- Goodman, M., Finnegan, R., Mohadjer, L., Krenzke, T., & Hogan, J. (2013). *Literacy, numeracy, and problem solving in technology-rich environments among U.S. adults: Results from the program for the international assessment of adult competencies 2012. First look* (NCES 2014-008). Washington, DC: U.S. Department of Education, National Center for Education Statistics. <https://nces.ed.gov/pubs2014/2014008.pdf>
- Greiff, S., & Fischer, A. (2013). Measuring complex problem solving: An educational application of psychological theories. *Journal for Educational Research Online*, 5(1), 38–58. <https://doi.org/10.25656/01:8019>
- Greiff, S., Fischer, A., Stadler, M., & Wüstenberg, S. (2015). Assessing complex problem-solving skills with multiple complex systems. *Thinking & Reasoning*, 21(3), 356–382. <https://doi.org/10.1080/13546783.2014.989263>
- Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M., & Martin, R. (2013). A multitrait–multimethod study of assessment instruments for complex problem solving. *Intelligence*, 41(5), 579–596. <https://doi.org/10.1016/j.intell.2013.07.012>
- Greiff, S., Molnár, G., Martin, R., Zimmermann, J., & Csapó, B. (2018). Students' exploration strategies in computer-simulated complex problem environments: A latent class

approach. *Computers & Education*, 126, 248–263.
<https://doi.org/10.1016/j.compedu.2018.07.013>

Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. *Computers in Human Behavior*, 61, 36–46. <https://doi.org/10.1016/j.chb.2016.02.095>

Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers & Education*, 91, 92–105.
<https://doi.org/10.1016/j.compedu.2015.10.018>

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts—Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, 105(2), 364–379. <https://doi.org/10.1037/a0031856>

Guo, X., Luo, Z., & Yu, X. (2020). A speed-accuracy tradeoff hierarchical model based on cognitive experiment. *Frontiers in Psychology*, 10, Article 2910.
<https://doi.org/10.3389/fpsyg.2019.02910>

Güre, Ö. B., Kayri, M., & Erdoğan, F. (2020). PISA 2015 matematik okuryazarlığını etkileyen faktörlerin eğitsel veri madenciliği ile çözümlenmesi. *Eğitim ve Bilim*, 45(202), 393–415. <https://doi.org/10.15390/EB.2020.8477>

Hamner, B., & Frasco, M. (2018). *Metrics: Evaluation metrics for machine learning* (R package version 0.1.4). <https://CRAN.R-project.org/package=Metrics>

Han, Y., & Wilson, M. (2022). Analyzing student response processes to evaluate success on a technology-based problem-solving task. *Applied Measurement in Education*, 35(1), 33–45. <https://doi.org/10.1080/08957347.2022.2034821>

- Han, Z., He, Q., & Von Davier, M. (2019). Predictive feature generation and selection using process data from PISA interactive problem-solving items: An application of random forests. *Frontiers in Psychology*, 10, Article 447759. <https://doi.org/10.3389/fpsyg.2019.02461>
- Hao, J., Smith, L., Mislevy, R., & Bauer, M. (2016). Taming log files from game/simulation-based assessments: Data models and data analysis tools. *ETS Research Report Series*, 2016(1), 1-17. <https://doi.org/10.1002/ets2.12096>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
- He, Q., Liao, D., & Jiao, H. (2019). Clustering behavioral patterns using process data in PIAAC problem-solving items. In B. P. Veldkamp & C. Sluijter (Eds.), *Theoretical and practical advances in computer-based educational measurement: Methodology of educational measurement and assessment* (pp. 189–212). https://doi.org/10.1007/978-3-030-18480-3_10
- Herde, C. N., Wüstenberg, S., & Greiff, S. (2016). Assessment of complex problem solving: What we know and what we don't know. *Applied Measurement in Education*, 29(4), 265–277. <https://doi.org/10.1080/08957347.2016.1209208>
- Hofner, B., Hothorn, T., Kneib, T., & Schmid, M. (2011). A framework for unbiased model selection based on boost. *Journal of Computational and Graphical Statistics*, 20(4), 956–971. <https://doi.org/10.1198/jcgs.2011.09220>
- Hothorn, T., & Bühlmann, P. (2006). Model-based boosting in high dimensions. *Bioinformatics*, 22(22), 2828–2829. <https://doi.org/10.1093/bioinformatics/btl462>
- Hothorn, T., & Zeileis, A. (2015). partykit: A modular toolkit for recursive partitioning in R. *The Journal of Machine Learning Research*, 16(1), 3905–3909.
- Hothorn, T., & Zeileis, A. (2023). A toolkit for recursive partitioning. *Package “partykit”*. <https://cran.r-project.org/web/packages/partykit/partykit.pdf>

- Hothorn, T., Bühlmann, P., Kneib, T., Schmid, M., & Hofner, B. (2010). Model-based boosting 2.0. *Journal of Machine Learning Research*, 11(71), 2109-2113. <http://jmlr.org/papers/v11/hothorn10a.html>
- Hothorn, T., Bühlmann, P., Kneib, T., Schmid, M., Hofner, B., Otto-Sobotka, F., Scheipl, F., & Mayr, A. (2024). *mboost: Model-based boosting* (R package version 2.9-11). <https://CRAN.R-project.org/package=mboost>
- Hothorn, T., Bühlmann, P., Kneib, T., Schmid, M., Hofner, B., Otto-Sobotka, F., Scheipl, F., & Mayr, A. (2024). Model-based boosting. <https://cran.r-project.org/web/packages/mboost/mboost.pdf>
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651–674. <https://doi.org/10.1198/106186006X133933>
- Hothorn, T., Hornik, K., & Zeileis, A. (2015). ctree: Conditional inference trees. *The Comprehensive R Archive Network*, 8.
- Hutter, F., Hoos, H., & Leyton-Brown, K. (2014, June 21–26). An efficient approach for assessing hyperparameter importance [Conference presentation]. In *Proceedings of the 31st International Conference on Machine Learning* (pp. 754-762). Beijing, China. <https://proceedings.mlr.press/v32/hutter14.pdf>
- Izrailev, S. (2024). *tictoc: Functions for timing R scripts, as well as implementations of "stack" and "stacklist" structures* (R package version 1.2.1). <https://CRAN.R-project.org/package=tictoc>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. Springer.
- Jiao, H., He, Q., & Veldkamp, B. P. (2021). Editorial: Process data in educational and psychological measurement. *Frontiers in Psychology*, 12, Article 793399. <https://doi.org/10.3389/fpsyg.2021.793399>

- Kerr, D., Chung, G. K. W. K., & Iseli, M. R. (2011). *The feasibility of using cluster analysis to examine log data from educational video games* (CRESST Report 790). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Klotzke, K., & Fox, J. (2019). Bayesian covariance structure modeling of responses and process data. *Frontiers in Psychology*, 10, Article 425160. <https://doi.org/10.3389/fpsyg.2019.01675>
- Korkmaz, B., Yiğit Meşe, D., & Arslan, H. T. (2023). *Ortaokul ve imam hatip ortaokulu 5. sınıf matematik ders kitabı*. Millî Eğitim Bakanlığı Yayınları.
- Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2), 151–162. <https://doi.org/10.32614/RJ-2014-031>
- Kroehne, U., Hahnel, C., & Goldhammer, F. (2019). Invariance of the response processes between gender and modes in an assessment of reading. *Frontiers in Applied Mathematics and Statistics*, 5, Article 422234. <https://doi.org/10.3389/fams.2019.00002>
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>
- Kuhn, M. (2019). *The caret package*. <https://topepo.github.io/caret/>
- Kuhn, M. (2023). *Classification and regression training*. <https://cran.r-project.org/web/packages/caret/caret.pdf>
- LeDell, E. (2018). *useR! Machine learning tutorial*. <https://koalaverse.github.io/machine-learning-in-R/index.html#resources>
- Lee, Y., Hao, J., Man, K., & Ou, L. (2019). How do test takers interact with simulation-based tasks? A response-time perspective. *Frontiers in Psychology*, 10, Article 422642. <https://doi.org/10.3389/fpsyg.2019.00906>

- Liao, D., He, Q., & Jiao, H. (2019). Mapping background variables with sequential patterns in problem-solving environments: An investigation of United States adults' employment status in PIAAC. *Frontiers in Psychology*, 10, Article 646. <https://doi.org/10.3389/fpsyg.2019.00646>
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomforest. *R News*, 2(3), 18–22.
- Liaw, A., & Wiener, M. (2022). *Breiman and Cutler's random forests for classification and regression*. <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
- Liu, H., Liu, Y., & Li, M. (2018). Analysis of process data of PISA 2012 computer-based problem solving: Application of the modified multilevel mixture IRT model. *Frontiers in Psychology*, 9, Article 359310. <https://doi.org/10.3389/fpsyg.2018.01372>
- Loh, W.Y. (2014a). Fifty years of classification and regression trees. *International Statistical Review/Revue Internationale de Statistique*, 82(3), 329–348. <https://doi.org/10.1111/insr.12016>
- Loh, W.Y. (2014b). Fifty years of classification and regression trees. *International Statistical Review/Revue Internationale de Statistique*, 82(3), 367–370. <https://doi.org/10.1111/insr.12057>
- Loh, Y. (2010). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 14–23. <https://doi.org/10.1002/widm.8>
- Ma, X. (2018). *Using classification and regression trees: A practical primer*. IAP.
- Man, K. (2024). Multimodal data fusion to detect preknowledge test-taking behavior using machine learning. *Educational and Psychological Measurement*, 84(4), 753–779. <https://doi.org/10.1177/00131644231193625>

- Mayr, A., Binder, H., Gefeller, O., & Schmid, M. (2014). The evolution of boosting algorithms—from machine learning to statistical modelling. *Methods of Information in Medicine*, 53(6), 419–427. <https://doi.org/10.3414/ME13-01-0122>
- Milborrow, S. (2024). *rpart.plot: Plot 'rpart' models: An enhanced version of 'plot.rpart'* (R package version 3.1.2). <https://CRAN.R-project.org/package=rpart.plot>
- Mingers, J. (1987). Expert systems – rule induction with statistical data. *Journal of the Operations Research Society*, 38(1), 39–47.
- Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity*, 9, 35–45. <https://doi.org/10.1016/j.tsc.2013.03.002>
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159–176. <https://doi.org/10.1002/j.2333-8504.1992.tb01436.x>
- Nicodemus, K. K., Malley, J. D., Strobl, C., & Ziegler, A. (2010). The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC Bioinformatics*, 11(1), 110–123. <https://doi.org/10.1186/1471-2105-11-110>
- Organisation for Economic Co-operation and Development (OECD). (2024). *PISA 2022 technical report*. OECD Publishing. <https://doi.org/10.1787/01820d6d-en>
- Organisation for Economic Co-operation and Development (OECD). (2014a). *PISA 2012 results: Creative problem solving: Students' skills in tackling real-life problems (Volume V)*. OECD Publishing. <https://doi.org/10.1787/9789264208070-en>
- Organisation for Economic Co-operation and Development (OECD). (2014b). *PISA 2012 technical report*. OECD Publishing.
- Organisation for Economic Co-operation and Development (OECD). (2014c). *PISA 2012 Results: What students know and can do – Student performance in mathematics*,

reading and science (Volume I, Revised edition, February 2014). OECD Publishing.
<https://doi.org/10.1787/9789264201118-en>

Organisation for Economic Co-operation and Development (OECD). (2009). *PISA 2009 Assessment framework: Key competencies in reading, mathematics and science.* OECD Publishing. <http://doi.org/10.1787/9789264062658-en>

Organisation for Economic Co-operation and Development (OECD). (2007). *PISA 2006: Science competencies for tomorrow's world.* OECD Publishing.
<http://doi.org/10.1787/9789264040014-en>

Organisation for Economic Co-operation and Development (OECD). (t.y.). *Codebook and instructions for merging files.*
<https://www.oecd.org/content/dam/oecd/en/data/datasets/pisa/pisa-2012-datasets/cba-pisa-2012/log-file-databases/Notes%20on%20log%20file%20data%20-%20releaseditems.pdf>

Özçelik, U. (2023). *Ortaokul ve imam hatip ortaokulu 6. sınıf matematik ders kitabı.* ATA Yayıncılık.

Pejić, A., & Molcer, P. S. (2021). Predictive machine learning approach for complex problem solving process data mining. *Acta Polytechnica Hungarica*, 18(1), 45–63.
<https://doi.org/10.12700/APH.18.1.2021.1.4>

Philipp, M., Rusch, T., Hornik, K., & Strobl, C. (2018). Measuring the stability of results from supervised statistical learning. *Journal of Computational and Graphical Statistics*, 27(4), 685–700. <https://doi.org/10.1080/10618600.2018.1473779>

Philipp, M., Zeileis, A., & Strobl, C. (2016). A toolkit for stability assessment of tree-based learners. In A. Colubi, A. Blanco, & A. Gatu (Eds.), *Proceedings of COMPSTAT 2016 – 22nd International Conference on Computational Statistics* (pp. 315–325).
<http://EconPapers.RePEc.org/RePEc:inn:wpaper:2016-11>

- Polikar, R. (2012). Ensemble learning. In C. Zhang & Y. Ma (Eds.), *Ensemble machine learning: Methods and applications* (pp. 1–34). Springer.
- Polya, G. (1945). *How to solve it*. Princeton University Press.
- Qiao, X., & Jiao, H. (2018). Data mining techniques in analyzing process data: A didactic. *Frontiers in Psychology*, 9, Article 374277. <https://doi.org/10.3389/fpsyg.2018.02231>
- R Core Team. (2024). *R: A language and environment for statistical computing*. <https://www.R-project.org/>
- Rabinowicz, A., & Rosset, S. (2022). Tree-based models for correlated data. *The Journal of Machine Learning Research*, 23(1), 11802–11832.
- Redelmeier, A., Jullum, M., & Aas, K. (2020). Explaining predictive models with mixed features using Shapley values and conditional inference trees. In A. Holzinger, P. Kieseberg, A. Tjoa, & E. Weippl (Eds.), *Machine learning and knowledge extraction: 4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, August 25–28, 2020, Proceedings* (pp. 117–137). Springer International Publishing. https://doi.org/10.1007/978-3-030-57321-8_7
- Reis Costa, D., Bolsinova, M., Tijmstra, J., & Andersson, B. (2021). Improving the precision of ability estimates using time-on-task variables: Insights from the PISA 2012 computer-based assessment of mathematics. *Frontiers in Psychology*, 12, Article 579128. <https://doi.org/10.3389/fpsyg.2021.579128>
- Ren, Y., Luo, F., Ren, P., Bai, D., Li, X., & Liu, H. (2019). Exploring multiple goals balancing in complex problem solving based on log data. *Frontiers in Psychology*, 10, Article 445854. <https://doi.org/10.3389/fpsyg.2019.01975>

- Revelle, W. (2024). *psych: Procedures for psychological, psychometric, and personality research*. Northwestern University, Evanston, Illinois (R package version 2.4.3). <https://CRAN.R-project.org/package=psych>
- Ridgeway, G. (2024). *Generalized boosted models: A guide to the gbm package*. R package. <https://cran.r-project.org/web/packages/gbm/vignettes/gbm.pdf>
- Ridgeway, G., Edwards, D., Kriegler, B., Schroedl, S., Southworth, H., Greenwell, B., Boehmke, B., Cunningham, J., & Developers, G. (2024). *Generalized boosted regression models* (R package version 2.2.2). <https://cran.r-project.org/web/packages/gbm/index.html>
- Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1–25. <https://doi.org/10.18637/jss.v017.i05>
- Robitzsch, A., Kiefer, T., & Wu, M. (2024). *TAM: Test analysis modules* (R package version 4.2-21). <https://CRAN.R-project.org/package=TAM>
- Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601–618. <https://doi.org/10.1109/TSMCC.2010.2053532>
- Saidani, O., Umer, M., Alshardan, Alshardan, A., Alturki, N., Nappi, M. & Ashraf, I. (2024). Student academic success prediction in multimedia-supported virtual learning system using ensemble learning approach. *Multimedia Tools and Application*, 1-26. <https://doi.org/10.1007/s11042-024-18669-z>
- Salis, K. L., & Kliem, S. (2014). Conditional inference trees: A method for predicting intimate partner violence. *Journal of Marital and Family Therapy*, 40(4), 430–441. <https://doi.org/10.1111/jmft.12089>

- Scalise, K., & Clarke-Midura, J. (2018). The many faces of scientific inquiry: Effectively measuring what students do and not only what they say. *Journal of Research in Science Teaching*, 55(10), 1469–1496. <https://doi.org/10.1002/tea.21464>
- Scherer, R., & Gustafsson, J. E. (2015). The relations among openness, perseverance, and performance in creative problem solving: A substantive-methodological approach. *Thinking Skills and Creativity*, 18, 4–17. <https://doi.org/10.1016/j.tsc.2015.04.004>
- Scherer, R., Greiff, S., & Hautamäki, J. (2015). Exploring the relation between time on task and ability in complex problem solving. *Intelligence*, 48, 37–50. <https://doi.org/10.1016/j.intell.2014.10.003>
- Schivinski, B. (2021). Eliciting brand-related social media engagement: A conditional inference tree framework. *Journal of Business Research*, 130, 594–602. <https://doi.org/10.1016/j.jbusres.2019.08.045>
- Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., & Crowley, J. (2024). *GGally: Extension to 'ggplot2'* (R package version 2.2.1). <https://CRAN.R-project.org/package=GGally>
- Schneider, L., Zeileis, A., & Strobl, C. (2023). *Descriptive and graphical analysis of the variable and cutpoint selection inside random forests*. The Comprehensive R Archive Network. <https://cran.r-project.org/web/packages/stablelearner/vignettes/forests.html>
- Sinharay, S. (2016). An NCME instructional module on data mining methods for classification and regression. *Educational Measurement: Issues and Practice*, 35(3), 38–54. <https://doi.org/10.1111/emip.12115>
- Sinharay, S., Zhang, M., & Deane, P. (2019). Prediction of essay scores from writing process and product features using data mining methods. *Applied Measurement in Education*, 32(2), 116–137. <https://doi.org/10.1080/08957347.2019.1577245>

- Sonnleitner, P., Keller, U., Martin, R., & Brunner, M. (2013). Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence*, 41(5), 289–305. <https://doi.org/10.1016/j.intell.2013.05.002>
- Stadler, M., Fischer, F., & Greiff, S. (2019). Taking a closer look: An exploratory analysis of successful and unsuccessful strategy use in complex problems. *Frontiers in Psychology*, 10, Article 424920. <https://doi.org/10.3389/fpsyg.2019.00777>
- Steinberg, D., & Colla, P. (1995). *CART: Tree-structured nonparametric data analysis*. Salford Systems.
- Stekhoven, D. J. (2022). *missForest: Nonparametric missing value imputation using random forest* (R package version 1.5). <https://cran.r-project.org/web/packages/missForest/missForest.pdf>
- Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118. <https://doi.org/10.1093/bioinformatics/btr597>
- Strobl, C. (2013). *Data mining*. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods in psychology: Vol. 2: Statistical analysis* (pp. 678–700). OUP USA.
- Strobl, C., & Zeileis, A. (2008). *Danger: High power! – Exploring the statistical properties of a test for random forest variable importance*. Department of Statistics, Technical Reports, No. 17. <https://doi.org/10.5282/ubm/epub.2111>
- Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1), 307–318. <https://doi.org/10.1186/1471-2105-9-307>
- Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 1–21. <https://doi.org/10.1186/1471-2105-8-25>

- Strobl, C., Hothorn, T., & Zeileis, A. (2009). *Party on! A new, conditional variable importance measure for random forests available in the party package*. University of Munich. <http://epub.ub.uni-muenchen.de/9387/1/techreport.pdf>
- Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14(4), 323–348. <https://doi.org/10.1037/a0016973>
- Tang, X., Wang, Z., He, Q., Liu, J., & Ying, Z. (2020). Latent feature extraction for process data via multidimensional scaling. *Psychometrika*, 85(2), 378–397. <https://doi.org/10.1007/s11336-020-09708-3>
- Tang, X., Zhang, S., Wang, Z., Liu, J., & Ying, Z. (2021). Procdata: An R package for process data analysis. *Psychometrika*, 86, 1058–1083. <https://doi.org/10.1007/s11336-021-09798-7>
- Teig, N., Scherer, R., & Kjærnsli, M. (2020). Identifying patterns of students' performance on simulated inquiry tasks using PISA 2015 log-file data. *Journal of Research in Science Teaching*, 57(9), 1400–1429. <https://doi.org/10.1002/tea.21657>
- Therneau, T. M., & Atkinson, E. J. (2023a). *An introduction to recursive partitioning using the RPART routines*. Mayo Foundation: Technical report.
- Therneau, T., & Atkinson, B. (2023b). *rpart: Recursive partitioning and regression trees* (R package version 4.1.23). <https://CRAN.R-project.org/package=rpart>
- Tierney, N., & Cook, D. (2023). Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations. *Journal of Statistical Software*, 105(7), 1–31. <https://doi.org/10.18637/jss.v105.i07>
- Toton, S. L., & Maynes, D. D. (2019). Detecting examinees with pre-knowledge in experimental data using conditional scaling of response times. *Frontiers in Education*, 4, Article 422727. <https://doi.org/10.3389/feduc.2019.00049>

- Tschargi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, 51(1), 1–10. <https://doi.org/10.2307/1129583>
- Valdiviezo, H. C., & Van Aelst, S. (2015). Tree-based prediction on incomplete data using imputation or surrogate decisions. *Information Sciences*, 311, 163–181. <https://doi.org/10.1016/j.ins.2015.03.018>
- Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science*, 20(1), 75–100. https://doi.org/10.1207/s15516709cog2001_3
- Wang, C., Weiss, D. J., & Su, S. (2019). Modeling response time and responses in multidimensional health measurement. *Frontiers in Psychology*, 10, Article 420471. <https://doi.org/10.3389/fpsyg.2019.00051>
- West, R. M. (2012). Regression and classification trees. In Y. K. Tu & D. C. Greenwood (Eds.), *Modern methods for epidemiology* (pp. 279–290). Springer Science & Business Media.
- Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12), 1–20. <https://doi.org/10.18637/jss.v021.i12>
- Wickham, H. (2023). *stringr: Simple, consistent wrappers for common string operations* (R package version 1.5.1). <https://CRAN.R-project.org/package=stringr>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A grammar of data manipulation* (R package version 1.1.4). <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., Miller, E., & Smith, D. (2023). *haven: Import and export 'SPSS', 'Stata' and 'SAS' Files* (R package version 2.5.4). <https://CRAN.R-project.org/package=haven>
- Wickham, H., Vaughan, D., & Girlich, M. (2024). *tidyr: Tidy messy data* (R package version 1.3.1). <https://CRAN.R-project.org/package=tidyr>

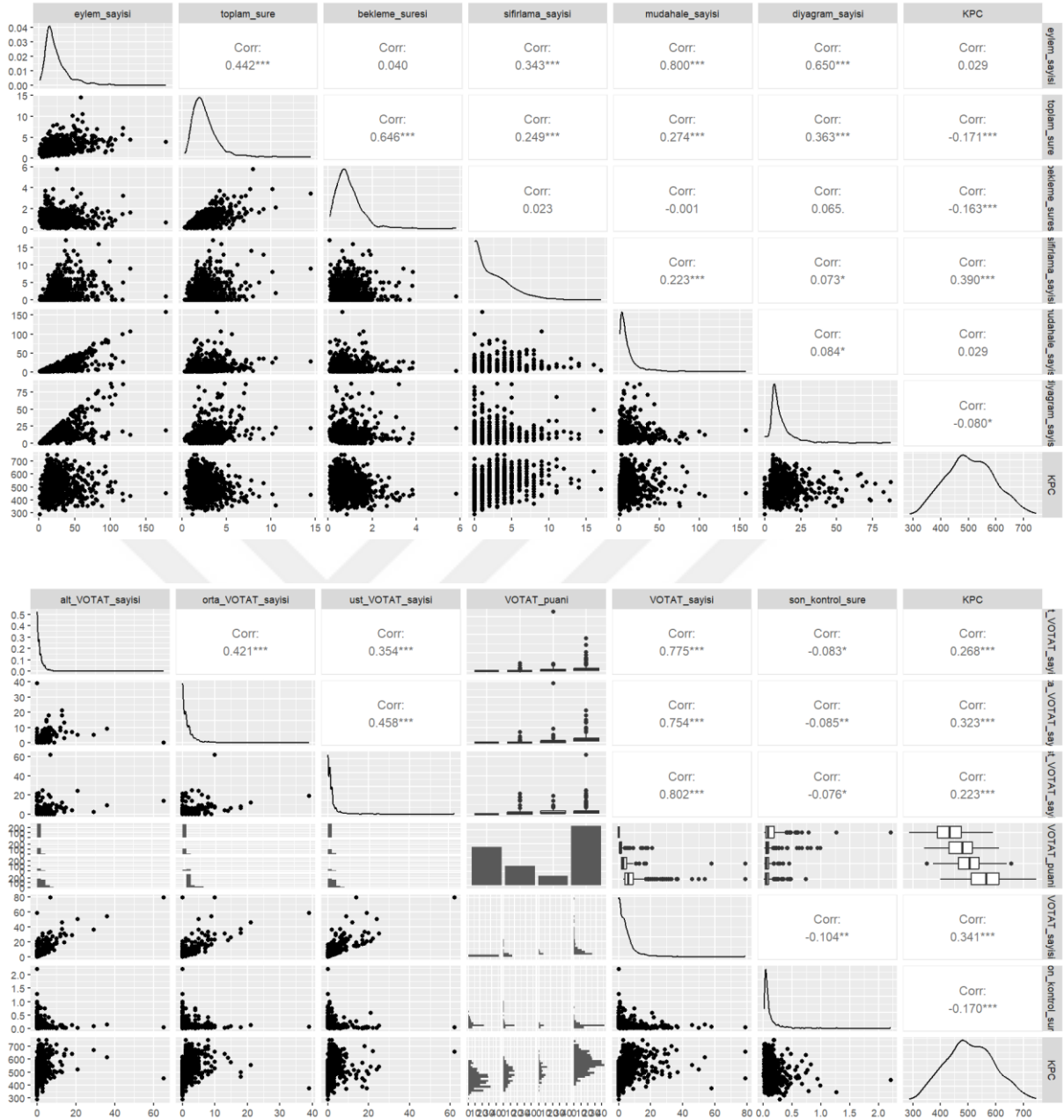
- Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving—More than reasoning? *Intelligence*, 40(1), 1–14. <https://doi.org/10.1016/j.intell.2011.11.003>
- Wüstenberg, S., Stadler, M., Hautamäki, J., & Greiff, S. (2014). The role of strategy knowledge for the application of strategies in complex problem solving tasks. *Technology, Knowledge and Learning*, 19(1–2), 127–146. <https://doi.org/10.1007/s10758-014-9222-8>
- Xu, H., Fang, G., Chen, Y., Liu, J., & Ying, Z. (2018). Latent class analysis of recurrent events in problem-solving items. *Applied Psychological Measurement*, 42(6), 478–498. <https://doi.org/10.1177/0146621617748325>
- Xu, X., Zhang, S., Guo, J., & Xin, T. (2024). Biclustering of log data: Insights from a computer-based complex problem-solving assessment. *Journal of Intelligence*, 12(1), 10. <https://doi.org/10.3390/jintelligence12010010>
- Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, 295–316. <https://doi.org/10.1016/j.neucom.2020.07.061>
- Yavuz, H. Ç. (2019). The effects of log data on students' performance. *Journal of Measurement and Evaluation in Education and Psychology*, 10(4), 378–390. <https://doi.org/10.21031/epod.564232>
- Yohannes, Y., & Webb, P. (1999). *Classification and regression trees, CART: A user manual for identifying indicators of vulnerability to famine and chronic food insecurity*. International Food Policy Research Institute.
- Yu, C. H., Kaprolet, C., Jannasch-Pennell, A., & DiGangi, S. (2012). A data mining approach to comparing American and Canadian grade 10 students' PISA science test performance. *Journal of Data Science*, 10(24), 441–464. [https://doi.org/10.6339/JDS.201207_10\(3\).0006](https://doi.org/10.6339/JDS.201207_10(3).0006)

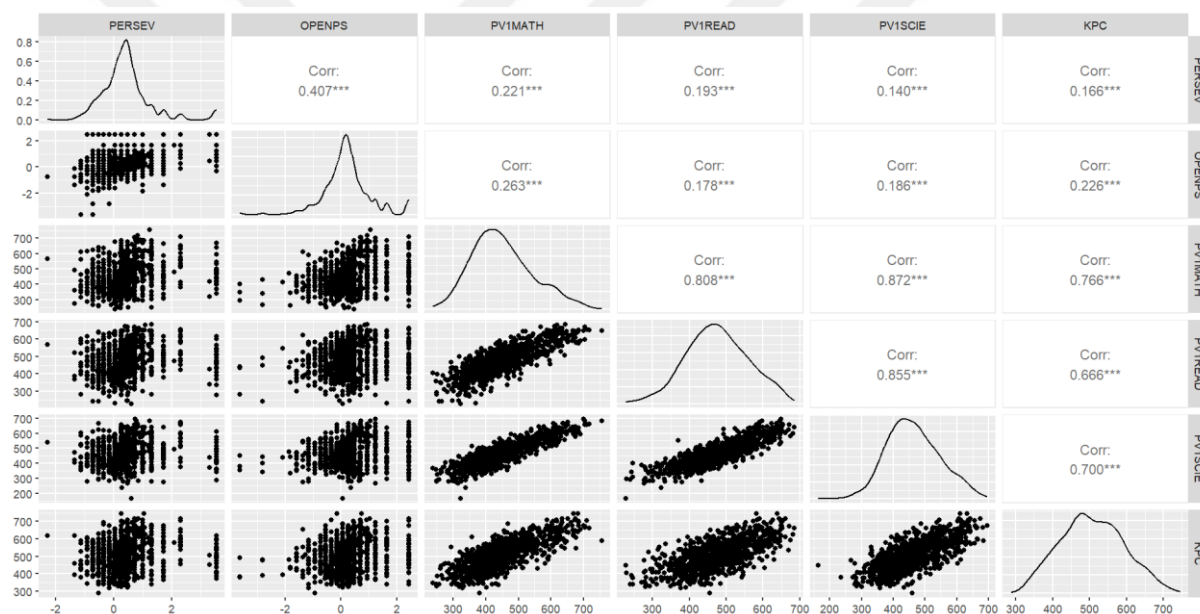
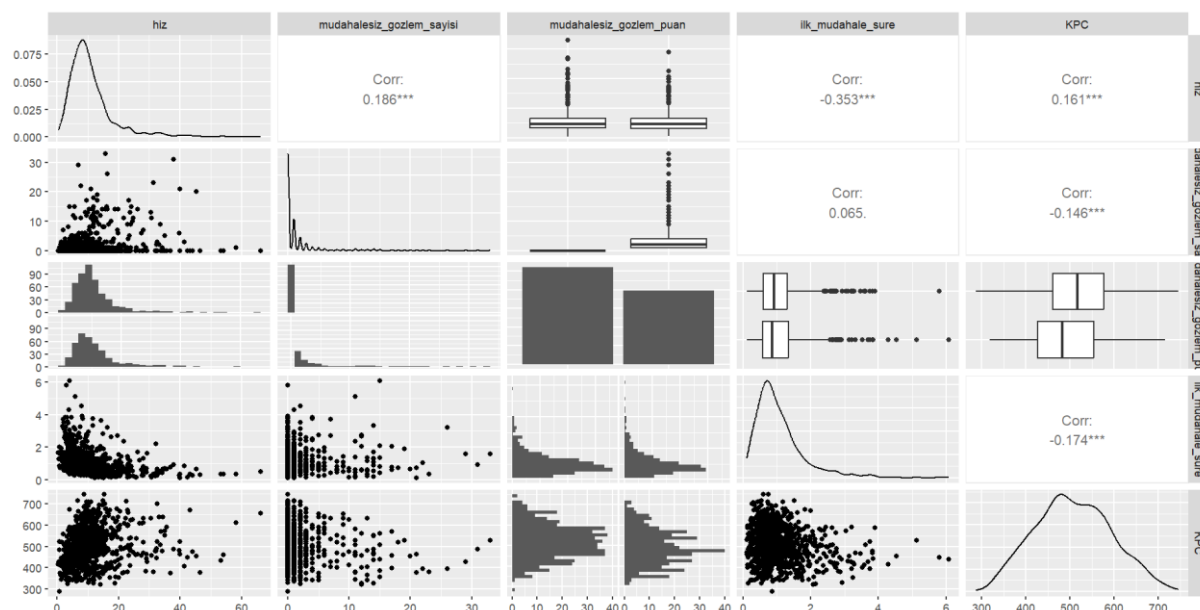
- Zhan, P., & Qiao, X. (2022). Diagnostic classification analysis of problem-solving competence using process data: An item expansion method. *Psychometrika*, 87(4), 1529–1547. <https://doi.org/10.1007/s11336-022-09855-9>
- Zhan, P., Jiao, H., Man, K., Wang, W. C., & He, K. (2021). Variable speed across dimensions of ability in the joint model for responses and response times. *Frontiers in Psychology*, 12, Article 469196. <https://doi.org/10.3389/fpsyg.2021.469196>
- Zhang, M., & Deane, P. (2015). *Process features in writing: Internal structure and incremental value over product features (Research Report No. RR-15-27)*. Educational Testing Service. <https://doi.org/10.1002/ets2.12075>
- Zhang, S., Wang, Z., Qi, J., Liu, J., & Ying, Z. (2023). Accurate assessment via process data. *Psychometrika*, 88(1), 76–97. <https://doi.org/10.1007/s11336-022-09880-8>
- Zhen, Y., & Zhu, X. (2024). An ensemble learning approach based on TabNet and machine learning models for cheating detection in educational tests. *Educational and Psychological Measurement*, 84(4), 780–809. <https://doi.org/10.1177/00131644231191298>
- Zhou, Z. H. (2009). Ensemble learning. In S. Z. Li & A. K. Jain (Eds.), *Encyclopedia of Biometrics* (pp. 270–273). <https://doi.org/10.1007/978-0-387-73003-5>

EK-A: Değişkenlere İlişkin Betimsel İstatistikler

	Ortalama	Standart Sapma	Medyan	Minimum	Maksimum	Açıklık	Çarpıklık	Basıklık	Standart Hata
Eylem sayısı	25.026	18.713	19.000	1.000	178.000	177.000	2.399	9.022	0.619
Toplam Süre	2.521	1.374	2.250	0.298	14.503	14.205	2.164	9.866	0.045
Bekleme Süresi	0.927	0.580	0.822	0.070	5.807	5.737	1.938	8.225	0.019
Sıfırlama Sayısı	2.247	2.626	1.000	0.000	17.000	17.000	1.564	3.245	0.087
Müdahale Sayısı	10.913	13.468	7.000	1.000	158.000	157.000	3.933	24.836	0.445
Diyagram Sayısı	11.867	10.832	9.000	0.000	88.000	88.000	3.325	14.647	0.358
Alt VOTAT Sayısı	1.477	3.342	1.000	0.000	65.000	65.000	10.203	160.669	0.110
Orta VOTAT Sayısı	1.343	2.458	1.000	0.000	39.000	39.000	6.271	70.572	0.081
Üst VOTAT Sayısı	1.737	3.505	1.000	0.000	62.000	62.000	7.876	104.188	0.116
VOTAT Puanı	1.706	1.322	2.000	0.000	3.000	3.000	-0.236	-1.720	0.044
VOTAT Sayısı	4.556	7.254	3.000	0.000	79.000	79.000	4.646	33.242	0.240
Son Kontrol Süresi	0.112	0.142	0.068	0.017	2.213	2.196	5.932	61.763	0.005
Hız	10.955	7.619	9.191	0.466	66.192	65.726	2.481	9.260	0.252
Müdahalesiz Gözlem Sayısı	1.694	3.854	0.000	0.000	33.000	33.000	3.877	18.369	0.127
Müdahalesiz Gözlem Puanı	0.431	0.495	0.000	0.000	1.000	1.000	0.280	-1.924	0.016
İlk Müdahale Süresi	1.074	0.740	0.887	0.110	6.087	5.977	2.095	6.938	0.024
KPÇ	507.449	86.111	503.399	287.532	745.456	457.924	0.129	-0.512	2.847
Azim	0.427	0.853	0.355	-2.288	3.529	5.817	1.439	3.762	0.028
Problem Çözmeye Açıklık	0.216	0.829	0.178	-3.633	2.446	6.080	0.130	2.628	0.027
Matematik Puanı	450.725	92.600	439.640	238.518	755.032	516.513	0.505	-0.096	3.061
Okuduğunu Anlama Puanı	476.046	84.816	474.922	227.173	685.652	458.480	-0.014	-0.248	2.804
Fen Puanı	463.826	82.369	458.542	163.875	696.046	532.171	0.229	-0.185	2.723

EK-B: Değişkenlere İlişkin Dağılım Grafikleri ve Korelasyon Değerleri



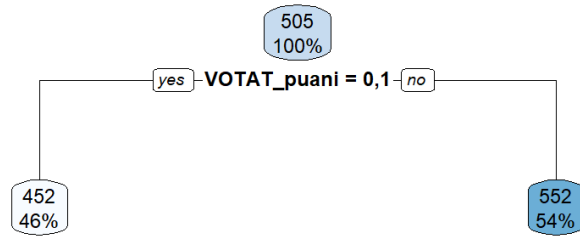


EK-C: Madde Parametre Kestirimlerine İlişkin Değerler

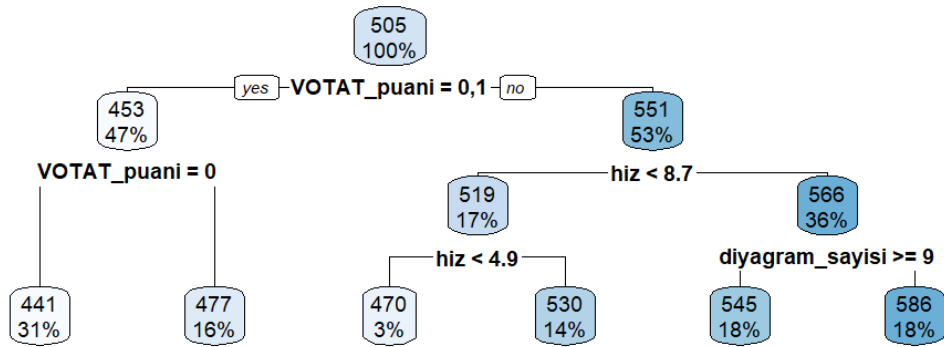
Madde	Kategori aralığı	Eğim	Konum	1. Eşik	2. Eşik	RMSE
CP014Q01	0-2	0,777	0,434	-2,761	2,761	0.04
CP014Q02	0-1	0,752	3,151	-	-	0.00
CP014Q06	0-1	1,583	0,831	-	-	-0.01
CP025Q01	0-2	0,901	-0,104	1,163	-1,163	0.01
CP025Q02	0-2	0,944	1,655	0,118	-0,118	0.01
CP027Q01T	0-2	0,359	0,008	-1,175	1,175	0.02
CP027Q02	0-2	0,564	3,841	-0,958	0,958	0.01
CP028Q01	0-1	0,688	-0,058	-	-	0.04
CP028Q02	0-2	1,009	-0,098	-0,130	0,130	0.01
CP029Q01	0-2	0,990	-0,174	0,702	-0,702	0.02
CP029Q02	0-2	0,944	0,404	0,166	-0,166	0.03
CP032Q01	0-1	1,858	1,059	-	-	-0.01
CP032Q02T	0-2	1,567	-0,846	-1,035	1,035	0.03
CP032Q04	0-1	0,330	0,924	-	-	0.04
CP034Q01T	0-1	1,229	0,978	-	-	0.01
CP034Q02	0-1	1,432	-1,911	-	-	0.02
CP034Q05	0-1	1,128	-0,947	-	-	-0.01
CP036Q01	0-1	1,286	0,228	-	-	-0.01
CP036Q03	0-1	1,553	0,641	-	-	-0.01
CP037Q01	0-1	1,024	-0,374	-	-	-0.01
CP037Q02	0-1	0,917	0,855	-	-	0.01
CP037Q03	0-1	1,503	1,825	-	-	0.01
CP038Q01	0-2	1,087	0,822	-1,127	1,127	-0.01
CP038Q02	0-1	1,599	0,898			0.01
CP038Q03	0-1	1,317	1,008			-0.01
CP041Q02	0-2	0,645	2,583	2,204	-2,204	0.01
CP041Q03	0-2	0,728	-1,705	-2,502	2,502	0.01

EK-D: Süreç Verisinden Elde Edilen Regresyon Ağaçları

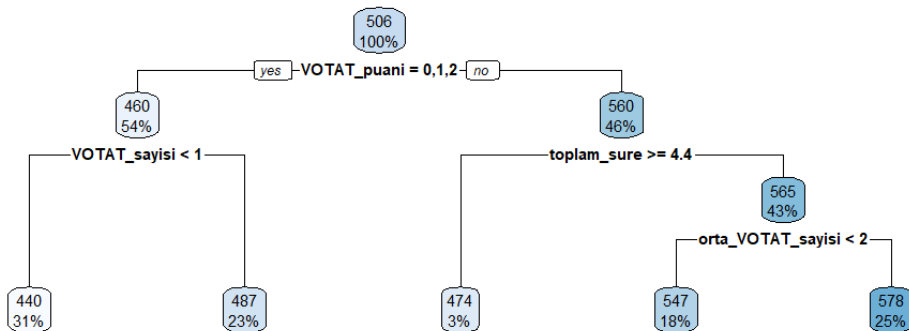
%50 Eğitim Verisi-%50 Test Verisi



%60 Eğitim Verisi-%40 Test Verisi

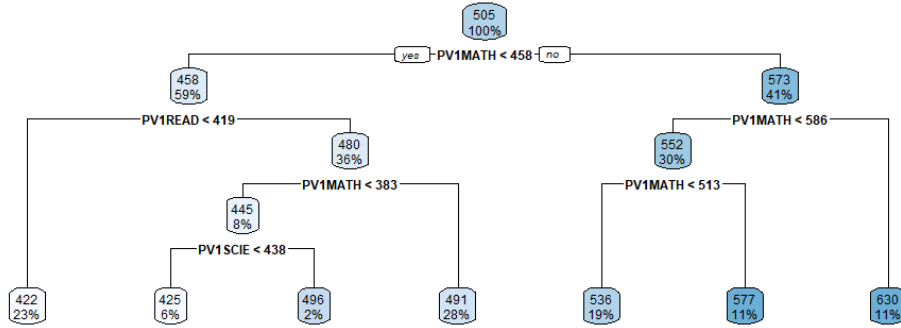


%70 Eğitim Verisi-%30 Test Verisi

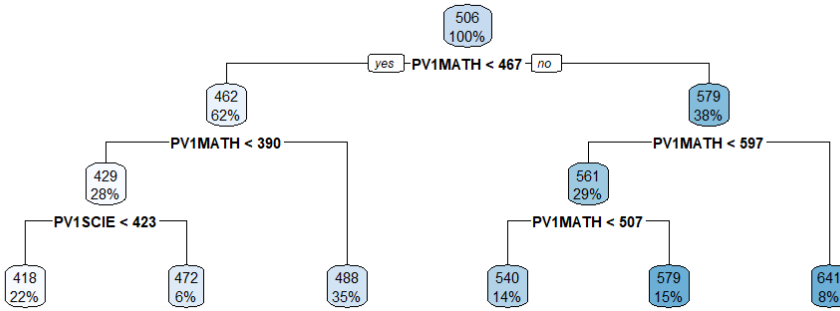


EK-E: Sonuç Verisinden Elde Edilen Regresyon Ağaçları

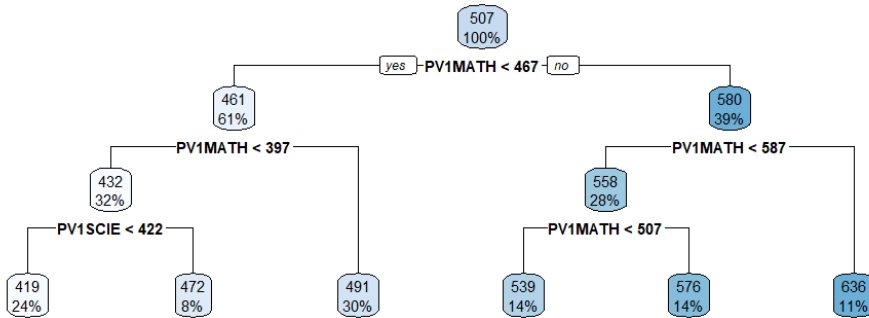
%50 Eğitim Verisi-%50 Test Verisi



%70 Eğitim Verisi-%30 Test Verisi

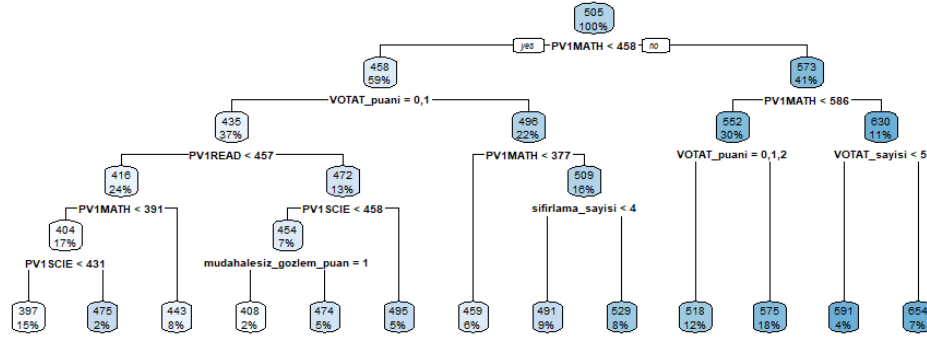


%80 Eğitim Verisi-%20 Test Verisi

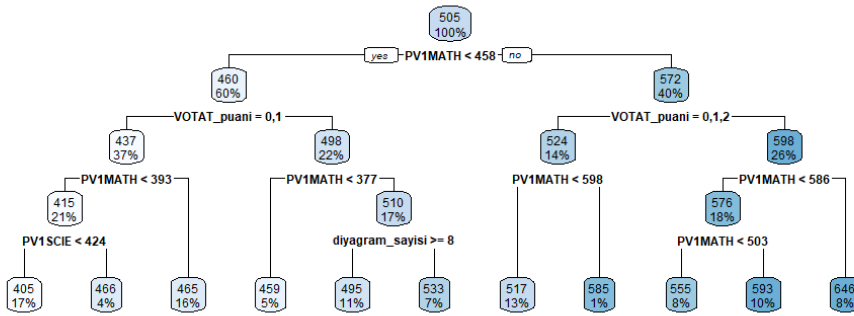


EK-F: Süreç+Sonuç Verisinden Elde Edilen Regresyon Ağaçları

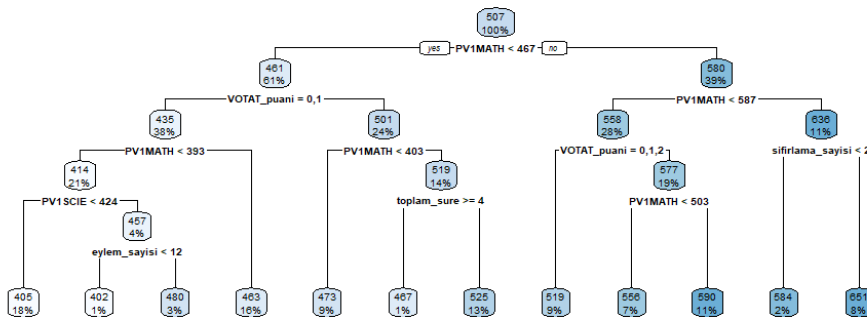
%50 Eğitim Verisi-%50 Test Verisi



%60 Eğitim Verisi-%40 Test Verisi

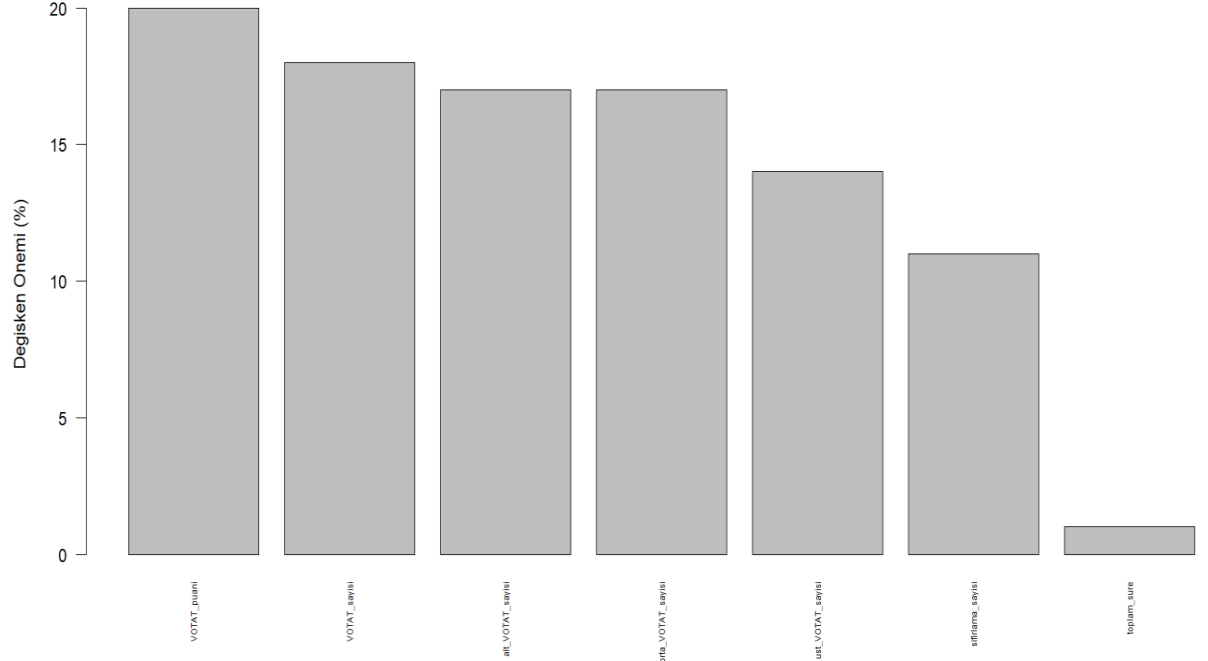


%80 Eğitim Verisi-%20 Test Verisi

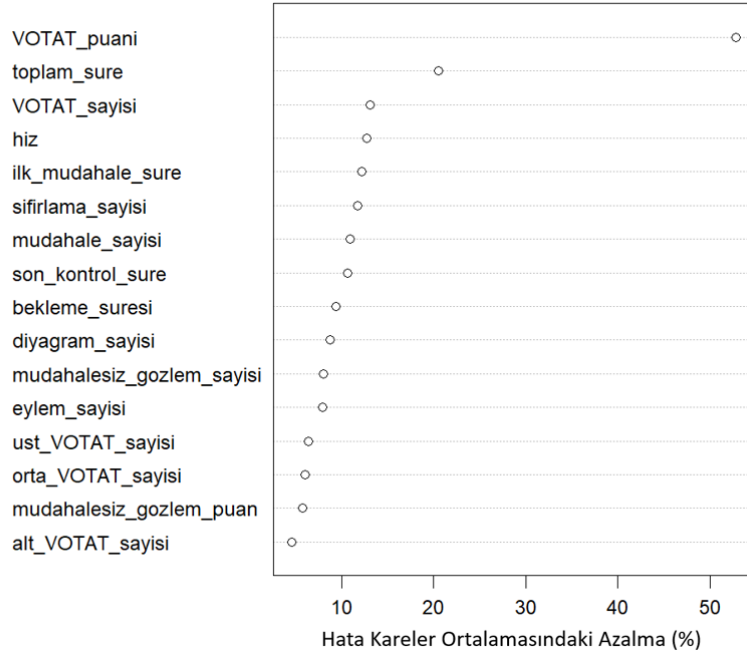


EK-G: Süreç Verisi-Değişken Önem Grafikleri

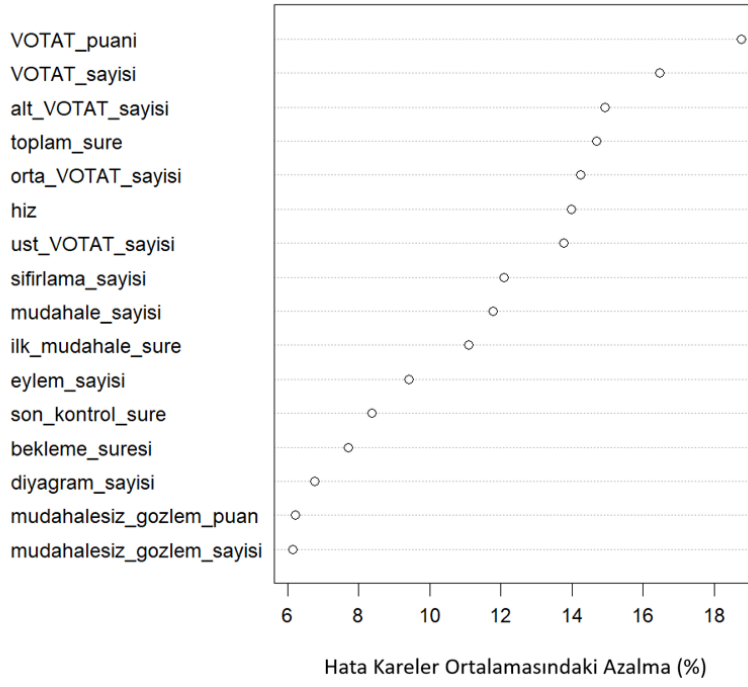
Regresyon Ağacı Değişken Önem Grafiği



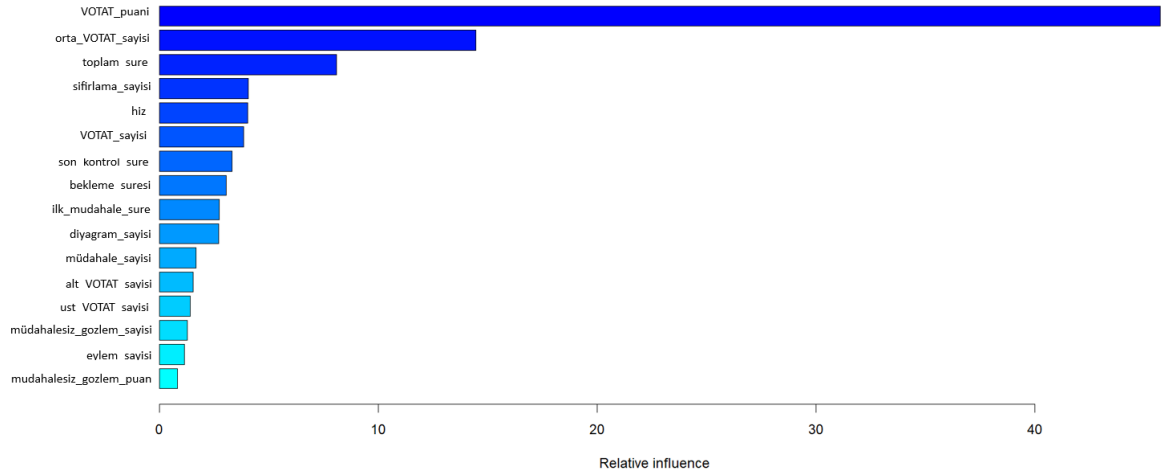
Torbalama Değişken Önem Grafiği



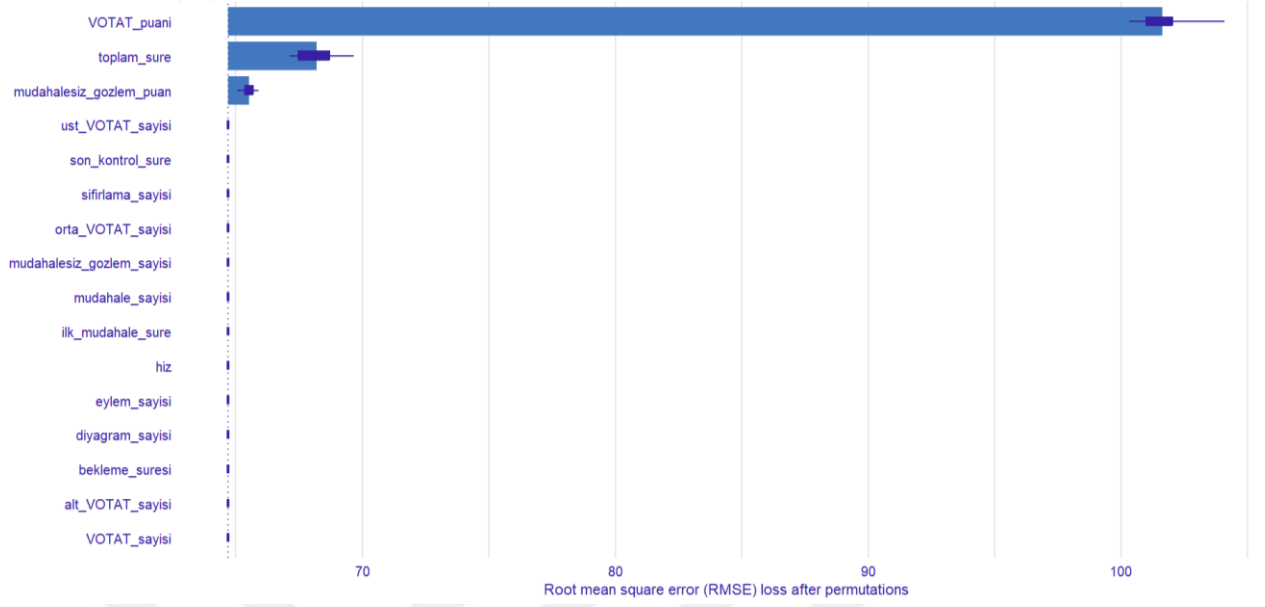
Rastgele Orman (Ölçeklenmiş/Varsayılan) Değişken Önem Grafiği



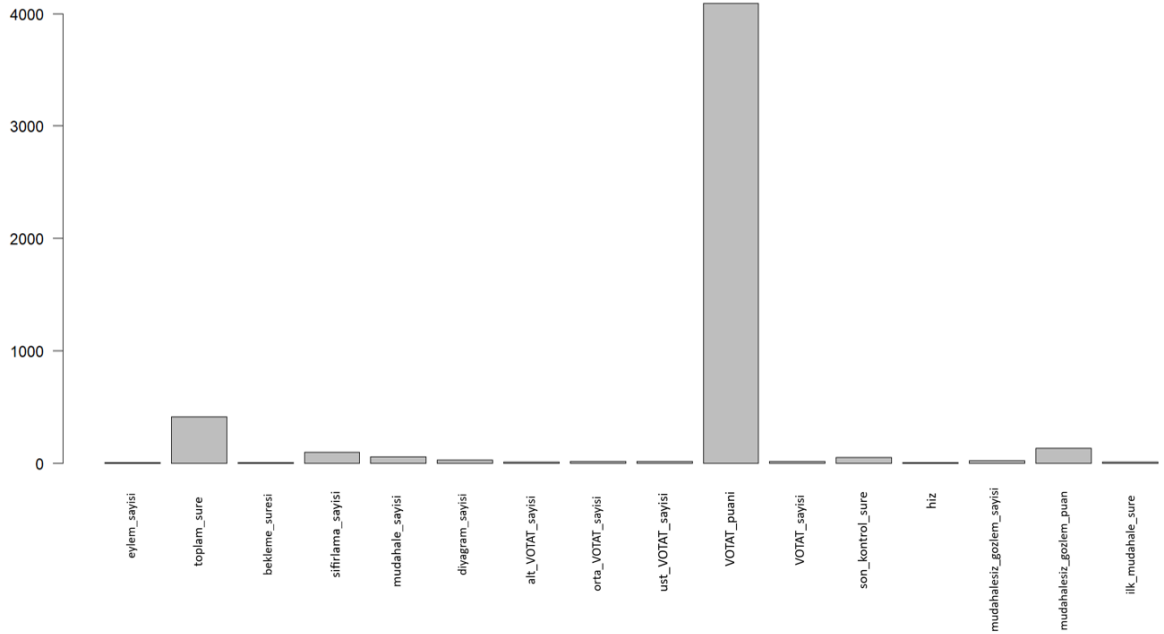
Gradyan Artırma Değişken Önem Grafiği



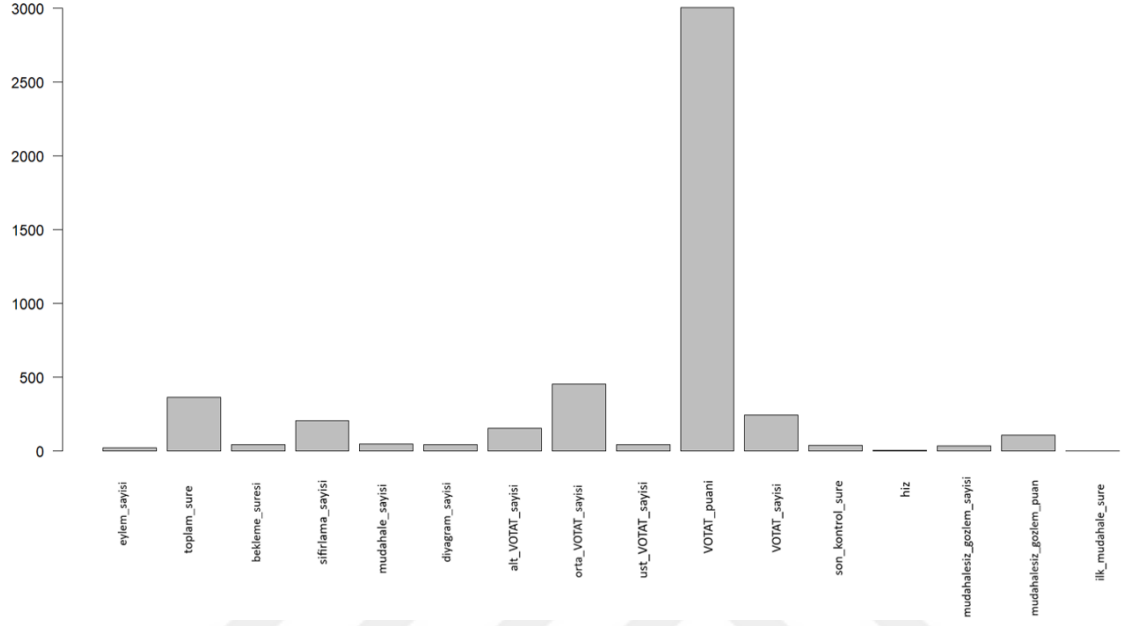
Koşullu Çıkarım Ağacı Değişken Önem Grafiği



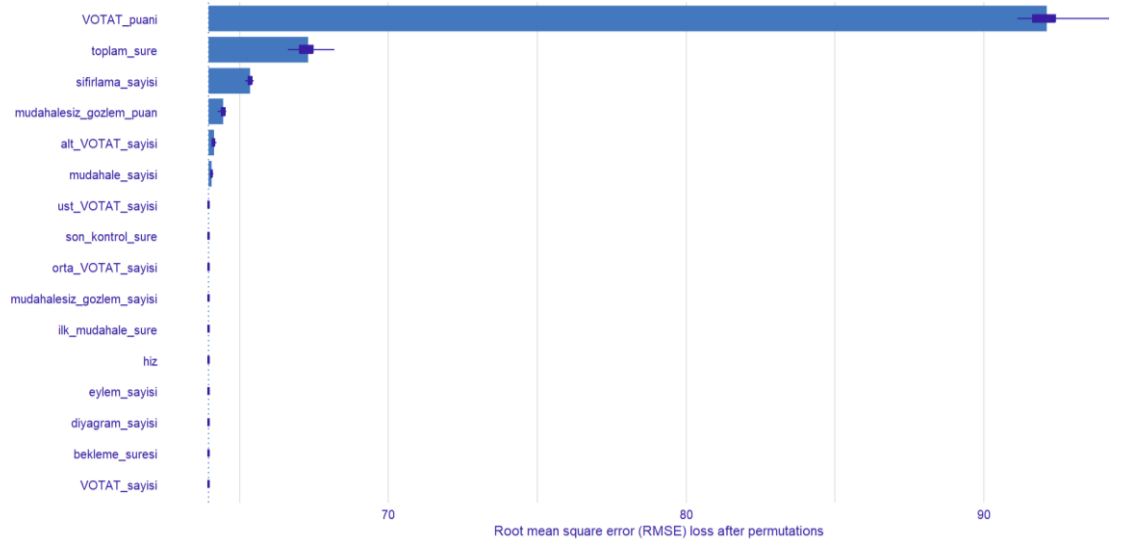
Koşullu Çıkarım Ağaçlarına Dayalı Torbalama Değişken Önem Grafiği



Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman (Koşullu Olmayan Versiyon/Varsayılan) Değişken Önem Grafiği

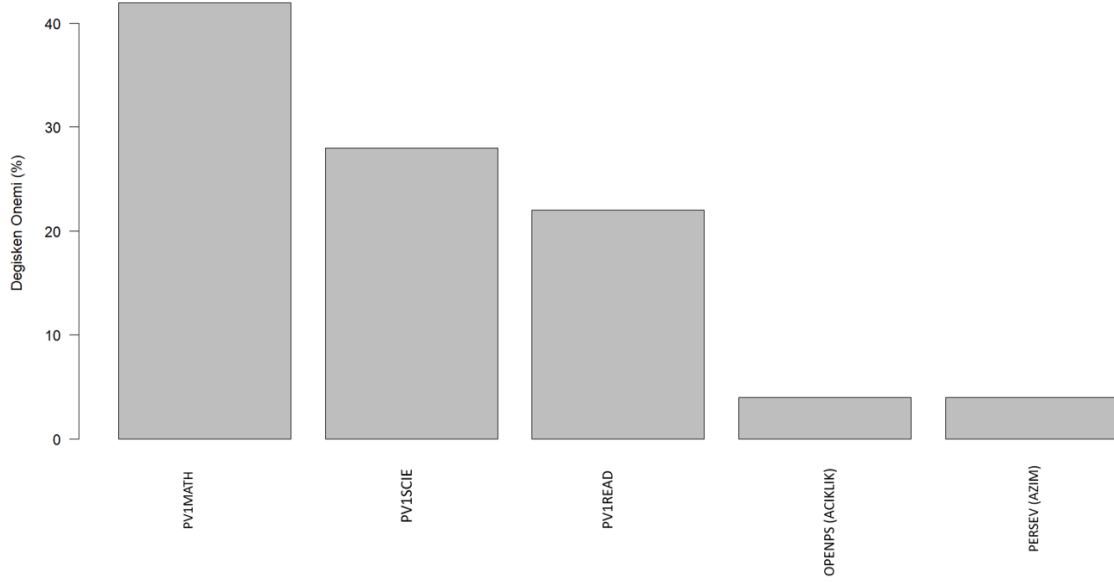


Koşullu Çıkarım Ağaçlarına Dayalı Gradyan Artırma Değişken Önem Grafiği

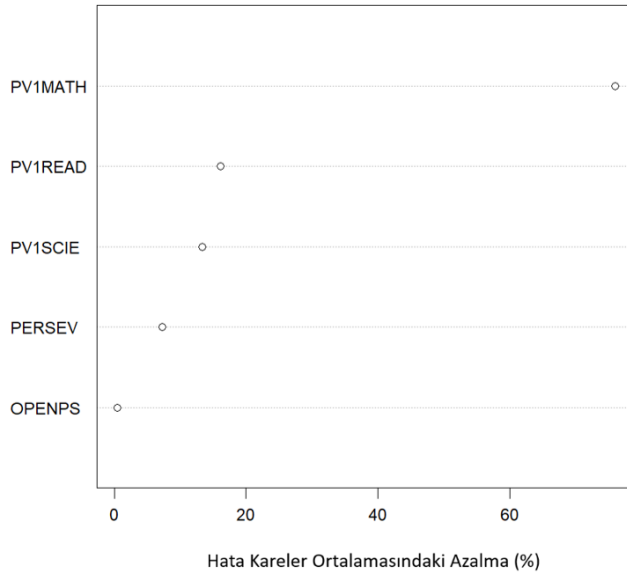


EK-H: Sonuç Verisi-Değişken Önem Grafikleri

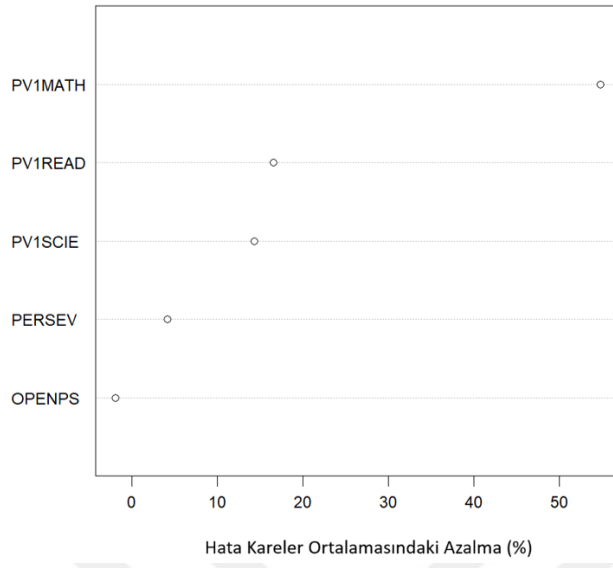
Regresyon Ağacı Değişken Önem Grafiği



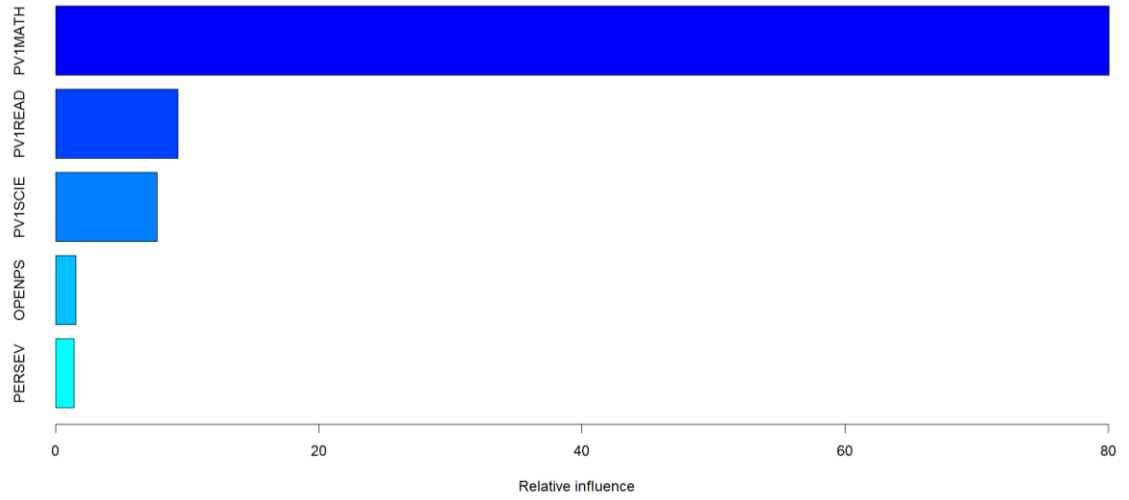
Torbalama Değişken Önem Grafiği



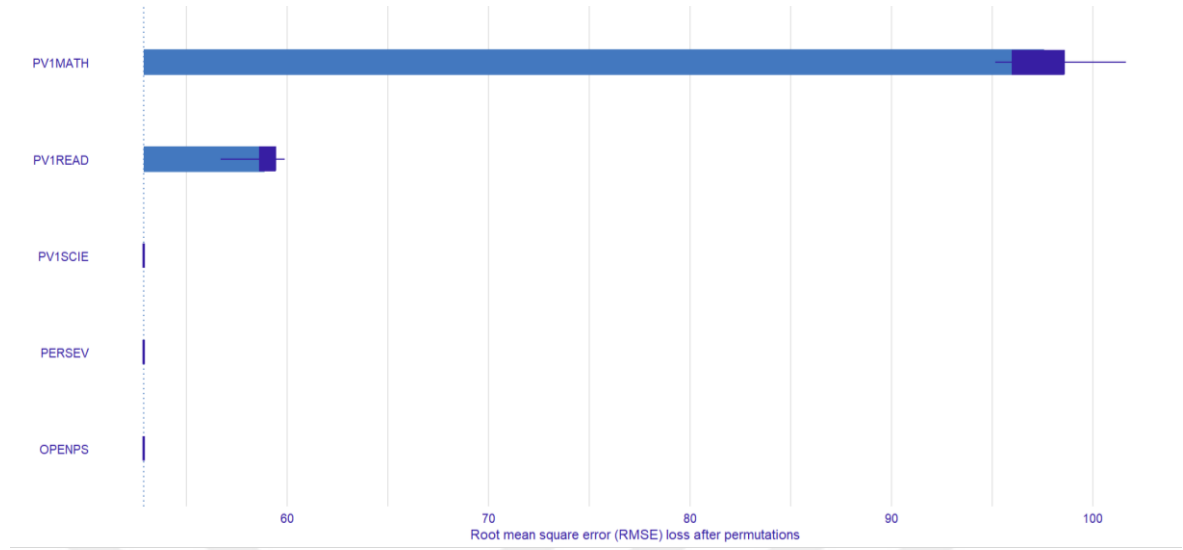
Rastgele Orman (Ölçeklenmiş/Varsayılan) Değişken Önem Grafiği



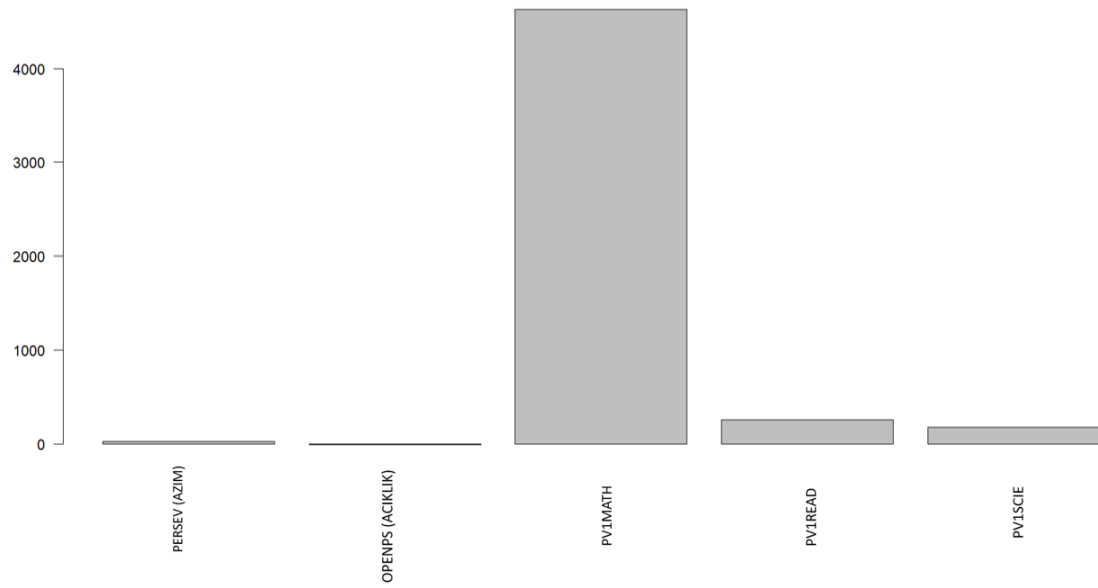
Gradyan Artırma Değişken Önem Grafiği



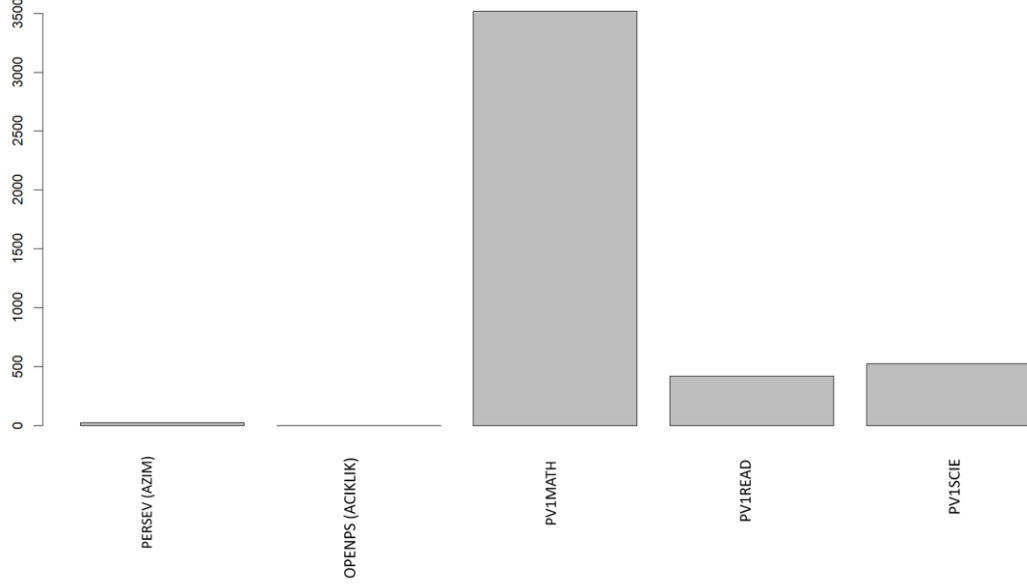
Koşullu Çıkarım Ağacı Değişken Önem Grafiği



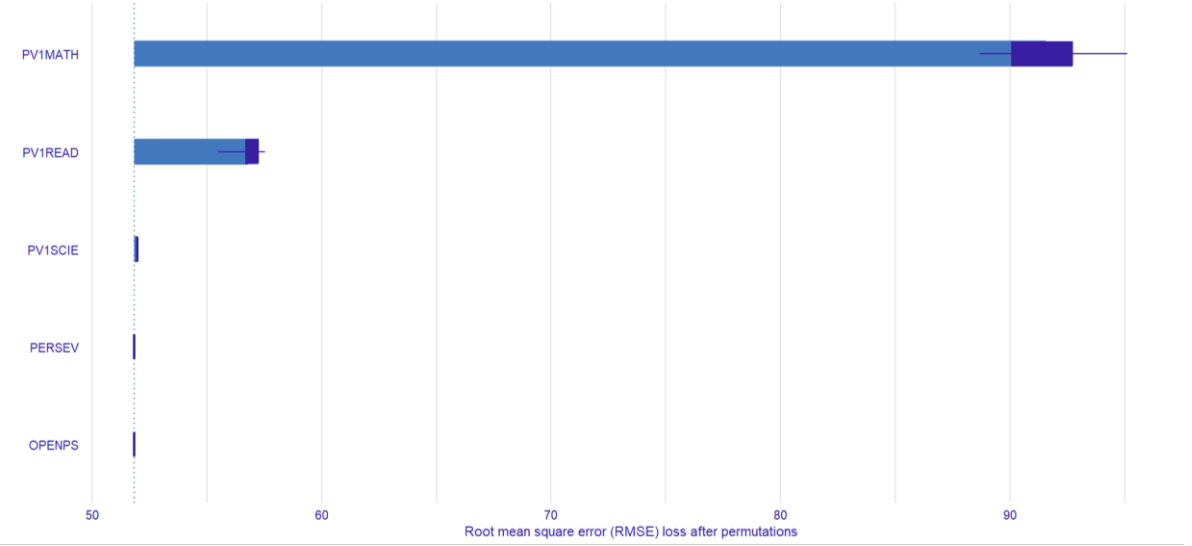
Koşullu Çıkarım Ağaçlarına Dayalı Torbalama Değişken Önem Grafiği



Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman (Koşullu Olmayan Versiyon/Varsayılan) Değişken Önem Grafiği

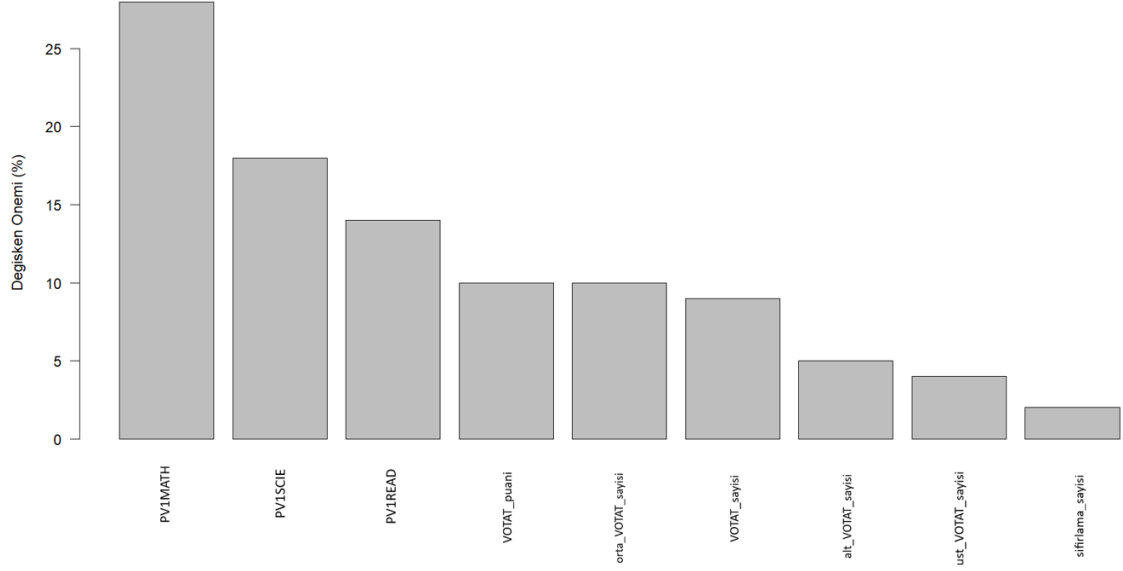


Koşullu Çıkarım Ağaçlarına Dayalı Gradyan Artırma Değişken Önem Grafiği

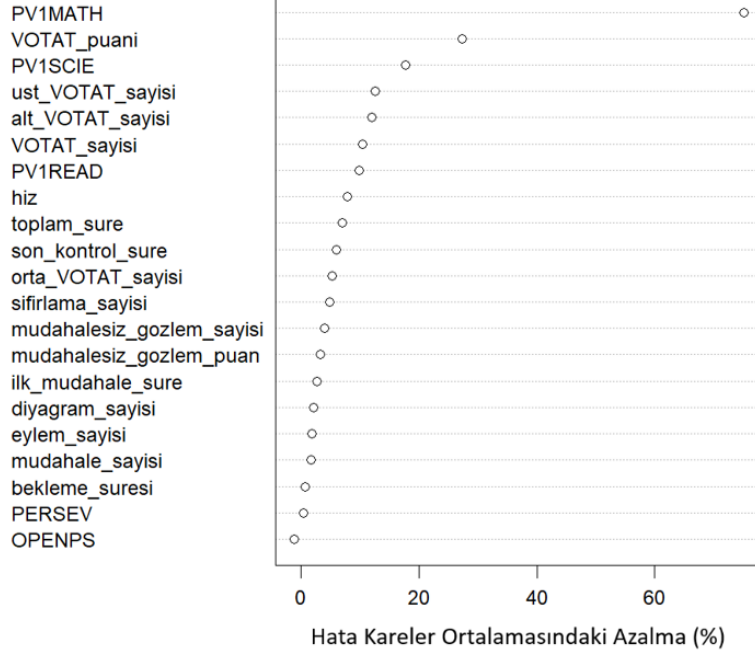


EK-I: Süreç+Sonuç Verisi-Değişken Önem Grafikleri

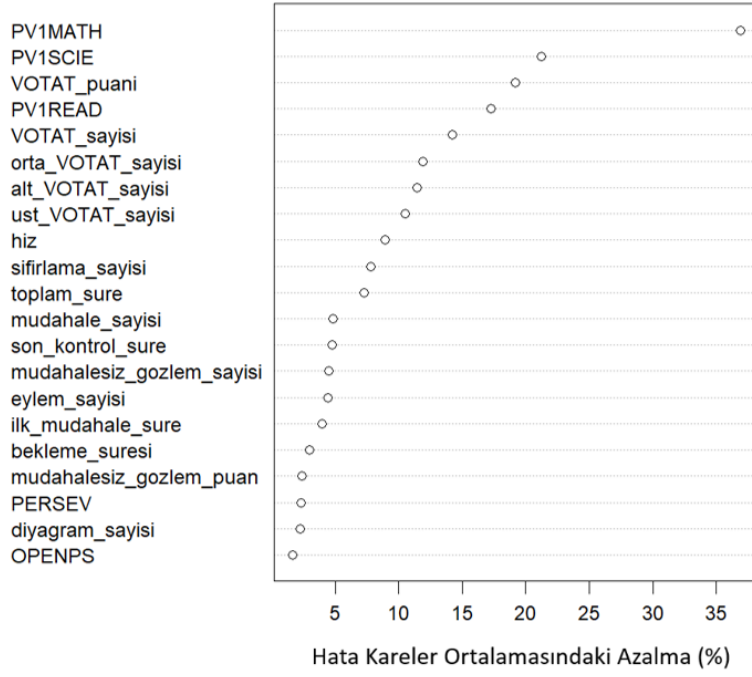
Regresyon Ağacı Değişken Önem Grafiği



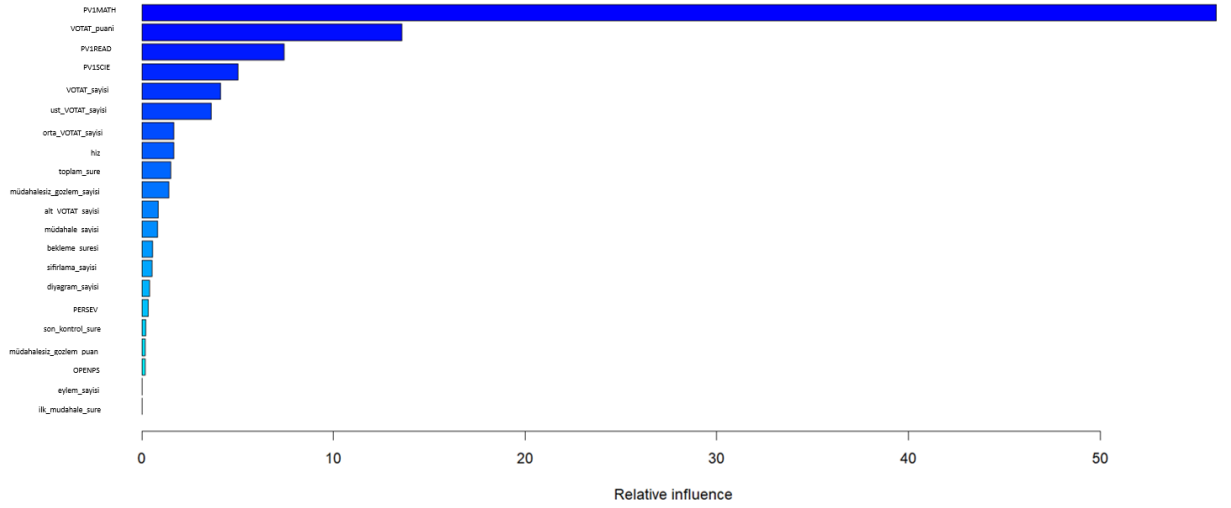
Torbalama Değişken Önem Grafiği



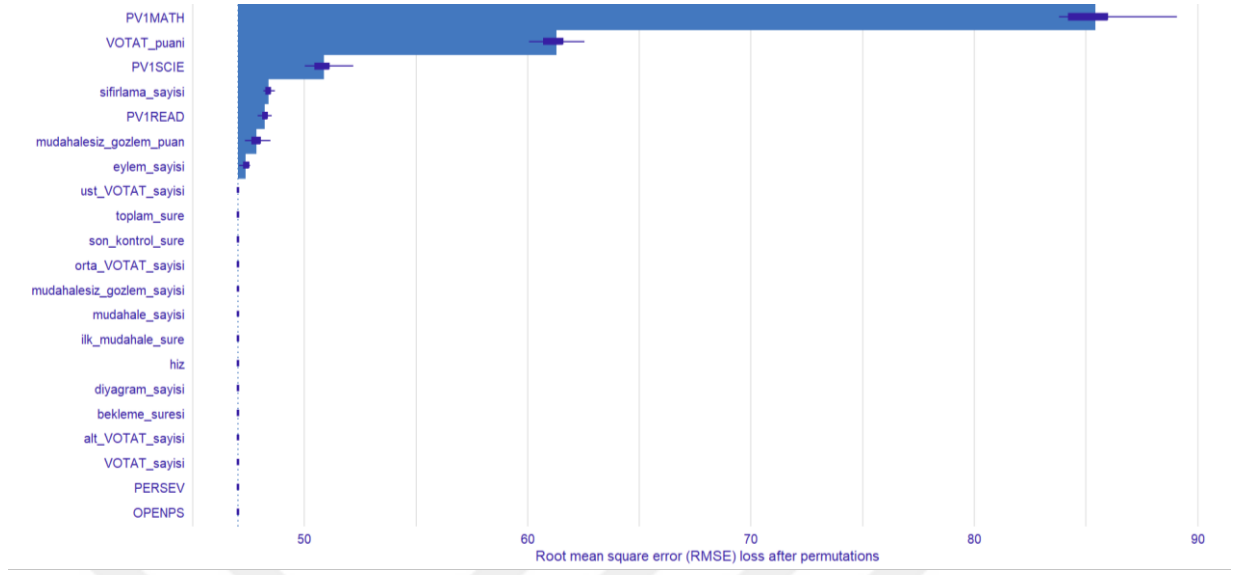
Rastgele Orman (Ölçeklenmiş/Varsayılan) Değişken Önem Grafiği



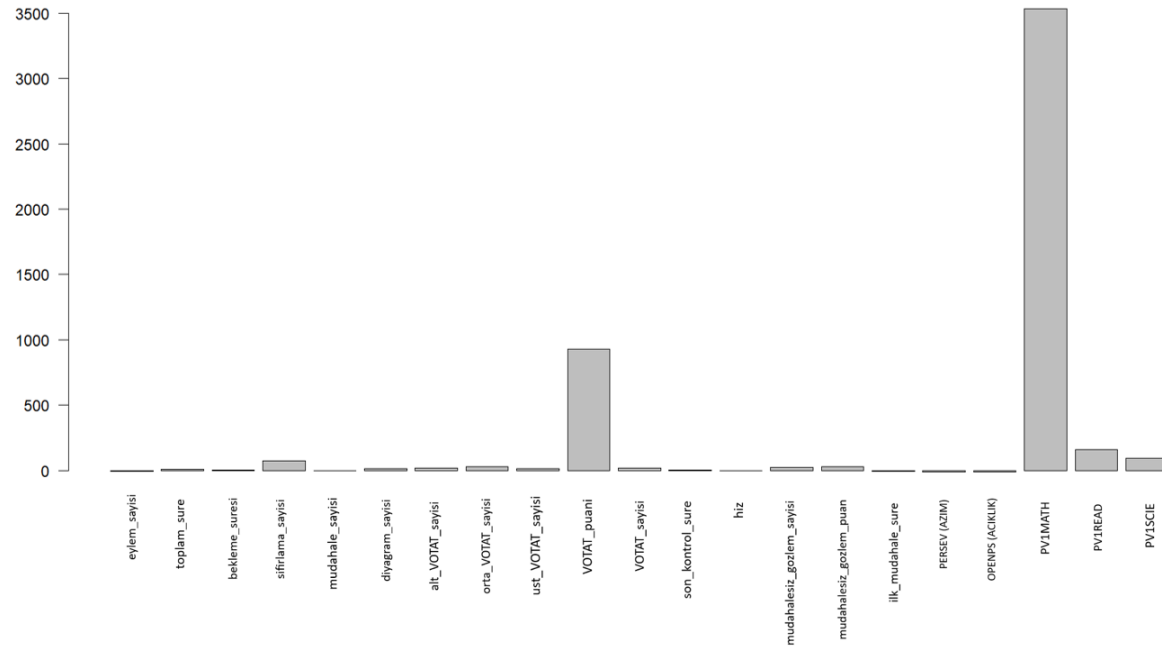
Gradyan Artırma Değişken Önem Grafiği



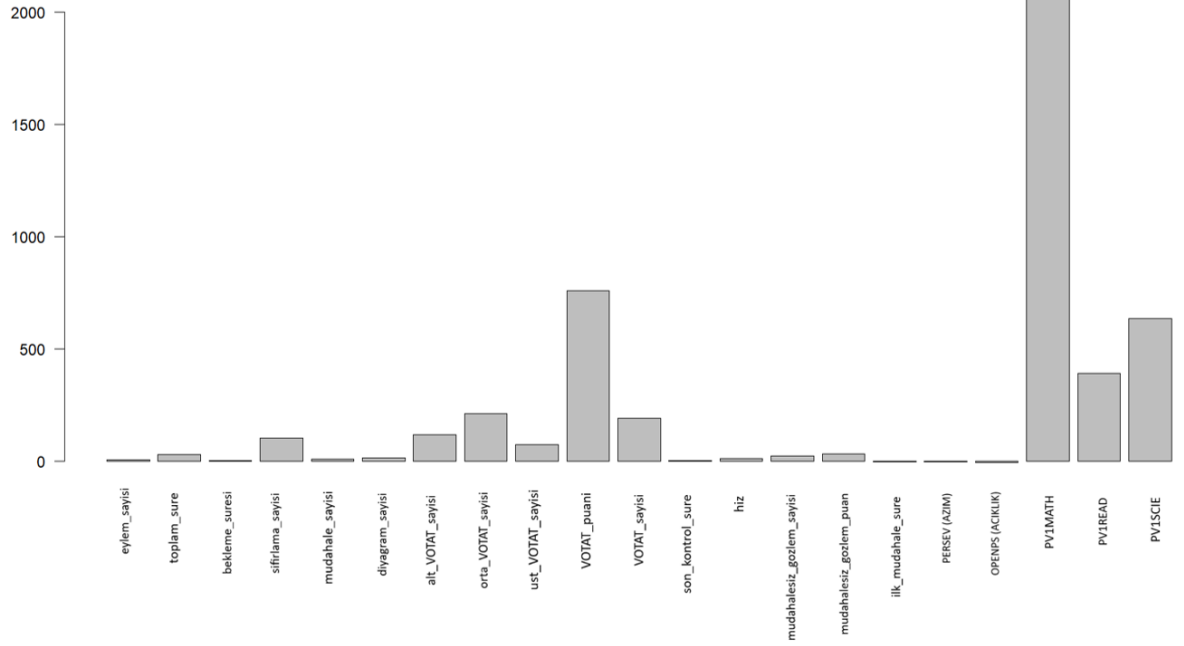
Koşullu Çıkarım Ağacı Değişken Önem Grafiği



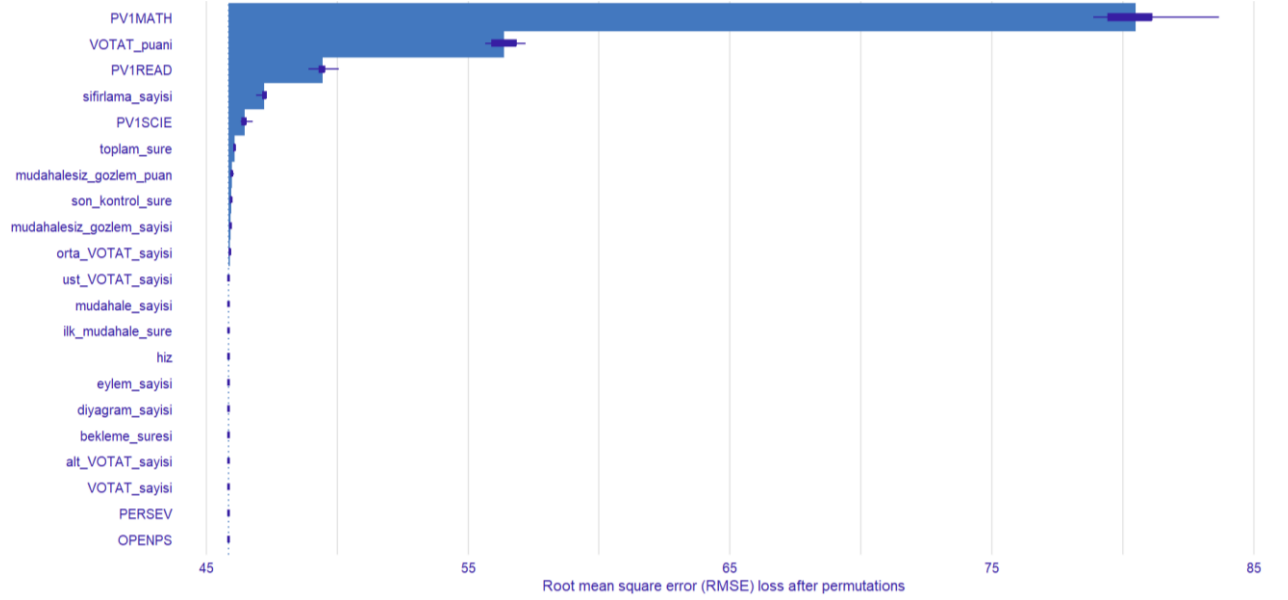
Koşullu Çıkarım Ağaçlarına Dayalı Torbalama Değişken Önem Grafiği



Koşullu Çıkarım Ağaçlarına Dayalı Rastgele Orman (Koşullu Olmayan Versiyon/Varsayılan) Değişken Önem Grafiği



Koşullu Çıkarım Ağaçlarına Dayalı Gradyan Artırma Değişken Önem Grafiği



EK-J: Araştırma Etik Komisyon İzin Muafiyeti Formu



Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Tez Çalışması/Araştırma Etik Komisyon İzin Muafiyeti Formu

F46

09 / 01 / 2023

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına

Tez/Araştırma Başlığı:	REGRESYON AĞAÇLARIYLA SÜREÇ ve SONUÇ VERİLERİNİN KARMAŞIK PROBLEM ÇÖZME BECERİSİNİ YORDAMA DÜZEYLERİNİN İNCELENMESİ
-------------------------------	---

Yukarıda başlığı/konusu verilen tez/araştırma çalışmam,

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır.
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir.
4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırmalar niteliğinde değildir.
5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir.

Çalışmada kullanacağım veriler:

- (X) Kamusal erişime açık (buraya yazınız): PISA 2012 değerlendirmesine ait veriler kullanılacaktır.
 () Özel izin ve onaya tabi (buraya yazınız):
 () Üretilmiş veri (buraya yazınız):
 () Diğer (buraya yazınız):

Yükseköğretim Kurumları Etik Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Komisyonundan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Fatma Nur Aydın
(Araştırmacı Adı Soyadı, İmzası)

Araştırmacı Bilgileri

Adı Soyadı	Fatma Nur Aydın
Öğrenci İse No	N20141134
Ana Bilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	☐ Yüksek Lisans ☒ Doktora ☐ Bütünleşik Dr. ☐ Diğer

Danışman Görüşü ve Onayı*

Doç. Dr. Kübra ATALAY KABASAKAL
(İmza)
(Danışmanın İmzası, Adı ve Soyadı)

*Tez ve tezden üretilen yayınlarda gerekli

EK-K: Etik Beyanı

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- * tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- * görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- * başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- * atıfta bulunduğum eserlerin bütününe kaynak olarak gösterdiğimi,
- * kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- * bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

18/11/2024

Fatma Nur AYDIN

EK-L: Doktora Tez Çalışması Orijinallik Raporu

18/11/2024

HACETTEPE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı : REGRESYON AĞAÇLARIYLA SÜREÇ VE SONUÇ VERİLERİNİN KARMAŞIK PROBLEM ÇÖZME BECERİSİNİ YORDAMA DÜZEYLERİNİN İNCELENMESİ
Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
23/09/2024	219	288,492	04/11/2024	%3	2463272303

Uygulanan filtreler:

- Kaynaklar hariç
- Alıntılar dâhil
- 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esaslarını inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Fatma Nur AYDIN

Öğrenci No.: N20141134

Ana Bilim Dalı: Eğitim Bilimleri

İmza

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: ☐ Y.Lisans ☒ Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

Doç. Dr. Kübra ATALAY KABASAKAL

EK-M: Dissertation Originality Report

18/11/2024

HACETTEPE UNIVERSITY
Graduate School of Educational Sciences
To The Department of Educational Sciences

Thesis Title: INVESTIGATION OF THE LEVELS OF PROCESS AND RESULT DATA TO PREDICTION COMPLEX PROBLEM SOLVING WITH REGRESSION TREES

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
23/09/2024	219	288,492	04/11/2024	%3	2463272303

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Name Lastname: Fatma Nur AYDIN

Student No.: N20141134

Department: Educational Sciences

Program: Measurement and Evaluation in Education

Status: ☐ Masters ☒ Ph.D. ☐ Integrated Ph.D.

Signature

ADVISOR APPROVAL

APPROVED
Assoc. Prof. Dr. Kübra ATALAY KABASAKAL

EK-N: Yayınlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına ilişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- ☐ Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- ☐ Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren ... ay ertelenmiştir. ⁽²⁾
- ☐ Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

18/11/2024

Fatma Nur AYDIN

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına ilişkin Yönerge"

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlerle ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir
*Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

