

T.C. İSTANBUL KÜLTÜR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**DİJİTAL EĞİTİMDE SÜREÇ MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ:
LMS VERİLERİ İLE ÖĞRENCİ DAVRANIŞ MODELLERİ VE AKADEMİK
PERFORMANS TAHMİNİ**

DOKTORA TEZİ
FETTAH KURTULUŞ
1410012002

Anabilim Dalı: İşletme

Programı: İşletme

Tez Danışmanı: Doç. Dr. Murat Taha BİLİŞİK

Haziran 2025

T.C. İSTANBUL KÜLTÜR ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**DİJİTAL EĞİTİMDE SÜREÇ MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ:
LMS VERİLERİ İLE ÖĞRENCİ DAVRANIŞ MODELLERİ VE AKADEMİK
PERFORMANS TAHMİNİ**

DOKTORA TEZİ

FETTAH KURTULUŞ

1410012002

Anabilim Dalı: İşletme

Programı: İşletme

Tez Danışmanı: Doç. Dr. Murat Taha BİLİŞİK

Jüri üyeleri: Prof. Dr. Akhan Akbulut

Doç. Dr. Çiğdem Kaya

Doç. Dr. Ersin Namlı

Dr. Bilge Turp Gölbaşı

Haziran 2025

ÖNSÖZ

Bu tez çalışması, akademik hayatımın önemli dönüm noktalarından biri olmuştur. Araştırma sürecinde edindiğim bilgi birikimi ve deneyimler hem akademik hem de kişisel gelişimime değerli katkılar sağlamıştır.

Tez çalışmam süresince bilgi birikimi, yol gösterici desteği ve kıymetli katkılarıyla bana rehberlik eden doktora danışmanım Doç. Dr. Murat Taha BİLİŞİK'e en içten teşekkürlerimi sunarım.

Ayrıca, yüksek lisans eğitimim sırasında danışmanlığımı üstlenen ve fikirleri her zaman benim için değerli ve yol gösterici olan Prof. Dr. Akhan AKBULUT'a şükranlarımı sunarım.

Araştırmalarımnda sağladığı veri desteği ve süreç boyunca sunduğu katkılarla çalışmalarımın şekillenmesinde önemli rol oynayan çalışma arkadaşım ve birim müdürüm Ahmet Gökhan ULUÇAY'a teşekkür ederim.

Aynı zamanda tez sürecimde gösterdiği destek ve teşvik edici yaklaşımı için İstanbul Kültür Üniversitesi Genel Sekreteri Ender Rıza EKİCİ'ye de şükranlarımı sunarım.

Tez sürecinde doğrudan veya dolaylı olarak katkı sunan İstanbul Kültür Üniversitesi İşletme Bölümü öğretim üyelerine ve tüm akademik camiaya teşekkür ederim.

Her zaman yanımda olan, sabırları ve destekleriyle bana güç veren sevgili eşime ve aileme en derin minnettarlığımı sunarım. Onların varlığı ve desteği, bu süreci daha anlamlı ve değerli kılmıştır.

Son olarak, bu çalışmanın eğitimde dijitalleşme, süreç madenciliği ve makine öğrenmesi gibi alanlarda literatüre katkı sağlamasını ve ilerleyen çalışmalara ilham vermesini temenni ederim.

İÇİNDEKİLER

ÖNSÖZ.....	i
KISALTMALAR	v
TABLO LİSTESİ	vi
ŞEKİL LİSTESİ.....	vii
DENKLEM LİSTESİ	ix
KISA ÖZET	x
ABSTRACT	xi
1. GİRİŞ	1
1.1 Araştırma Amacı.....	2
1.2 Araştırma Kapsamı.....	2
1.3 Araştırma Motivasyonu ve Hipotezler	3
1.4 Araştırma Soruları.....	4
2. KAVRAMSAL ÇERÇEVE VE LİTERATÜR TARAMASI.....	6
2.1 Öğrenim Yönetim Sistemi	6
2.2 Makine Öğrenmesi.....	7
2.2.1 Makine Öğrenmesinin Türleri.....	7
2.2.2 Eğitimde Makine Öğrenmesinin Temelleri	8
2.3 Veri Madenciliği ve Eğitimsel Veri Madenciliği	9
2.4 Süreç Madenciliği (PM).....	10
2.4.1 Süreçlerin Keşfi	11
2.4.2 Uygunluk Kontrolü	12
2.4.3 Süreçlerin İyileştirilmesi.....	13
2.4.4 Süreç Madenciliğinde Kullanılan Araçlar	13
2.5 Eğitimsel Süreç Madenciliği (EPM)	14
2.5.1 EPM ve ML Entegrasyonu	18
3. ARAŞTIRMA TASARIMI VE YÖNTEMİ	21
3.1 Veri Kaynağı Özellikleri.....	21
3.1.1 Veri Kaynağı.....	21
3.1.2 Toplanan Verilerin Miktarı ve Türleri	23
3.1.3 Veri Türleri.....	24
3.1.4 Anonimleştirme ve Veri Güvenliği	25

3.1.5	Verinin Geçerliliği ve Güvenilirliği	25
3.2	Veri Toplama ve Ön işleme	26
3.2.1	Veri Toplama	26
3.2.2	Eksik Veri Doldurma	27
3.2.3	Aykırı Değer Temizleme	27
3.2.4	Normalizasyon ve Özellik Mühendisliği	29
3.3	Süreç Madenciliği Teknikleri	32
3.4	Makine Öğrenmesi Modelleri	34
3.4.1	Kullanılan Algoritmalar	34
3.4.2	Performans Ölçütleri	36
3.4.3	Doğrulama Yöntemi	38
3.4.4	Tutarlılık Değerlendirme Kriterleri	38
3.4.5	Hiperparametre Ayarlama Süreci	42
3.5	Süreç Madenciliği ve Makine Öğrenmesi Tekniklerinin Seçimi	43
3.6	Araçlar ve Yazılımlar	43
3.7	Etik Beyan	44
4.	UYGULAMA VE DENEYSEL ÇALIŞMALAR	45
4.1	Makine Öğrenmesi Modellerinin Eğitimi ve Performansı	45
4.1.1	Rastgele Orman (RF)	45
4.1.2	K-En Yakın Komşu (KNN)	46
4.1.3	Karar Ağaçları (DT)	47
4.1.4	Destek Vektör Makineleri (SVM)	48
4.1.5	LightGBM	49
4.1.6	XGBoost	50
4.1.7	Model Performanslarının Bütünsel Değerlendirmesi	51
4.1.8	Model Açıklanabilirlik Analizi	54
4.2	Süreç Madenciliği Tekniklerinin Uygulanması	57
4.2.1	Genel Süreç Yapısı	57
4.2.2	Başarılı Öğrencilerin Süreç Modeli (İdeal Öğrenme Süreci)	65
4.2.3	Başarısız Bir Öğrenci: Vaka Analizi	68
4.2.4	Süreç Uyum Skorlarının (Fitness) Hesaplanması	71
4.3	EPM Destekli Veri Seti Üzerinde Makine Öğrenmesi Modellerinin Eğitimi ve Performansı	72
4.3.1	Rastgele Orman (RF)	72
4.3.2	K-En Yakın Komşu (KNN)	74

4.3.3	Karar Ağacı (DT)	75
4.3.4	Destek Vektör Makinesi (SVM).....	76
4.3.5	LightGBM.....	77
4.3.6	XGBoost	79
4.3.7	Model Performanslarının Bütünsel Değerlendirilmesi.....	80
4.3.8	Süreç Uygunluk Skorlarının Başarıyla İlişkisi.....	83
4.3.9	EPM Destekli Model Açıklanabilirlik Analizi.....	84
4.4	Sonuçların Performans Ölçütleriyle Değerlendirilmesi	88
4.4.1	Sonuçların Performans Ölçütleriyle Değerlendirilmesi	88
4.4.2	Tezde Önerilen Model Yaklaşımının Değerlendirilmesi	93
4.4.3	Genel Yorum ve Akademik Katkı	93
5.	TARTIŞMALAR VE SONUÇLAR.....	95
5.1	Araştırma Sorularının Yanıtlanması	95
5.2	Hipotezin Test Edilmesi ve Yorumu.....	97
5.3	Literatürle Karşılaştırma	98
5.4	Katkılar ve Sınırlılıklar	99
5.5	Gelecek Çalışmalar için Öneriler	101
5.6	Genel Değerlendirme ve Sonuçlar	102
	KAYNAKÇA	104

KISALTMALAR

DM	: Veri Madenciliđi
EDM	: Eđitimsel Veri Madenciliđi
PM	: Sre Madenciliđi
EPM	: Eđitimsel Sre Madenciliđi
LMS	: đrenme Ynetim Sistemi
ERP	: Kurumsal Kaynak Planlama
CMS	: Mřteri Ynetim Sistemi
ML	: Makine đrenmesi
KNN	: K En Yakın Komřu
RF	: Rastgele Orman
SVM	: Destek Vektr Makinesi
MOOCs	: Byk Aık evrimii Dersler
TPR	: Dođru Pozitif Oranı
FPR	: Yanlıř Pozitif Oranı
SHAP	: Shapley Deđeri Tabanlı Katkı Aıklamaları
CV	: apraz Dođrulama
K-FOLD CV	: K-Katlı apraz Dođrulama
ROC	: Alıcı İřletim Karakteristiđi
AUC	: Eđri Altı Alan

TABLO LİSTESİ

Tablo 3.1 EDM Veri Seti Özellikleri	22
Tablo 3.2 Eğitimsel Süreç Madenciliği için LMS Olay Günlükleri Özellikleri	22
Tablo 3.3 Verilerin İstatistiksel Değerlendirmeleri	24
Tablo 3.4 - Veri Setinde Kayıtlı Olay Sayıları.....	25
Tablo 3.5 Tutarlılık Kategorileri	39
Tablo 3.6 Hiperparametre Arama Alanları	42
Tablo 4.1 - Model Performans Tablosu (Split & K-Fold)	51
Tablo 4.2 - İdeal Öğrenci İle Kalan Öğrenci Davranışları Frekans Karşılaştırması..	71
Tablo 4.3 - Süreç Uyum Skoru Örnek Veri Seti	72
Tablo 4.4 - EPM Destekli Model Performans Tablosu (Split & K-Fold).....	80



ŞEKİL LİSTESİ

Şekil 2.1 - Süreç madenciliğinin üç ana türünün konumlandırılması: keşif, uygunluk ve iyileştirme.....	10
Şekil 2.2- LMS olay günlüğü örneği.....	11
Şekil 2.3- Süreç keşfi girdi ve çıktı açısından açıklanması.....	12
Şekil 2.4- Süreç uygunluk kontrolü girdi ve çıktı açısından açıklanması.....	12
Şekil 2.5- Süreçlerin iyileştirmesi girdi ve çıktı açısından açıklanması	13
Şekil 2.6- Eğitimsel Süreç Madenciliği Çerçevesi: türleri ve bileşenleri	15
Şekil 4.1 - Rastgele Orman Karmaşıklık Matrisi	46
Şekil 4.2 - Rastgele Orman Roc Eğrisi Grafiği.....	46
Şekil 4.3- K-En Yakın Komşu Karmaşıklık Matrisi.....	47
Şekil 4.4 - K-En Yakın Komşu Roc Eğrisi Grafiği.....	47
Şekil 4.5- Karar Ağaçları Karmaşıklık Matrisi	48
Şekil 4.6 - Karar Ağaçları Roc Eğrisi Grafiği.....	48
Şekil 4.7 - Destek Vektör Makineleri Karmaşıklık Matrisi	49
Şekil 4.8 - Destek Vektör Makineleri Roc Eğrisi Grafiği.....	49
Şekil 4.9 - Lightgbm Karmaşıklık Matrisi	50
Şekil 4.10 - Lightgbm Roc Eğrisi Grafiği.....	50
Şekil 4.11 - XGboost Karmaşıklık Matrisi	51
Şekil 4.12 - XGBoost Roc Eğrisi Grafiği.....	51
Şekil 4.13 - F1 skorlarına göre 80-20 Split ve K-Fold CV yöntemlerinin karşılaştırmalı Grafiği	52
Şekil 4.14 - Karşılaştırmalı Roc Eğrileri Grafiği	53
Şekil 4.15- SHAP Özellik Önem Grafiği	54
Şekil 4.16 - SHAP Şelale Grafiği – Bireysel Tahmin.....	55
Şekil 4.17 - Genel Eğitim Süreç Modeli	57
Şekil 4.18- Ödev Gönderme Süreci	60
Şekil 4.19 - İdeal Öğrenci Grubu Süreç Modeli	66
Şekil 4.20 - Başarısız Bir Öğrenci Süreç Modeli.....	69
Şekil 4.21- EPM Destekli Rastgele Orman Karmaşıklık Matrisi	73
Şekil 4.22 EPM Destekli Rastgele Orman Roc Eğrisi Grafiği	73
Şekil 4.23- EPM Destekli Rastgele K-En Yakın Komşu Matrisi	74
Şekil 4.24 - EPM Destekli K-En Yakın Komşu Roc Eğrisi Grafiği	75
Şekil 4.25 - EPM Destekli Rastgele Karar Ağacı Matrisi.....	76
Şekil 4.26 - EPM Destekli Karar Ağacı Roc Eğrisi Grafiği	76
Şekil 4.27 - Birleştirilmiş Veri Seti Rastgele Destek Vektör Makinesi Matrisi	77
Şekil 4.28 - EPM Destekli Destek Vektör Makinesi Roc Eğrisi Grafiği	77
Şekil 4.29- EPM Destekli Rastgele LightGBM Matrisi	78
Şekil 4.30 - EPM Destekli LightGBM Roc Eğrisi Grafiği	78
Şekil 4.31- Birleştirilmiş Veri Seti Rastgele XGBoost Matrisi	79
Şekil 4.32 – EPM Destekli XGBoost Roc Eğrisi Grafiği	80
Şekil 4.33 - EPM Destekli Model F1 skorlarına göre 80-20 Split ve K-Fold CV Yöntemleri ile Karşılaştırma Grafiği	82
Şekil 4.34 - EPM Destekli Model Karşılaştırmalı Roc Eğrileri Grafiği	83
Şekil 4.35- Süreç Uyum Skorunun Başarı Durumuna Göre Dağılımı	84

Şekil 4.36 - EPM Destekli Model SHAP Özellik Önem Grafiği.....	85
Şekil 4.37 - EPM Destekli Model SHAP Şelale Grafiği – Bireysel Tahmin.....	86
Şekil 4.38 - EPM Desteği Sonrası Model Performanslarının Karşılaştırması (F1 Split & CV).....	89
Şekil 4.39 - EDM ve EPM AUC (ROC) Değerlerinin Karşılaştırılması	90
Şekil 4.40 - Doğrulama Duyarlılığı ($\Delta F1$) Karşılaştırması – EDM ve EPM.....	91
Şekil 4.41 - Model Performans Gelişim Isı Haritası	92



DENKLEM LİSTESİ

Denklem 3.1- Z-skoru	28
Denklem 3.2- Min-Max Normalizasyon	29
Denklem 3.3- F1 Skor	36
Denklem 3.4- True Positive Rate	36
Denklem 3.5- False Positive Rate	37
Denklem 3.6- Kesinlik (Precision)	37
Denklem 3.7- Duyarlılık (Recall)	38



KISA ÖZET

COVID-19 pandemisi ile birlikte çevrim içi eğitim platformlarının, özellikle Öğrenme Yönetim Sistemleri'nin (LMS) kullanımı büyük ölçüde artmıştır (Fitriani, 2020). Bu dönüşüm, öğrencilerin dijital ortamlardaki davranışlarının analiz edilmesini ve akademik başarı ile olan ilişkilerinin daha derinlemesine incelenmesini gerekli kılmıştır.

Bu tezde, bir üniversitenin LMS verileri kullanılarak, öğrencilerin öğrenme süreçleri eğitimsel süreç madenciliği (EPM) teknikleriyle analiz edilmiş ve elde edilen süreçsel davranış göstergeleri, makine öğrenmesi algoritmaları ile akademik başarı tahmini için modellenmiştir. Süreç madenciliği aracılığıyla elde edilen süreç uyum skoru gibi değişkenler, geleneksel eğitimsel veri madenciliği (Educational Data Mining) değişkenleriyle birleştirilerek EPM destekli veri setleri oluşturulmuştur.

Farklı makine öğrenmesi modelleri (Random Forest, XGBoost, LightGBM vb.) üzerinde yapılan karşılaştırmalı analizler sonucunda, süreç temelli değişkenlerin modele dahil edilmesinin F1 skoru ve ROC-AUC gibi performans metriklerinde anlamlı iyileşmeler sağladığı görülmüştür. Ayrıca SHAP analizi ile model çıktılarının açıklanabilirliği artırılmış; süreç uyum skorunun başarı tahmininde en yüksek katkıya sahip değişken olduğu belirlenmiştir.

Bu çalışma, klasik öğrenme analitiği yaklaşımlarının ötesine geçerek sürece dayalı öğrenme davranışlarının modellenmesine katkı sağlamaktadır. Elde edilen bulgular, LMS tabanlı karar destek sistemlerinin geliştirilmesi ve öğrenci başarısını artırmaya yönelik kişiselleştirilmiş müdahalelerin tasarlanması açısından önemli bir zemin sunmaktadır.

Anahtar Kelimeler: Eğitimsel Süreç Madenciliği, Makine Öğrenmesi, Öğrenim Yönetim Sistemleri

ABSTRACT

With the COVID-19 pandemic, the use of online learning platforms—particularly Learning Management Systems (LMS)—has significantly increased (Fitriani, 2020). This transformation has made it necessary to analyze students' behaviors in digital environments and examine their relationship with academic success in greater depth.

This thesis analyzes student learning processes using Educational Process Mining (EPM) techniques based on LMS data obtained from a university. Behavioral indicators derived from process mining, such as the fitness score, were integrated with traditional Educational Data Mining (EDM) variables to construct EPM-enhanced datasets, which were then modeled using various machine learning algorithms to predict academic performance.

Comparative analysis across different machine learning models (Random Forest, XGBoost, LightGBM, etc.) revealed that including process-based variables led to statistically significant improvements in performance metrics such as F1 score and ROC-AUC. Furthermore, SHAP analysis enhanced the interpretability of model outputs and identified the fitness score as the most influential feature in predicting student success.

By moving beyond traditional clickstream analysis, this study contributes to the modeling of process-oriented learning behaviors. The findings provide a strong foundation for developing LMS-based decision support systems and designing personalized educational interventions to improve student performance.

Keywords: Educational Process Mining, Machine Learning, Learning Management System

BİRİNCİ BÖLÜM

1. GİRİŞ

Pandemi süresince dünya genelinde öğrenme yönetim sistemlerinin (LMS) kullanımı önemli ölçüde artmıştır. Eğitim kurumlarının uzaktan eğitime geçişi, LMS platformlarının kritik bir rol oynamasına neden olmuştur. Örneğin, COVID-19 pandemisi sırasında uzaktan eğitim, öğrencilerin eğitim süreçlerini sürdürebilmesi için bir zorunluluk haline gelmiş ve bu süreçte LMS'ler yaygın olarak kullanılmıştır (Fitriani, 2020). LMS sistemlerinin kullanımının artması bu sistemler içerisinde toplanan verilerin analiz edilmesi ve yorumlanması ihtiyacını beraberinde getirmiştir.

COVID-19 pandemisi sırasında öğrenme yönetim sistemleri (LMS), yalnızca öğretim süreçlerini sürdürmek için değil, aynı zamanda öğrenci davranışlarının ve akademik performanslarının izlenmesi için de kritik bir araç haline gelmiştir. Araştırmalar, uzaktan eğitim sürecinde öğrenci davranışlarında değişiklikler olduğunu ve bu değişikliklerin akademik performans üzerinde önemli etkiler yarattığını göstermektedir. Örneğin, pandemi döneminde LMS'ye daha az zaman harcayan öğrencilerin akademik performanslarının, daha fazla zaman geçiren öğrencilere kıyasla düştüğü gözlemlenmiştir (Lobos, et al., 2021). Ayrıca, dijital araçların ve çevrimiçi eğitimin kullanımının öğrenci başarısı ile pozitif bir ilişki içinde olduğu tespit edilmiştir (Atlam, Ewis, El-Raouf, Ghoneim, & Gad, 2022).

Öğrenme Yönetim Sistemlerinde (LMS) biriken büyük hacimli veriler, süreç madenciliği ve makine öğrenmesi teknikleri kullanılarak analiz edilebilmektedir. Süreç madenciliği, öğrenci davranışlarını modellemeye ve eğitim süreçlerindeki gizli kalıpları ortaya çıkarmaya olanak tanır. Özellikle COVID-19 pandemisi döneminde LMS sistemleri üzerinden biriken veriler, öğrenci davranışlarındaki eğilimlerin daha iyi anlaşılmasını sağlamış ve akademik performansı öngörme noktasında değerli bilgiler sunmuştur. Araştırmalar, LMS'den elde edilen verilerin süreç madenciliği teknikleri ile işlenerek, öğrenci etkileşimlerinin akademik başarı ile ilişkisini anlamada büyük katkı sağladığını göstermektedir (Brozina, Knight, Kinoshita, & Johri, 2019).

Makine öğrenmesi ise, bu süreçlerden elde edilen verilerin işlenmesiyle gelecekteki öğrenci performanslarını tahmin eden modeller geliştirilmesine olanak tanımaktadır. Bu tür modeller, öğrenci davranışlarını izleyerek eğitim süreçlerine dair öngörüler sağlayabilir ve öğretim stratejilerinin kişiselleştirilmesine olanak tanır. LMS verilerine dayalı olarak oluşturulan makine öğrenmesi modelleri, öğrencilerin düşük performans gösterme risklerini önceden tespit edebilmekte ve eğitimcilerin müdahalelerde bulunmalarına olanak sağlamaktadır. Örneğin, LMS verileri üzerinden geliştirilen bir makine öğrenmesi modeli, öğrenci katılımının artmasıyla akademik başarı arasında pozitif bir ilişki olduğunu göstermiştir (Vautier, Enns, & Cadaret, 2023).

1.1 Araştırma Amacı

Bu araştırmanın temel amacı, pandemi süreci ile birlikte yaygınlaşan Öğrenme Yönetim Sistemleri (LMS) üzerinden elde edilen öğrenci davranış verilerini, eğitimsel süreç madenciliği (EPM) ve eğitimsel veri madenciliği (EDM) teknikleri kullanarak analiz etmektir. Çalışmada, öğrencilerin öğrenme süreçlerinde sergiledikleri davranışların süreç madenciliği yöntemleriyle modellenmesi ve bu süreç verilerinin makine öğrenmesi algoritmalarıyla bütünleştirilerek akademik başarı tahmini yapabilecek bir model geliştirilmesi hedeflenmektedir. Bu amaç doğrultusunda, klasik EDM değişkenlerinin yanı sıra EPM ile elde edilen süreç uyum skorları da modele dahil edilerek, öğrenci davranışlarının daha derinlikli ve süreç temelli bir bakış açısıyla analiz edilmesi sağlanacaktır. Elde edilen sonuçlar ile, LMS ortamlarında öğrencilerin akademik başarılarının erken dönemde tahmin edilmesi, riskli öğrenci gruplarının belirlenmesi ve eğitim süreçlerinin bireyselleştirilmiş geri bildirim mekanizmaları yoluyla iyileştirilmesi amaçlanmaktadır.

1.2 Araştırma Kapsamı

Bu çalışma, Türkiye'deki bir vakıf üniversitesinde kullanılan Sakai LMS platformu aracılığıyla elde edilen öğrenci davranış verilerine odaklanmaktadır. Araştırmanın veri seti, bir yarı yıl boyunca platform üzerinde kaydedilen olay günlükleri (event logs) ve performans verilerinden oluşmaktadır. Toplanan veriler; öğrencilerin içerik okuma, ödev gönderme, sınavlara katılım, duyuru takibi gibi davranışlarını içermekte olup, süreç madenciliği teknikleri (ProM, Inductive Miner vb.) kullanılarak modellenmiştir.

Çalışmanın kapsamı dahilinde:

- Öğrenci davranışlarının süreç madenciliği yöntemleri ile çıkarılması,
- Süreç uyum skorlarının (fitness score) hesaplanması,
- Bu süreç verilerinin klasik EDM değişkenleri ile birleştirilmesi,
- EPM destekli ve desteklenmemiş veri setleri kullanılarak farklı makine öğrenmesi algoritmalarının (Random Forest, XGBoost, KNN, SVM vb.) uygulanması,
- Elde edilen modellerin doğruluk, F1 skoru ve ROC-AUC metrikleri ile performanslarının karşılaştırılması amaçlanmıştır.

Araştırma, bu kapsamda süreç madenciliği ile elde edilen özelliklerin öğrenci akademik başarısının öngörülmesindeki katkısını ortaya koymayı ve eğitimsel veri analitiği alanında yeni bir bakış açısı geliştirmeyi hedeflemektedir.

1.3 Araştırma Motivasyonu ve Hipotezler

Makine öğrenmesi algoritmaları, benzer özelliklere sahip öğrenciler arasında başarıyı ve başarısızlığı ayırt etme noktasında zorluklar yaşamaktadır. Özellikle öğrenci davranış verilerinin yalnızca giriş-çıkış zamanları, materyal erişim sıklıkları veya ödev teslim oranları gibi yüzeysel istatistiklerle sınırlı kalması, algoritmaların başarı oranlarının düşmesine neden olmaktadır. Bu çalışmanın temel problemi, LMS verilerine süreç madenciliği tekniklerinin entegre edilmemesi ve öğrenci öğrenme süreçlerinin dikkate alınmamasıdır.

Bu hipotezler, süreç madenciliği tekniklerinin makine öğrenmesi modellerine entegrasyonunun, öğrencilerin öğrenme süreçlerini anlamaya ve akademik başarılarını tahmin etmeye nasıl katkı sağlayabileceğini test etmek için formüle edilmiştir:

H0 (Null Hypothesis):

Süreç madenciliği teknikleri kullanılarak öğrenme davranışlarının veri setine eklenmesinin, makine öğrenmesi algoritmalarının doğruluk, F1 skoru veya ROC-AUC metriklerinde anlamlı bir etkisi yoktur.

H1 (Alternative Hypothesis):

Süreç madenciliği teknikleri kullanılarak öğrenme davranışlarının veri setine eklenmesi, makine öğrenmesi algoritmalarının doğruluk, F1 skoru veya ROC-AUC metriklerinde anlamlı bir iyileşme sağlar.

Bu hipotezler, 1.4 Araştırma Soruları bölümünde tanımlanan sorulara verilen yanıtlar üzerinden test edilmiş ve analiz edilmiştir.

1.4 Araştırma Soruları

Bu çalışmanın temel amacı, eğitimsel süreç madenciliği (EPM) ile zenginleştirilen veri setleri kullanılarak, makine öğrenmesi algoritmaları ile öğrenci akademik başarısının tahmin edilebilme düzeyini ortaya koymaktır. Ayrıca, süreç madenciliğinden türetilen yeni özelliklerin, geleneksel eğitimsel veri madenciliği (EDM) değişkenlerine kıyasla model performansına olan katkısı ve bu modellerin karar mekanizmalarının açıklanabilirliği de incelenmiştir.

Soru No Araştırma Sorusu

- S1** Eğitimsel süreç madenciliği teknikleri ile desteklenen veri setleri, makine öğrenmesi algoritmaları kullanılarak öğrenci akademik başarısını ne düzeyde doğru tahmin edebilir?
- S2** Eğitimsel süreç madenciliği (EPM) ile elde edilen süreç uyum skorunun (fitness score) öğrenci akademik başarısına etkisi nedir?
- S3** Klasik eğitimsel veri madenciliği (EDM) göstergeleri ile süreç madenciliğinden türetilen göstergelerin başarı tahmini üzerindeki katkıları nasıl farklılaşmaktadır?
- S4** EPM destekli ve desteklenmemiş veri setleri kullanıldığında makine öğrenmesi modellerinin (Random Forest, XGBoost, SVM, KNN vb.) performansları nasıl değişmektedir?
- S5** SHAP açıklanabilirlik analizi kullanılarak öğrenci başarı tahmin modellerinin karar mekanizmaları nasıl yorumlanabilir?

Soru No Arařtırma Sorusu

S6 Öğrencilerin aktif katılım göstergeleri (ödev teslimi, sınavlara katılım) ile pasif katılım göstergeleri (okunan içerik sayısı) arasındaki farklar, başarı tahmininde nasıl bir rol oynamaktadır?

S7 Süreç uyumu (fitness score) ile öğrenci davranışlarının (okuma, tamamlama, sınava katılma) bütüncül ilişkisi başarı tahminlerinde nasıl bir etkileşim göstermektedir?

S8 Eğitimde makine öğrenmesi modellerinin açıklanabilirliğini artırmak amacıyla SHAP gibi yöntemlerin kullanımı, eğitimcilere öğrenci destek süreçlerinde nasıl katkı sağlayabilir?

Bu çalışmada, yukarıda tanımlanan araştırma sorularının yanıtları hipotezlerin test edilmesine temel oluşturmaktadır.

İKİNCİ BÖLÜM

2. KAVRAMSAL ÇERÇEVE VE LİTERATÜR TARAMASI

Kavramsal çerçeve ve literatür taraması, bu araştırmanın dayandığı teorik temelleri ve daha önce yapılmış çalışmaları ele alarak, çalışmanın akademik bağlamını ortaya koymaktadır. Öğrenme Yönetim Sistemleri (LMS), süreç madenciliği ve makine öğrenmesi gibi teknolojilerin eğitimde kullanımı, son yıllarda giderek artan bir araştırma alanı haline gelmiştir. Bu bölümde, LMS'lerin öğrenci davranışlarını izleme, süreç madenciliği teknikleri ile bu davranışların modellenmesi ve makine öğrenmesi algoritmaları ile akademik başarı öngörülerinin yapılması üzerine mevcut literatür incelenecektir. Literatür taraması, öğrenci davranışlarının analiz edilmesi ve süreç madenciliği ile elde edilen bilgilerin makine öğrenmesi ile nasıl entegre edilebileceği konusundaki boşlukları tespit ederek bu araştırmanın dayanağını oluşturmaktadır.

2.1 Öğrenim Yönetim Sistemi

Öğrenim Yönetim Sistemi (LMS), çevrimiçi eğitim içeriklerinin yönetilmesi, dağıtılması ve izlenmesi için kullanılan dijital bir platformdur. Bu sistemler, eğitim materyallerinin kolayca paylaşılmasına olanak tanır, öğrenci katılımını ve ilerlemesini izler ve eğitim süreçlerini daha verimli hale getirir (Akhmedova & Nargisa, 2024).

LMS'nin temel amaçları şunlardır:

1. Eğitim içeriklerini organize etmek ve sunmak.
2. Öğrencilerin ilerlemelerini izlemek ve performanslarını değerlendirmek.
3. Çevrimiçi etkileşimleri kolaylaştırmak ve eğitim süreçlerini kişiselleştirmek.

LMS'nin özellikleri arasında özelleştirilebilir ders tasarımı, entegre değerlendirme araçları, raporlama ve analiz özellikleri yer alır. Bu platformlar hem eğitim kurumlarında hem de iş dünyasında öğrenme ve geliştirme programlarında kullanılmaktadır.

Yaygın kullanılan LMS sistemleri aşağıda belirtilmiştir.

Moodle: Açık kaynaklı bir LMS platformu olup, kullanıcı dostu ve özelleştirilebilir bir yapıya sahiptir. Eğitim kurumlarında sıkça tercih edilmektedir (Dougiamas & Taylor, 2003).

Canvas: Kullanıcı dostu arayüzü ve geniş entegrasyon olanakları ile dikkat çeker. Özellikle üniversiteler ve yükseköğrenim kurumları tarafından kullanılmaktadır (Nguyen & Winman, 2020).

Blackboard: Özellikle büyük eğitim kurumlarında yaygın olarak kullanılan bir platformdur. Gelişmiş raporlama ve analiz özellikleri sunar (Alyadumi & Falcıoglu, 2023).

Google Classroom: Eğitimciler ve öğrenciler arasında kolay etkileşim sağlayan, kullanıcı dostu bir platformdur (Syakur, Sugirin, & Widiarni, 2020).

Sakai LMS: Açık kaynaklı bir öğrenme yönetim sistemidir ve özellikle üniversiteler ile yüksek öğrenim kurumlarında yaygın olarak kullanılmaktadır. Sakai, geniş bir topluluk desteği ile işbirlikçi öğrenme, belge paylaşımı, değerlendirme ve raporlama gibi özellikler sunar (Tagoe & Cole, 2020).

2.2 Makine Öğrenmesi

Makine öğrenmesi (ML), bilgisayarların verilerden öğrenmesine ve deneyimlerle kendini geliştirmesine olanak tanıyan bir yapay zekâ alt dalıdır (Mitchell, 1997). Diğer bir tanımı Makine öğrenimi, büyük verilerdeki kalıpları veya yapıları keşfederek tahmine dayalı içgörüler sunan güçlü bir hesaplama yaklaşımıdır (Sarker, 2021). İnsan müdahalesine gerek kalmadan, verilerden anlamlı desenler çıkararak tahminler veya kararlar almak için kullanılır. Makine öğrenmesi, büyük veri setlerini analiz etmek, öngörude bulunmak, öneriler sunmak ve otomatikleştirilmiş sistemler geliştirmek gibi birçok farklı alanda uygulanmaktadır (Alpaydın, 2020). Makine öğrenmesi ve onun bir alt dalı olan öngörücü analitik, geçmiş verilerden öğrenerek mevcut ve gelecekteki sonuçları iyileştirmeye odaklanır (Sarker, 2021). Bu bölümde makine öğrenmesi türlerine ve eğitimde kullanımı ile ilgili araştırmalara değinilecektir.

2.2.1 Makine Öğrenmesinin Türleri

Makine öğrenmesi üç ana kategoriye ayrılır:

Denetimli Öğrenme (Supervised Learning): Giriş ve çıkış verilerinin eşleştiği etiketlenmiş veri setleri üzerinde çalışır. Yaygın Algoritmalar: Doğrusal regresyon,

lojistik regresyon, karar ağaçları, rastgele ormanlar, destek vektör makineleri (Bishop, 2006).

Denetimsiz Öğrenme (Unsupervised Learning): Etiketlenmemiş veri setlerinden desenler ve gruplamalar bulur. Örnekler: kümeleme ve boyut indirgeme. Yaygın Algoritmalar: K-means, hiyerarşik kümeleme, temel bileşen analizi (PCA) (Hastie, Tibshirani, & Friedman, 2009).

Pekiştirmeli Öğrenme (Reinforcement Learning): Bir ajanın ödül mekanizmaları ile bir ortamda en iyi sonuçları öğrenmesini sağlar.

Yaygın Algoritmalar: Q-learning, derin pekiştirmeli öğrenme (Sutton & Barto, 2018).

2.2.2 Eğitimde Makine Öğrenmesinin Temelleri

Makine öğrenmesi, dijital eğitime uygulandığında, öğrenme deneyimini daha iyi anlamak ve akademik performansı iyileştirmek mümkün hale gelir (Iglesias-Pradas, Hernández-García, Chaparro-Peláez, & Prieto, 2021). Dijital eğitimde, öğrencilerin akademik performansını, özellikle üniversite öğrencilerinin not başarılarını tahmin etmek için yaş, cinsiyet, öğrenme etkinlikleri, ödev puanları ve değerlendirme puanları gibi özellikler kullanılarak çeşitli makine öğrenmesi algoritmaları uygulanmıştır (Yağcı, 2022). Kullanılan algoritmalar arasında, dijital öğrenme ortamlarından geliştirilen tahmin veri setlerinde uygulamak üzere doğrusal regresyon analizi, karar ağaçları, rastgele ormanlar, destek vektör makineleri, k-en yakın komşu, Gauss süreçleri, naive Bayes ve sinir ağları gibi yöntemler bulunmaktadır.

Makine öğrenmesi algoritmalarının akademik performans tahmininde kullanımının en önemli avantajlarından biri, öğrencilere yönelik öğrenme yollarının kişiselleştirilmesine olanak sağlamasıdır. Örneğin, bir karar ağacı algoritmasının kullanılmasıyla, akademik başarı açısından risk altında olan öğrencilerin belirgin özellikleri tespit edilebilir ve bu öğrencilere yönelik kişiselleştirilmiş müdahale programları oluşturulabilir. Bu tür bir yaklaşım, daha yüksek veya düşük akademik performansa katkıda bulunan faktörlerin etkisini belirleme ve bu faktörleri önem sırasına göre sıralama imkânı sunmaktadır (Bhutoria, 2022). Bu öngörücü öğrenci modelleri, öğrenenlerin ilerleme durumu, ilgi düzeyi veya katılımına bağlı olarak öğretim içeriğinin çeşitlendirildiği ve değiştirildiği uyarlanabilir öğrenme sistemleri

oluşturmak için kullanılabilir (Strielkowski, Grebennikova, Lisovski, Rakhimova, & Vasileva, 2024).

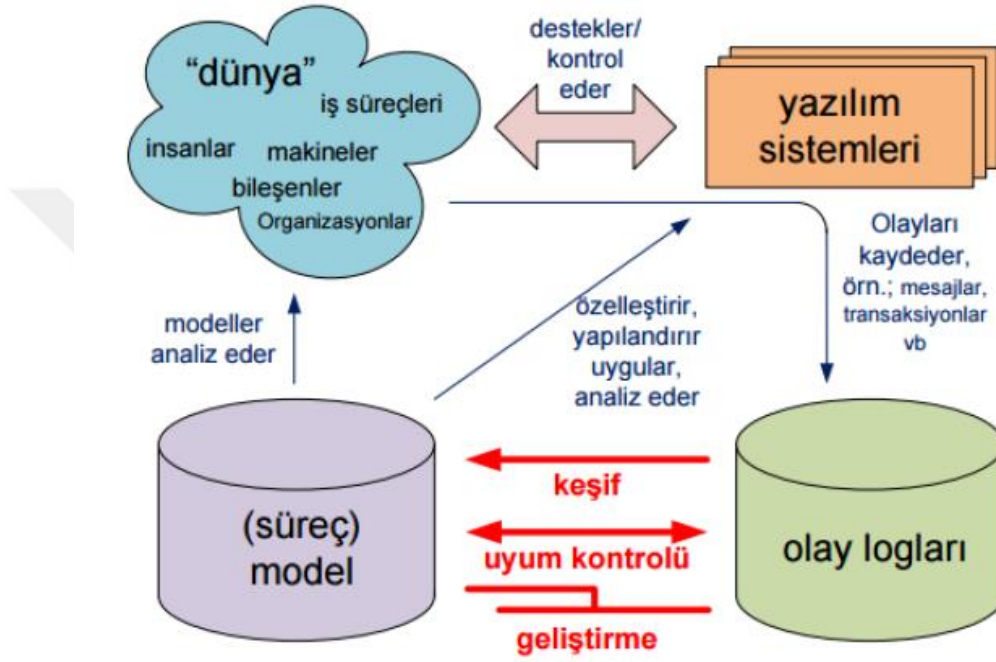
2.3 Veri Madenciliği ve Eğitimsel Veri Madenciliği

Veri madenciliği (DM), büyük veri setlerinden anlamlı desenler, ilişkiler ve bilgilerin çıkarılmasını amaçlayan bir süreçtir ve çeşitli alanlarda karar verme süreçlerini iyileştirmek için kullanılan analitik bir süreçtir (Han, Kamber, & Pei, 2011). Bu teknik, genellikle makine öğrenimi, istatistiksel analiz ve veri tabanı yönetim sistemlerini bir araya getirir. Veri madenciliğinin temel amacı, geçmiş verilere dayanarak tahmin modelleri oluşturmak ve karar verme süreçlerini desteklemektir (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Kullanım alanları arasında pazarlama stratejileri geliştirme, sağlık sektöründe hastalık tahmini, finans sektöründe dolandırıcılık tespiti ve eğitim alanında öğrencilerin başarı faktörlerini analiz etmek gibi çok çeşitli uygulamalar bulunmaktadır.

Son yıllarda veri madenciliği (DM) tekniklerinin eğitim verilerinde kullanımı giderek artmıştır; bu durum, Eğitimsel Veri Madenciliği (EDM) ve Öğrenme Analitikleri (LA) alanlarının oluşumuna yol açmıştır (Romero & Ventura, 2020). Özellikle eğitim verilerinde, öğrenci davranışlarını modelleme ve öğrenme süreçlerini optimize etme gibi önemli roller üstlenmektedir. EPM'in Temel amacı, eğitimde karar destek sistemlerini güçlendirmek ve bireyselleştirilmiş öğrenme yolları sunarak öğrencilerin ihtiyaçlarına daha iyi yanıt vermektir (Baker & Siemens, 2014). Kullanım alanları arasında öğrenme analitiği geliştirme, müfredat tasarımı optimizasyonu ve öğrenci başarı/başarısızlık oranlarını tahmin etme gibi önemli uygulamalar bulunmaktadır (Han, Kamber, & Pei, 2011). Bilici ve Özdemir, 2014-2020 yılları arasında veri madenciliği yöntemlerinin eğitim alanındaki uygulamalarını inceleyen bir literatür taraması yapmışlardır. Çalışma, EDM'nin akademik performans tahmini ve kişiselleştirilmiş öğrenme stratejilerinde önemli sonuçlar sağladığını ortaya koymuştur (Bilici & Özdemir, 2022). AlQaheri ve Panda, bir EPM çerçevesi sunarak öğrencilerin öğrenme davranışlarını anlamayı ve öğretim kalitesini artırmayı hedeflemiştir. Bulgular, olay günlüklerinden anlamlı bilgiler elde edilebileceğini göstermektedir (AlQaheri & Panda, 2022). Yağcı, makine öğrenimi algoritmalarını kullanarak öğrencilerin akademik performanslarını tahmin etmiş ve modelin doğruluğunu artırmayı amaçlamıştır. Çalışma, sınıflandırma algoritmalarının %90 üzerinde

doğruluk sağladığını göstermiştir (Yağcı, 2022). Batool ve diğerleri, EDM tekniklerini kullanarak öğrencilerin akademik performansını tahmin etmeye yönelik bir anket çalışması gerçekleştirmişlerdir. Araştırma, EDM'nin öğretim yöntemlerini iyileştirme ve öğrencilerin başarısını artırmada etkili olduğunu bulmuştur (Batool, Rashid, Nisar, Kim, & Kwon, 2023).

2.4 Süreç Madenciliği (PM)



Şekil 2.1 - Süreç madenciliğinin üç ana türünün konumlandırılması: keşif, uygunluk ve iyileştirme. (van der Aalst, 2011)

Süreç madenciliği, ERP, CRM ve LMS gibi dijital sistemlerde toplanan olay kayıtlarını kullanarak iş süreçlerinin detaylı analizini gerçekleştiren bir yöntemdir (van der Aalst, 2012). Bu teknik, dijital sistemlerdeki kullanıcı etkileşimlerinden elde edilen verileri analiz ederek, süreçlerin mevcut işleyişini ortaya çıkarmayı ve performans iyileştirmeleri sağlamayı amaçlar. Van der Aalst ve diğerleri, süreç madenciliğini veri madenciliğinin bir alt dalı olarak tanımlamış, ancak süreç madenciliğinin gerçek süreçleri analiz etmek, izlemek ve iyileştirmek için olay günlüklerinden bilgi çıkarma üzerine odaklandığını vurgulamışlardır. Çalışmada, süreç madenciliğinin veri madenciliğinden farklı olarak desenleri yalnızca verilerden değil, süreçlerin işleyişini anlamak için olay günlüklerinden türetme amacı taşıdığı belirtilmiştir (van der Aalst, Bolt, & Van Zelst, 2017). Süreç madenciliği, organizasyonel süreçlerin görünür hale

getirilmesi ve optimize edilmesi için kritik bir araçtır. Van der Aalst, süreç madenciliğini "günümüz bilgi sistemlerinde kolayca erişilebilen olay günlüklerinden bilgi çıkararak gerçek süreçleri keşfetmeyi, izlemeyi ve iyileştirmeyi amaçlar" olarak tanımlamaktadır (van der Aalst, 2011). Süreç madenciliğinde kullanılan olay günlükleri, süreç analizinin temelini oluşturur. Olay günlükleri, bilgi sistemlerinde gerçekleştirilen işlemleri kronolojik olarak kaydeden bir veri yapısıdır ve süreçlerin davranışlarını anlamak, modellemek ve iyileştirmek için kritik bilgiler sağlar (van der Aalst, 2016). Bu kayıtlar genellikle olay kimliği, etkinlik türü, zaman damgası gibi temel özelliklerden oluşur ve bu yapı, süreç madenciliği algoritmalarının doğru ve etkili sonuçlar üretmesi için gereklidir (Rozinat & van der Aalst, 2008).

Zaman	IP Adresi	Öğrenci No	Etkinlik Türü	Ek Bilgi
8.01.2025 09:15	192.168.1.10	100123	Kaynak Görüntüleme	Konu 1
8.01.2025 09:16	192.168.1.10	100123	Kaynak Görüntüleme	Konu 2
8.01.2025 09:30	192.168.1.15	100456	Video İzleme	Ders Videosu - Bölüm 1
8.01.2025 09:45	203.123.45.67	100789	Ödev Gönderme	Ödev 1 - Çözüm Yükleme
8.01.2025 10:00	203.123.45.67	100789	Kaynak İndirme	Ders Notları.pdf
8.01.2025 10:15	145.124.25.65	100123	Forum Gönderisi	Ders Videosu - Bölüm 1
8.01.2025 10:20	155.123.23.14	100456	Klasör Görüntüleme	Pratik Dersler (Tanı)
8.01.2025 10:35	192.168.1.12	100123	Sınav Denemesi Başlatma	Sınav 1
8.01.2025 11:00	192.168.1.12	100123	Sınav Denemesi Tamamlama	Sınav 1
8.01.2025 11:30	192.168.1.15	100456	Video İzleme	Ders Videosu - Bölüm 2

Zaman Damgası

Olay Kimliği

Etkinlik Türü

Ek Bilgiler

Şekil 2.2- LMS olay günlüğü örneği

Dijital olay günlükleri her yerde bulunmaktadır ve üstel bir şekilde büyümeye devam edecektir (Manyika, et al., 2011). Bu tür verilerin her yerde bulunması, el yapımı modeller yerine gözlemlenen gerçeklere dayalı yeni süreç analizi biçimlerine olanak tanır. Süreç madenciliği için başlangıç noktası bir olay günlüğüdür (van der Aalst, 2012). Süreç madenciliği üç ana türe ayrılır: süreç keşfi, uygunluk kontrolü ve iyileştirme.

2.4.1 Süreçlerin Keşfi

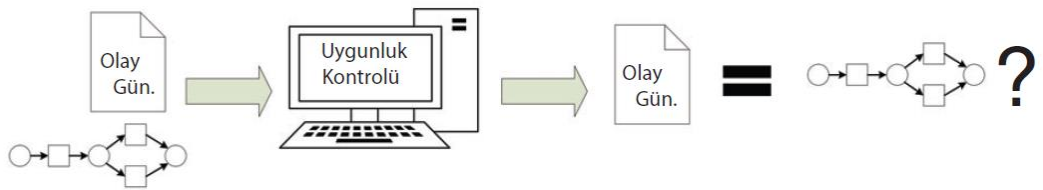


Şekil 2.3- Süreç keşfi girdi ve çıktı açısından açıklanması (Bogarín, Cerezo, & Romero, 2017)

Süreç keşfi, veri tabanındaki olay kayıtlarını kullanarak sürecin nasıl işlediğine dair bir model çıkarmayı amaçlar. Yani, veri setine dayalı olarak süreç modelleri oluşturulur ve bu modeller, sürecin işleyişini görselleştirir. Keşif madenciliği, sürecin mevcut işleyişi hakkında ön bilgiye ihtiyaç duymadan doğrudan veri tabanından çıkarım yapar. Van der Aalst, süreç madenciliği çalışmalarında süreç keşfi adımının en önemli adım olduğunu belirtmiş ve bu tekniğin olay kayıtlarından süreç modelleri elde etmek için kullanıldığını vurgulamıştır (van der Aalst, 2016).

Süreç keşfi algoritmaları, olay kayıtlarından otomatik olarak süreç modelleri çıkarmada oldukça etkilidir. Bu algoritmalar, model doğruluğunu artırma ve ölçeklenebilirlik açısından avantajlar sağlarken, aynı zamanda hesaplama karmaşıklığı ve olayların ilişkilendirilmesi için kimlik belirleyicilere bağımlılık gibi dezavantajlara sahiptir (Adams, et al., 2023). Alpha algoritması, belirli kalıpları bulmak için olay kayıtlarını tarar. Alpha algoritması basit ama verimlidir. Ancak, karmaşık yönlendirme yapıları ve gürültülerle başa çıkma konusunda sorunlar yaşamaktadır (Roci & Davidrajuh, 2018). Fuzzy miner (Potavin, Jongswat, & Premchaiswadi, 2012) ve Heuristic miner (Sarno, Haryadita, Sunaryono, & Munif, 2015) algoritmaları gürültü ve eksik veri ile başa çıkmada başarılı algoritmalarlardır.

2.4.2 Uygunluk Kontrolü

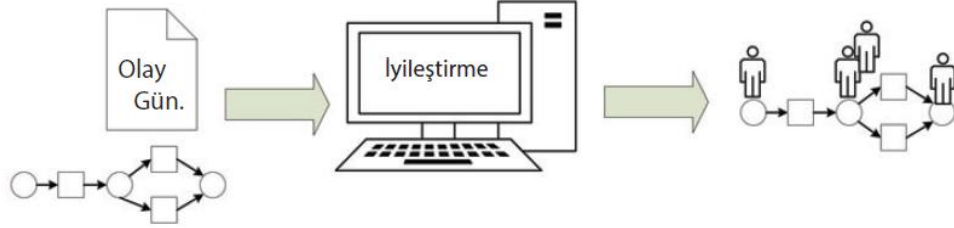


Şekil 2.4- Süreç uygunluk kontrolü girdi ve çıktı açısından açıklanması (Bogarín, Cerezo, & Romero, 2017)

Uygunluk kontrolü, önceden keşfedilmiş süreç modelleri ile gerçek olay kayıtları arasındaki farkları bulmayı amaçlar (van der Aalst, 2016). Bu teknik, belirli bir sürecin nasıl işlemesi gerektiğine dair modelle (beklenen model) gerçek işleyiş arasında karşılaştırma yapar. Uygunluk kontrolü, sürecin mevcut işleyişi ile olması gereken işleyiş arasındaki sapmaları ortaya koyarak, süreçlerin doğruluğunu ve uygunluğunu

kontrol eder. Rozinat ve van der Aalst, uygunluk kontrolü ile süreç modelleri ve olay kayıtları arasında doğrulama yapılmasının süreçlerin performansını ölçmede ve iyileştirme alanlarını belirlemede etkili olduğunu belirtmiştir (Rozinat & van der Aalst, 2008).

2.4.3 Süreçlerin İyileştirilmesi



Şekil 2.5- Süreçlerin iyileştirmesi girdi ve çıktı açısından açıklanması (Bogarín, Cerezo, & Romero, 2017)

İyileştirme, süreçlerin performansını artırmayı amaçlar. Bu teknik, mevcut süreç modellerinin iyileştirilmesi için kullanılır ve olay kayıtlarından elde edilen veriler kullanılarak süreçlerin yeniden yapılandırılmasını sağlar. İyileştirme madenciliği, süreç performansını daha verimli hale getirmek için süreç modeline yeni bilgiler veya performans metrikleri eklemeyi hedefler. Van der Aalst, iyileştirme madenciliğinin süreç performansını optimize etmek ve süreç modellerine yeni bilgiler eklemek için kullanıldığını belirtmiştir (van der Aalst, 2011).

Keşif, uygunluk ve iyileştirme türleri birbirini tamamlayan analiz adımlarıdır. Örneğin, keşif madenciliği ile çıkarılan modeller, uygunluk kontrolü ile doğrulanabilir ve iyileştirme madenciliği ile optimize edilebilir (Rozinat & van der Aalst, 2008).

2.4.4 Süreç Madenciliğinde Kullanılan Araçlar

Süreç madenciliğinde süreç keşfi için birçok algoritma kullanılmaktadır. Bölüm 2.4.1 içerisinde kullanılan popüler algoritmalara değinilmiştir. Süreç madenciliğinde en yaygın kullanılan platformlardan biri olan ProM, açık kaynaklı bir araçtır ve süreç keşfi, uygunluk kontrolü ve iyileştirme gibi çok sayıda analitik işlemi destekler (van der Aalst, Dongen, Günther, Verbeek, & Weijters, 2009). Bir diğer popüler araç olan Disco, kullanıcı dostu arayüzü ve yüksek performanslı algoritmalarıyla süreç madenciliği projelerinde özellikle veri keşfi ve görselleştirme açısından sıklıkla tercih edilmektedir (Günther & Rozinat, 2012). Bir diğer uygulama Celonis, süreç

madenciliği alanında lider bir platform olup, gerçek zamanlı süreç analizi ve iyileştirme olanakları sağlayan kullanıcı dostu bir araçtır. Gelişmiş veri görselleştirme özellikleri ve kurumsal entegrasyon yetenekleri sayesinde Celonis, süreç madenciliği projelerinde sıklıkla tercih edilen bir çözüm sunmaktadır (van der Aalst, 2018).

ProM, esnek bir açık kaynak platform olmasına karşın, büyük veri setleri üzerinde işlem süresi açısından sınırlamalar yaşayabilir. Disco, daha hızlı ve kullanıcı dostu bir arayüz sunarken, Celonis, güçlü görselleştirme araçları ve kurumsal entegrasyon yetenekleriyle öne çıkar (Günther & Rozinat, 2012) (van der Aalst, 2018).

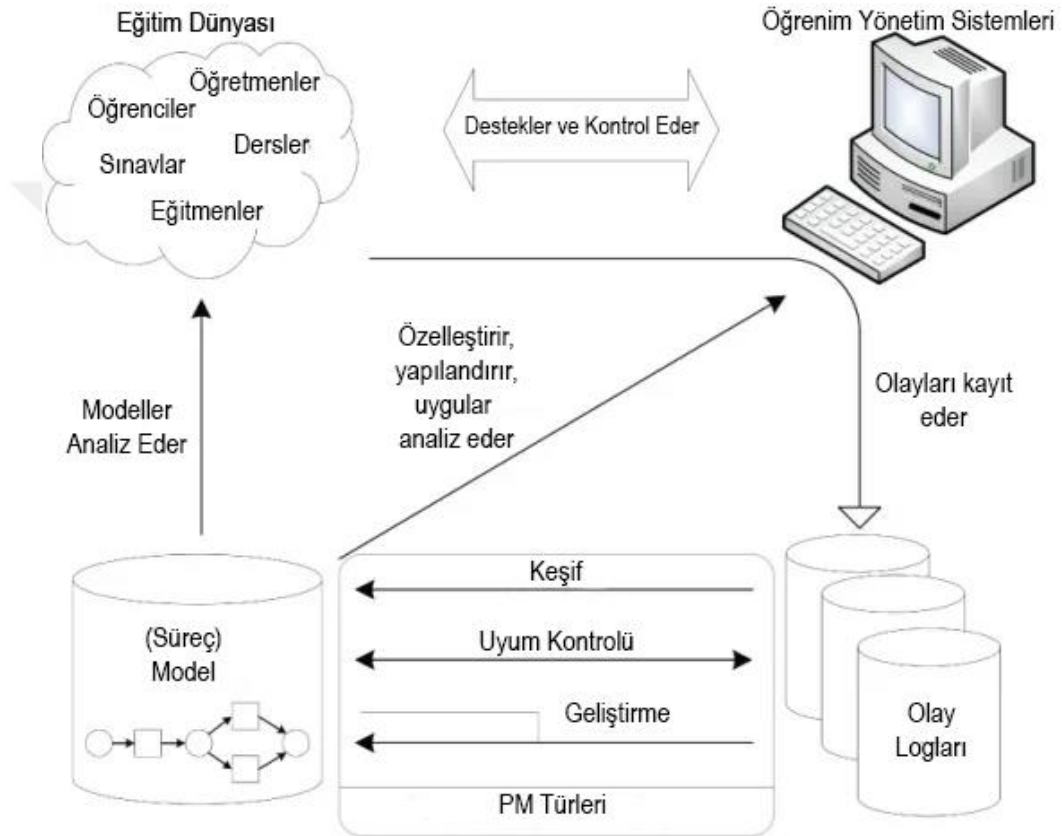
Süreç madenciliği yazılımları arasında ProM, Disco ve Celonis farklı ama tamamlayıcı yetenekler sunmaktadır. Çelik ve Ayçetin tarafından yapılan araştırmaya göre, ProM açık kaynaklı doğasıyla akademik esneklik sunarken, Disco basit kullanımı ve hızlı modelleme yetenekleriyle iş kullanıcıları için uygundur. Celonis ise kurumsal düzeyde büyük veri işleme kapasitesi ve entegrasyon yetenekleriyle dikkat çekmektedir. Ancak, her aracın belirli sınırlamalarının olduğu vurgulanmaktadır ve seçim, kullanım amacına göre yapılmalıdır (Çelik & Akçetin, 2018).

Farklı birçok yazılım bulunmasına rağmen, ProM, Disco ve Celonis gibi araçlar süreç madenciliği alanında akademik araştırmalarda yaygın olarak tercih edilmektedir. Bu çalışmada, süreç analizi ve modelleme süreçlerini gerçekleştirmek amacıyla, bu popüler yazılımlardan faydalanılacaktır.

2.5 Eğitimsel Süreç Madenciliği (EPM)

Eğitimsel süreç madenciliği, eğitim süreçlerinin dijital ortamlarda toplanan öğrenci davranış verileriyle analiz edilmesini amaçlayan, hızla gelişmekte olan bir araştırma alanıdır. Bu yaklaşım, özellikle Öğrenme Yönetim Sistemleri (LMS) gibi platformlardan elde edilen verilerden, öğrencilerin öğrenme süreçlerini ve performanslarını anlamak, modellemek ve tahmin etmek için kritik bilgiler sağlamaktadır. Eğitimsel süreç madenciliği, öğrencilerin öğrenme davranışlarını daha derinlemesine anlamak ve kişiselleştirilmiş öğrenme stratejileri geliştirmek için yenilikçi çözümler sunar (Baker & Siemens, 2014). EPM, özellikle sorunlu öğrencilerin öğrenme süreçlerindeki engelleri tespit ederek öğretim stratejilerini optimize etme imkânı sunar. Bu bağlamda, Bogarín ve diğerleri tarafından yapılan bir

çalışmada, Moodle platformunda öğrencilerin gezinme yolları keşfedilerek, farklı öğrenci gruplarının davranış modelleri analiz edilmiştir. Bu analizler, öğrencilerin öğrenme süreçlerindeki potansiyel zorlukların belirlenmesine ve buna bağlı olarak öğretim stratejilerinin iyileştirilmesine katkı sağlamaktadır (Bogarín, Cerezo, & Romero, 2015). EPM, sadece öğrenci başarısını artırmakla kalmaz, aynı zamanda öğretmenlere ve yöneticilere veri odaklı kararlar alma konusunda önemli bir destek sağlar (Siemens & Long, 2011).



Şekil 2.6- Eğitimsel Süreç Madenciliği Çerçevesi: türleri ve bileşenleri (Bogarín, Cerezo, & Romero, 2017)

Literatürde yapılan çalışmalar, DM ve PM tekniklerinin öğrenci başarısını tahmin etme, öğrenme stratejilerini belirleme ve eğitimi optimize etme konularında etkili olduğunu göstermektedir (Romero & Ventura, 2010).

Chen ve diğerleri, süreç madenciliği tekniklerini kullanarak öğrenme aktivitelerini optimize etmek için hedef odaklı bir navigasyon sistemi geliştirmiştir. Araştırma, öğrencilerin bireysel hedeflerine ulaşmasını destekleyecek araçlar sunmaktadır (Chen, Man, Jin, & Huang, 2011).

Awatef ve diğerkleri, süreç madencilii ile öğrencilerin öğrenme alışkanlıklarını ve performans faktörlerini analiz ederek eğitim müfredatlarının geliştirilmesine yönelik bulgular elde etmişlerdir. Araştırma, eğitimde sosyal ağ madencilii ve kümeleme yaklaşımları ile performans göstergelerini değerlendirmiştir (Hicheur Cairns, et al., 2015).

van der Aalst, süreç madenciliiğinin veri bilimi uygulamalarını ele alarak, öğrenme davranışlarının süreç modellerine dönüştürülmesi ve öğrencilerin ihtiyaçlarının daha iyi anlaşılmasını sağlamıştır (van der Aalst, 2016).

Grigороva ve diğerkleri, süreç madenciliiği ve veri madenciliiği yaklaşımlarını e-öğrenme süreçlerini iyileştirmek için kullanmışlardır. Araştırma, süreç modellerini geliştirmek ve eğitimsel süreçlerin verimliliğini artırmak için öneriler sunmaktadır (Grigороva, Malysheva, & Bobrovskiy, 2017).

Bogarín ve diğerkleri tarafından yapılan çalışmada, öğrencilerin öğrenme yollarını modelleyerek bireyselleştirilmiş öğrenme stratejileri geliştirilmesine yönelik bulgulara erişilmiştir (Bogarín, Cerezo, & Romero, 2015).

Bonde ve Kirange, eğitim verilerini analiz ederek öğrenci performansını etkileyen faktörleri sınıflandırmış ve tahmin modelleri geliştirmişlerdir. Çalışma, yükseköğretimde öğrencilerin başarısını artıracak stratejiler önermektedir (Bonde & Kirange, 2018).

Intayoad ve diğerkleri, süreç madenciliiği tekniklerini kullanarak öğrencilerin öğrenme yollarını keşfetmiş ve kişiselleştirilmiş öğrenme modelleri geliştirmiştir. Araştırma, öğrenme süreçlerinin görselleştirilmesiyle öğrencilerin akademik başarılarını artırmayı hedeflemiştir (Intayoad, Kamyod, & Temdee, 2018).

Dolak, LMS Moodle olay günlüklerini analiz etmek için süreç madenciliiği tekniklerini kullanarak öğrenci aktivitelerini ve navigasyon yollarını keşfetmiştir. Bulgular, öğrenme davranışlarının daha iyi anlaşılmasını desteklemiştir (Dolak, 2019).

AlQaheri ve Panda, otomatik süreç madenciliiği tekniklerini kullanarak öğrencilerin öğrenme davranışlarını analiz etmiş ve öğretim kalitesini artırmaya yönelik bir çerçeve

önermiştir. Araştırma, sınıf aktivitelerini ve performans verilerini değerlendirerek etkili öğrenme modelleri geliştirmiştir (AlQaheri & Panda, 2022).

Loder, yüksek öğrenci terk oranlarını azaltmak amacıyla eğitim süreç madenciliği tekniklerini kullanarak Graz Üniversitesi'nde 2012/13 ile 2022/23 yılları arasındaki 53.839 öğrenci ve 78.495 çalışmanın verilerini analiz etmişlerdir. Çalışmada, terk eden ve mezun olan öğrencilerin ders sıralamalarını karşılaştırmak için etkileşimli bir web uygulaması geliştirilmiştir. Süreç haritaları, paydaş gruplarıyla yapılan tartışmalarda ve müfredat tasarımı süreçlerinde kullanılmıştır. Bulgular, süreç haritalarının öğrencilerin sınav sıralamalarını gerçek zamanlı olarak karşılaştırma ve izleme imkânı sağladığını, özellikle erken dönem terk değerlendirmesi için etkili bir araç olduğunu göstermiştir (Loder, 2024).

Diamantini ve diğerleri, İtalyan üniversitelerindeki lisans programlarını süreç olarak modelleyerek, öğrencilerin başarısızlık ve gecikme nedenlerini analiz etmişlerdir. Araştırmada süreç analizi teknikleri kullanılarak mezuniyet sürelerini etkileyen darboğazlar tespit edilmiş ve başarılı öğrenciler ile geç mezun olan öğrencilerin kariyer yolları karşılaştırılmıştır. Elde edilen bulgular, derslerde ortaya çıkan darboğazların nedenlerini anlamaya ve müfredatın genel iyileştirilmesine yönelik önerilerin geliştirilmesine önemli katkılar sağlamıştır. (Diamantini, Genga, Mircoli, Potena, & Zannone, 2024).

Yazar ve diğerleri, eğitim bilgi sistemlerindeki büyük ölçekte artan olay günlüklerini analiz etmek ve iyileştirmek amacıyla Eğitimsel Süreç Madenciliği'ni (EPM) kullanarak bir çalışma gerçekleştirmiştir. Çalışmalarında Fuzzy miner, Integer Linear Programming miner, Heuristics miner ve Inductive miner gibi algoritmaları uygulayarak veri setinden süreç modelleri çıkarmışlardır. Elde edilen sonuçlar, basitlik, uyum ve doğruluk gibi metriklerle değerlendirilmiş ve süreç modeli, öğrenciler için öğrenme yolları rehberliği, öğretim programlarının iyileştirilmesi ve eğitim yönetimi için referans önerileri sunacak şekilde geliştirilmiştir. Araştırma, süreç madenciliğinin eğitim süreçlerinde yenilikçi çözümler sunma potansiyelini vurgulamaktadır (Feng & Chen, 2025).

2.5.1 EPM ve ML Entegrasyonu

Eğitimsel Süreç Madenciliği (EPM), olay verilerini analiz etmek için halihazırda yeterince güçlü bir yaklaşımdır; ancak, EPM'nin en yeni makine öğrenmesi yöntemleriyle birleştirilmesi, çeşitli yeni yönere genişletilebilme potansiyeline sahiptir. Bu birleşim, daha bilgilendirici veri odaklı içgörüler elde edilmesine ve öğretim ile öğrenme sonuçlarını geliştirebilecek makine öğrenmesi modellerinin geliştirilmesine olanak sağlayabilir (Patil, 2021). Eğitimsel veri madenciliğinin temel amacı—öğrenciler için eğitim süreçlerini iyileştirmek ve öğrenme çıktılarını optimize etmek (Yağcı, 2022) — göz önüne alındığında, eğitimsel verilerin analizi ile eğitimsel süreç madenciliğinin birleştirilmesi, bu etkilerin daha da artırılmasını sağlayabilir.

Umer ve diğerleri, açık çevrimiçi derslerde (MOOCs) öğrencilerin öğrenme süreçlerini iyileştirmek ve düşük tamamlama oranlarını azaltmak için süreç madenciliği özelliklerini içeren veri odaklı bir yaklaşım önermişlerdir. Çalışmalarında, geleneksel ve süreç uygunluk testi özelliklerinden elde edilen iki farklı veri seti ile lojistik regresyon, Naïve Bayes, rastgele orman ve K-en yakın komşu gibi makine öğrenmesi tekniklerini kullanarak öğrencilerin performansını tahmin etmişlerdir. Sonuçlar, süreç madenciliği ile zenginleştirilmiş özelliklerin, makine öğrenmesi tekniklerinin doğruluğunu anlamlı ölçüde artırdığını ve lojistik regresyon ile Naïve Bayes'in diğer tekniklere üstün geldiğini ortaya koymuştur. Araştırma, öğrencilerin kurs boyunca desteklenmesine ve başarısızlık riski taşıyanların erken tespitine yönelik önemli içgörüler sunmaktadır (Umer, Susnjak, Mathrani, & Suriadi, 2017).

Feng ve diğerleri, modern eğitim bilgi sistemlerinden elde edilen verilerin süreç madenciliği ve makine öğrenmesi teknikleri ile analiz edilerek öğrenci davranış kalıplarını anlamayı ve bu davranışlara dayalı tahmin modelleri geliştirmeyi amaçlamışlardır. Çalışmalarında öğrenme davranış kalıplarını tahmin etmek için, destek vektör makineleri ve çeşitli optimizasyon yöntemleri kullanılmış; bulgular, önerilen yarı yarıya istatistik yönteminin kümeleme etkisini iyileştirdiğini ortaya koymuştur (Feng, Fan, & Ao, 2022).

Liu ve diğerleri, olay günlüklerinden elde edilen bilgilerle öğrencilerin öğrenme süreçlerini proaktif bir şekilde iyileştirmek için iki aşamalı bir problem çözme beceri modelleme yaklaşımı önermişlerdir. Çalışmalarında Gradient Boosting Decision Tree

(GBDT) algoritmasını kullanarak öğrencilerin ders öncesi bilgi düzeyini belirlemek için "ön model" ve ders sonrası öğrenme çıktısını analiz etmek için "sonraki model" geliştirmişlerdir. Deneysel sonuçlar, bu modellerin yüksek sınıflandırma doğruluğuna sahip olduğunu ve öğrencilerin büyük bir çoğunluğunun ders sonunda problem çözme becerilerinde önemli gelişmeler kaydettiğini ortaya koymuştur (Liu, ve diğerleri, 2022).

Racca ve diğerleri, süreç madenciliği yöntemlerini kullanarak bilgisayar programlama eğitiminde öğrenme süreçlerinin değerlendirilmesi üzerine bir çalışma gerçekleştirmiştir. Araştırmada, NetLogo ile rehberli bir egzersiz sırasında öğrencilerin davranış verileri (örneğin, sayfalarda geçirilen süre, fare hareketleri ve tıklamalar) süreç keşfi teknikleri ile analiz edilmiştir. Öğrenme çıktılarının görselleştirilmesine odaklanan çözüm, ısı haritaları ve doğrudan takip grafikleri içerirken, ilk bulgular süreç madenciliği tekniklerinin öğrenme sürecinin değerlendirilmesini iyileştirme potansiyelini doğrulamaktadır (Racca, Sulis, & Capecchi, 2022).

Cenka ve diğerleri, süreç madenciliği teknikleriyle LMS verilerinden elde edilen öğrenci davranış modellerini analiz etmiştir. Çalışma, süreç madenciliği ile elde edilen verilerle destek vektör makineleri (SVM) ve karar ağaçları gibi algoritmaların bir arada kullanımını içermektedir. Sonuçlar, SVM'nin özellikle aktif öğrencilerin akademik başarısını tahmin etmede etkili olduğunu göstermiştir (Cenka, Santoso, & Junus, 2022).

Khan ve diğerleri, LMS günlüklerinden öğrenci performansını tahmin etmek amacıyla bir çalışma gerçekleştirmiştir. Çalışmada, rastgele orman, destek vektör makineleri (SVM), karar ağaçları ve k-en yakın komşu (k-NN) gibi çeşitli makine öğrenmesi algoritmaları kullanılmıştır. Sonuçlar, rastgele orman algoritmasının diğer yöntemlere kıyasla daha yüksek doğruluk oranı sağladığını göstermiştir. Bu çalışma, öğrencilerin başarı tahmininde LMS verilerinin etkili bir veri kaynağı olduğunu ortaya koymuştur (Khan, ve diğerleri, 2023).

Literatürde LMS olay loglarının analizi genellikle temel kullanım istatistikleriyle sınırlı kalmıştır. Çoğu çalışma, öğrencilerin giriş-çıkış zamanları, materyal erişimleri

veya sınav tamamlama oranları gibi yüzeysel metriklerle odaklanmakta, ancak öğrenci davranışlarının süreç madenciliği teknikleriyle detaylı bir şekilde modellenmesini ihmal etmektedir. Süreç madenciliği ve makine öğrenmesi entegrasyonu konusunda yapılan çalışmalar oldukça sınırlıdır ve bu iki alan genellikle birbirinden bağımsız olarak ele alınmaktadır. Bu bağlamda, bu çalışma, LMS olay loglarından elde edilen süreç modellerinin makine öğrenmesi algoritmalarına entegre edilmesiyle literatürdeki bu boşluğu doldurmayı hedeflemektedir. Bu yaklaşım, öğrenci davranışlarının zamansal ve dinamik yönlerini anlamayı, akademik başarı tahminlerini iyileştirmeyi ve kişiselleştirilmiş öğrenme stratejileri geliştirmeyi amaçlamaktadır.



ÜÇÜNCÜ BÖLÜM

3. ARAŞTIRMA TASARIMI VE YÖNTEMİ

Bu bölümde, çalışmanın temel veri kaynağı olan öğrenme yönetim sistemi (LMS) olay günlüklerinin toplanma süreci ve bu verilerin analiz için hazırlanması açıklanacaktır. LMS günlükleri, öğrencilerin sisteme giriş-çıkış zamanları, quiz tamamlama oranları ve materyal etkileşimleri gibi etkinlikleri kayıt altına alarak çalışmanın temel veri setini oluşturmuştur. Bu veri setinin analiz için kullanılabilir hale gelmesi amacıyla eksik veri doldurma, aykırı değer temizleme ve normalizasyon gibi ön işleme adımları gerçekleştirilmiştir. Ayrıca, özellik mühendisliği yöntemleri ile veri setine yeni özellikler eklenmiş ve makine öğrenmesi algoritmalarına uygun hale getirilmiştir.

3.1 Veri Kaynağı Özellikleri

Bu çalışma kapsamında kullanılan veriler, bir öğrenme yönetim sistemi (LMS) üzerinden toplanmış ve öğrencilerin öğrenme süreçlerini detaylı bir şekilde analiz etmek amacıyla kullanılmıştır. Veriler, öğrencilerin platform üzerindeki etkileşimlerini ve davranışlarını yansıtan olay günlükleri oluşmaktadır.

3.1.1 Veri Kaynağı

Veriler, üniversitenin eğitim faaliyetlerinde kullanılan bir LMS (Sakai) platformundan elde edilmiştir. Veriler, platformda gerçekleştirilen kullanıcı eylemlerini ve etkinliklerini kaydeden olay günlükleri şeklinde toplanmıştır. Toplama süreci aşağıdaki adımları içermektedir:

- Eğitimsel Veri Madenciliği (EDM) kapsamında analiz edilen temel veri türleri listelenmiştir. Bu veriler, öğrenci davranışlarını, akademik performanslarını ve öğrenme süreçlerini anlamak amacıyla LMS (Öğrenme Yönetim Sistemi) ve diğer eğitim platformlarından elde edilmiştir. EDM teknikleri, bu veriler üzerinden öğrenme desenlerini keşfetmek, akademik başarıyı tahmin etmek ve kişiselleştirilmiş öğrenme stratejileri geliştirmek için kullanılır. Tablo, demografik bilgilerden sosyal etkileşimlere, etkinlik sürelerinden tamamlama durumlarına kadar geniş bir veri yelpazesi sunmaktadır ve her veri türü literatürdeki etkili çalışmalarla desteklenmiştir.

Tablo 3.1 EDM Veri Seti Özellikleri

Veri Türleri	Açıklama	Kaynaklar
Demografik Bilgiler	Öğrencilerin yaş, cinsiyet gibi genel özellikleri.	(Romero & Ventura, 2020)
Akademik Performans	Öğrencilerin devamsızlık durumları, sınava katılımları ve ödev gönderme.	(Baker & Siemens, 2014)
Etkinlik Frekansı	LMS'deki giriş-çıkış sayısı, kaynaklara erişim sıklığı gibi kullanıcı etkinlikleri.	(Batool, Rashid, Nisar, Kim, & Kwon, 2023)
Etkinlik Türleri	Materyal görüntüleme, ödev teslimi, quiz tamamlama gibi spesifik eylemler.	(Yağcı, 2022)
Öğrenme Materyali İndirme	Öğrencilerin PDF, video veya diğer materyalleri indirme etkinlikleri.	(Strielkowski, Grebennikova, Lisovskiy, Rakhimova, & Vasileva, 2024)
Oturum Süresi	Öğrencinin LMS'de oturumda geçirdiği toplam süre.	(Romero & Ventura, 2020)
Tamamlama Durumu	Bir kurs veya modülün tamamlanıp tamamlanmadığı bilgisi.	(AlQaheri & Panda, 2022)

- **Olay Günlükleri:** LMS'nin kayıt sisteminden öğrenci davranışlarını yansıtan zaman damgaları, etkinlik türleri, kaynaklar ve diğer ilgili bilgiler elde edilmiştir. Veriler, eğitimsel süreç madenciliği ve veri madenciliği teknikleriyle işlenerek öğrenme süreçlerinin modellenmesi ve akademik performans tahmini için kullanılmaktadır. Her bir olay türü, kaynaklara ve literatürdeki yüksek etkili çalışmalara dayalı olarak açıklanmıştır.

Tablo 3.2 Eğitimsel Süreç Madenciliği için LMS Olay Günlükleri Özellikleri

Etkinlik	Olay	Açıklama	Kaynaklar
Sisteme Giriş	user_loggedin	Kullanıcının sisteme giriş yapması.	(Romero & Ventura, 2010)

Matertal Görüntüleme	content.read Gradebook.student View	Ders materyalinin incelenmesi. Video içeriklerinin izlenmesi.	(Baker & Siemens, 2014) (Dolak, 2019)
Sınav Katılımı	exam.midterm exam.final exam.makeup	Öğrencinin katılım bilgisi.	(Yağcı, 2022)
Duyuru & Ders İzlenesi Takibi	Announcement.read syllabus.read	Yapılan duyuruların takip frenkansı.	(Yağcı, 2022)
Ödev Teslimi	asn.read.assignme nt asn.new.submissio n asn.save.submissi on asn.revise.submis sion asn.read.submissi on asn.submit.submis sion	Ödevin LMS'e yüklenmesi.	(Batool, Rashid, Nisar, Kim, & Kwon, 2023)
Geçme / Kalma	status.pass status.fail	Dersten geçme kalma durumu	

3.1.2 Toplanan Verilerin Miktarı ve Türleri

Bu çalışmada kullanılan veri seti, Sakai LMS'ten elde edilen 13.302 adet işlem kaydı ve bunlarla ilişkilendirilmiş toplam 101 öğrencinin akademik ve davranışsal bilgilerini içermektedir. Veri setinde 48 öğrenci başarılı olurken, 53 öğrenci başarısız olmuştur. Öğrenci davranışları içerik okuma, ödev teslimi, duyuru takibi gibi eylemler üzerinden sayısallaştırılmış ve her bir öğrenciye ilişkin 13 değişken elde edilmiştir.

- 101 öğrenciden oluşan anonimleştirilmiş veri seti,
- LMS sisteminden alınmış 13.302 olay (olay günlüğü),
- Bu olaylara ait 13 özellik (demografik, davranışsal, sınav katılımı, etiket bilgileri)
- Sınıf dağılımı:
 - Başarılı (status = 1): 48 öğrenci
 - Başarısız (status = 0): 53 öğrenci

Ek Özellikler:

- course_attend_ratio: LMS’e aktif katılım oranı (0–1)
- avg_events_per_day: Günlük ortalama etkileşim sayısı
- participated_*: Vize, final ve büt sınavına katılım bilgileri (0/1)
- gender, age: Demografik değişkenler
- status: Öğrencinin ders sonucunda başarılı (1) veya başarısız (0) olması

Tablo 3.3 Verilerin İstatistiksel Değerlendirmeleri

Değişken	Açıklama	Yorum
gender	0: Kadın, 1: Erkek	Veride %77 erkek öğrenci yer alıyor. Cinsiyet dengesi asimetrik.
age	Yaş (tamsayı)	Öğrenciler 18–24 yaş aralığında, ortalama ≈ 19.7 .
course_attend_ratio	LMS’e etkinlik katılım oranı (0–1)	Ortalama oran ≈ 0.69 . Bu, öğrencilerin %69 oranında derse katıldığını gösteriyor.
content_read_count	İçerik okuma sayısı	Geniş dağılım var (0–107). Ortalama ≈ 33 .
assignment_submit_count	Teslim edilen ödev sayısı	Maksimum 6 ödev, ortalama ≈ 3.6 .

3.1.3 Veri Türleri

Toplanan veriler, aşağıdaki kategorilere ayrılabilir:

- **Demografik Bilgiler:**
 - Öğrenci ID’si: Anonimleştirilmiş, eşsiz tanımlayıcı.
 - Yaş: 18 ile 24 arasında değişmektedir, ortalama ≈ 19.7 .
 - Cinsiyet: Binary (0: Kadın, 1: Erkek) olarak temsil edilmektedir.
- **Davranışsal Veriler:**

Sakai LMS’den çıkarılmış 13.302 işlem den oluşur. Örnek olaylar:

Tablo 3.4 - Veri Setinde Kayıtlı Olay Sayıları

Özellik	Açıklama	Toplam
content_read_count	Okunan içerik sayısı (PDF, sayfa vb.)	3.402
assignment_submit_count	Teslim edilen ödev sayısı	364
assignment_read_count	Okunan ödev tanımı sayısı	1.027
assignment_completion_count	Tamamlanan ödev sayısı	364
annc_read_count	Okunan duyuru sayısı	2.356

- **Performans Verileri:**

- Öğrencilerin dönem içi sınavlara (vize, final, bütünleme) katılım durumları (participated_midterm, participated_final, participated_makeup),
- Modelleme amacıyla oluşturulan başarı etiketi (status: pass/fail),
- Derse katılım (course_attend_ratio) gibi ek açıklayıcı değişkenler olarak sistemde yer almıştır.

3.1.4 Anonimleştirme ve Veri Güvenliği

Kullanıcıların gizliliğini korumak amacıyla tüm kişisel bilgiler (ör. öğrenci numarası, Ad soyad) anonim hale getirilmiştir. Bu süreç, Kişisel verileri koruma kanunu (KVKK) ilkelerine uygun olarak gerçekleştirilmiştir. Verilerin güvenli bir şekilde işlenmesi ve analiz edilmesi için aşağıdaki adımlar uygulanmıştır:

- **Anonimleştirme:** Tüm kullanıcı kimlikleri (ör. öğrenci ID'si) rastgele kodlarla değiştirilmiştir.
- **Veri Şifreleme:** LMS'den çıkarılan ham veriler, güvenli bir sunucuda şifrelenmiş olarak saklanmıştır.
- **Yetkilendirme:** Verilere yalnızca bu çalışma için yetkilendirilmiş araştırma ekibi erişebilmiştir.

3.1.5 Verinin Geçerliliği ve Güvenilirliği

Toplanan veriler, LMS sisteminin standart kayıt mekanizmaları kullanılarak toplanmıştır. Bu nedenle, veri setinin doğruluğu ve güvenilirliği yüksektir. Ayrıca,

toplanan veri miktarı, süreç madenciliği ve makine öğrenmesi analizleri için yeterli bir büyüklüğe sahiptir.

3.2 Veri Toplama ve Önleme

Veri ön işleme disiplini, ham veriyi makine öğrenimi algoritmaları tarafından anlaşılabilir hale getirmek amacıyla uygulanır ve makine öğrenimi görevlerinin en kritik ve önemli aşaması olarak yaygın bir şekilde kabul edilmektedir. Veri ön işlemenin başarılı bir şekilde gerçekleştirilmesi, doğruluğu artırabilir ve modelin kararlılığını sağlayabilir (Olisah, Smith, & Smith, 2022). Veri ön işlemenin önemi, eğitim verileri söz konusu olduğunda azalmamakta; aksine, eğitim verileri üzerinde makine öğrenimi uygulandığında çok daha büyük bir önem kazanmaktadır (Hilbert, ve diğerleri, 2021).

3.2.1 Veri Toplama

Bu çalışmada, öğrenme yönetim sistemi (LMS) aracılığıyla toplanan üç farklı veri seti kullanılmıştır. Her bir veri seti, araştırma sorularını yanıtlamak ve eğitimsel süreçleri detaylı bir şekilde analiz etmek için farklı amaçlara hizmet etmiştir. Veriler anonimleştirilerek etik kurallar doğrultusunda toplanmıştır.

- **Eğitimsel Veri Madenciliği Veri Seti:** Moodle LMS'ten elde edilen bu veri seti, öğrencilerin akademik performanslarının belirlenmesinde temel oluşturmuştur. Kullanılan veriler şunlardır:
 - **Demografik Bilgiler:** Öğrenci yaşı, cinsiyet (Romero & Ventura, 2010).
 - **Etkinlik Verileri:** Sisteme giriş-çıkış zamanları, sınavlara girme durumları, duyuru kuma sayıları, kaynak okuma ve indirme sayıları, derse katılım oranları (Baker & Siemens, 2014).
- **Eğitimsel Süreç Madenciliği Veri Seti:** Bu veri seti, öğrencilerin öğrenme davranışlarını ve öğrenme yollarını keşfetmek için kullanılan olay günlüklerinden oluşmaktadır. Süreç madenciliği araçları (örneğin, ProM) ile analiz edilen bu loglar şu verileri içermektedir:
 - **Zaman Damgaları (Timestamps):** Öğrencilerin etkinlik başlama zamanları.
 - **Etkinlik Türleri:** Sınav tamamlama, kaynak okuma, materyal indirme gibi eylemler.
 - **Süreç Akışları:** Bir etkinlikten diğerine geçiş sıklıkları ve süreleri.

Bu veri seti, öğrenme süreçlerindeki davranış kalıplarını ortaya çıkarmak ve sapmaları analiz etmek için temel teşkil etmektedir (van der Aalst, 2016).

- **Birleştirilmiş Veri Seti:** İlk iki veri setinin birleştirilmesiyle oluşturulan bu veri seti, öğrenme davranışlarının ve süreç keşfi sonuçlarının makine öğrenmesi algoritmaları üzerindeki etkisini araştırmak için kullanılmıştır. Bu veri seti şu verileri içermektedir:
 - Süreç madenciliği ile elde edilen davranış özellikleri (örneğin, süreç uygunluk (fitness score)).
 - LMS tabanlı etkinlik özellikleri (örneğin, toplam sınav tamamlama oranı, forum etkileşim sayısı).

Birleştirilmiş veri seti, klasik eğitimsel veri madenciliği yöntemleriyle süreç madenciliği tekniklerinin entegrasyonunun etkilerini araştırmak için tasarlanmıştır.

3.2.2 Eksik Veri Doldurma

Eksik veri sorununu gidermek için literatürde yaygın olarak ortalama, medyan veya istatistiksel tahminleme yöntemleri kullanılmaktadır (Kang, 2013). Bunun yanı sıra, daha karmaşık veri setlerinde K-Nearest Neighbors (KNN) tabanlı imputation yöntemi öne çıkmaktadır. Bu yöntem, eksik verileri tamamlamak için benzer davranışlar sergileyen diğer veri noktalarının bilgilerini kullanır (Troyanskaya, ve diğerleri, 2001).

Bu çalışmada, KNN imputation yöntemi aşağıdaki adımlarla uygulanmıştır:

1. Eksik veri içeren gözlemler tespit edilmiştir.
2. Her bir eksik gözlem için, veri setindeki en yakın komşular belirlenmiştir.
3. Eksik değerler, seçilen komşuların ağırlıklı ortalamalarıyla doldurulmuştur.

KNN imputation, veri setinde eksiklikler nedeniyle oluşabilecek bilgi kaybını azaltmış ve veri analizinin doğruluğunu artırmıştır.

3.2.3 Aykırı Değer Temizleme

Aykırı değerler, öğrenci davranışlarının tutarlı bir şekilde analiz edilmesi için temizlenmiştir. Literatürde sıklıkla kullanılan iki temel yöntem Z-Skoru ve IQR (Interquartile Range) filtreleme yöntemleridir (Tukey, 1977). Bu çalışmada Z-Skoru yöntemi uygulanmış, $Z > 3$ veya $Z < -3$ olan değerler aykırı olarak işaretlenmiştir. Bu

sınır, normal dağılımda tüm veri noktalarının %99.7'sinin -3 ve +3 arasında olması prensibine dayanır (Chandola, Banerjee, & Kumar, 2009).

Z-skoru, bir veri noktasının, dağılımdaki ortalama dan kaç standart sapma uzaklıkta olduğunu ölçen istatistiksel bir yöntemdir. Z-skoru, özellikle normal dağılıma sahip veri setlerinde aykırı değerlerin tespiti için etkili bir araçtır. Formülasyonu şu şekildedir:

$$Z = \frac{(X - \mu)}{\sigma}$$

Denklem 3.1- Z-skoru

Burada:

- Z: Z-skoru değerini ifade eder.
- X: Değerlendirilen veri noktası.
- μ : Veri setinin ortalaması.
- σ : Veri setinin standart sapması.

Z-skoru yöntemi ile yapılan aykırı değer analizinde yalnızca age ve content_read_count değişkenlerinde düşük oranlı (%2'nin altında) sapmalar gözlenmiştir. Diğer tüm davranışsal değişkenler, ortalama çevresinde simetrik dağılmaktadır. Bu durum, veri setinin genel olarak homojen olduğunu ve modelleme sürecine doğrudan uygun olduğunu göstermektedir.

Veri Setlerinde Aykırı Değer Örnekleri;

Anormal Yüksek Sınav Puanları: Bir öğrencinin derse katılımları sınırlı, duyuru takibi veya ödev teslimi az olmasına rağmen yüksek not alması veri giriş hatası veya kopya çekme gibi durumları gösterebilir.

Aşırı Uzun Oturum Süreleri: Bir öğrencinin LMS'de oturum açma süresinin, sınıf ortalamasının çok üzerinde olması, sistemde oturumu açık unutma veya teknik bir sorunu işaret edebilir.

Beklenmedik Etkinlik Sıralamaları: Bir öğrencinin ders materyallerini tamamlamadan sınavı tamamlaması gibi mantıksal sıralama hataları, veri toplama sürecindeki sorunlara işaret edebilir.

Sınav Tamamlama Süreleri: Belirli bir sınavı tamamlamak için harcanan sürelerdeki anormallikler.

3.2.4 Normalizasyon ve Özellik Mühendisliği

Normalizasyon, veri setindeki değişkenlerin ölçeklerini uyumlu hale getirmek için kullanılan bir tekniktir. Bu süreç, özellikle farklı birimlere veya ölçeklere sahip özelliklerin karşılaştırılabilirliğini artırmak ve model performansını iyileştirmek için gereklidir (Han, Kamber, & Pei, 2011). Eğitim veri setlerinde normalizasyon genellikle, mesafe tabanlı algoritmaların doğruluğunu artırmak için kullanılır. Örneğin, KNN ve Destek Vektör Makineleri (SVM) gibi algoritmalar, farklı ölçeklerdeki değişkenler nedeniyle performans kaybına uğrayabilir.

Min-Max Scaling, verileri [0,1] aralığına dönüştürmek için sık kullanılan bir yöntemdir. Formülü şu şekildedir:

$$X' = \frac{(X - X_{\min})}{(X_{\max} - X_{\min})}$$

Denklem 3.2- Min-Max Normalizasyon

Burada X orijinal değer, X min minimum değer, X max ise maksimum değerdir. Min-Max Scaling, özellikle eğitimde kullanılan etkinlik süreleri, quiz başarı oranları ve oturum uzunlukları gibi sürekli değişkenler üzerinde etkili bir şekilde uygulanmaktadır (Chandola, Banerjee, & Kumar, 2009).

Eğitimsel veri madenciliği ve eğitimsel süreç madenciliği alanlarında, veri analizinin doğruluğunu ve model performansını artırmak amacıyla belirli veri türlerinin normalizasyonu kritik bir adımdır. Benzer akademik çalışmalar incelendiğinde, aşağıdaki veri türlerinin normalizasyonunun yaygın olduğu gözlemlenmiştir:

- **Sürekli Nicel Veriler:**

Öğrenci Notları ve Sınav Puanları: Farklı sınavların veya değerlendirmelerin farklı maksimum puanları olabileceğinden, bu puanların normalize edilmesi, öğrencilerin performanslarının tutarlı bir şekilde karşılaştırılmasını sağlar. Örneğin, farklı derslerdeki sınav puanlarının $[0,1]$ aralığına ölçeklendirilmesi, performans analizlerinde tutarlılığı artırır.

Etkinlik Süreleri: Öğrencilerin öğrenme yönetim sistemlerinde (LMS) geçirdikleri süreler, oturum uzunlukları gibi süre bazlı verilerin normalize edilmesi, öğrencilerin platform kullanım yoğunluklarının karşılaştırılmasında önemlidir. Bu tür verilerin normalizasyonu, öğrencilerin öğrenme davranışlarının analizi için gereklidir.

- **Kategorik Veriler:**

Demografik Bilgiler: Cinsiyet, yaş grubu, eğitim seviyesi gibi kategorik verilerin sayısal analizlerde kullanılabilmesi için uygun kodlama yöntemleriyle dönüştürülmesi gereklidir. Bu dönüşüm, modellerin bu değişkenleri etkili bir şekilde işlemesine olanak tanır.

Bu araştırmada kategorik verilerin normalize edilmesi için One-hot encoding yöntemi kullanılmıştır. One-hot encoding, kategorik verilerin, makine öğrenmesi modelleri tarafından işlenebilmesi için sayısal bir forma dönüştürülmesini sağlayan bir yöntemdir. Bu yöntemde, her bir kategorik değer, ilgili kategoriye temsil eden ikili bir vektörle ifade edilir (Bishop, 2006).

- **Etkinlik Frekansları:**

Sistem Giriş-Çıkış Sayıları: Öğrencilerin LMS'ye giriş yapma sıklıkları gibi frekans verilerinin normalize edilmesi, öğrencilerin platform etkileşimlerinin analizinde tutarlılığı sağlar. Bu tür verilerin normalizasyonu, öğrencilerin öğrenme süreçlerindeki etkileşim düzeylerinin karşılaştırılmasında kullanılır.

Özellik mühendisliği, ham verilerden anlamlı ve modelin öğrenme kapasitesini artıracak yeni özelliklerin türetilmesi sürecidir. Eğitim veri setlerinde özellik mühendisliği, öğrencilerin öğrenme davranışlarının daha iyi anlaşılmasını sağlar

(Chatti, Dyckhoff, Schroeder, & Thüs, 2012). Bu çalışmada aşağıdaki özellikler türetilmiştir:

- **Etkinlik Sıklıkları:** Bir öğrencinin belirli bir etkinlikte harcadığı toplam süre.
- **Geçiş Süreleri:** Bir etkinlikten diğerine geçiş için harcanan ortalama süre.
- **Davranış Modelleri:** LMS loglarından çıkarılan süreç modelleri ile öğrencilerin öğrenme desenlerinin belirlenmesi.

Özellik Mühendisliğinde Kullanılan Teknikler

1. Davranış Temelli Özellikler

Davranış temelli özellikler, öğrenme yönetim sistemlerinden (LMS) elde edilen log verilerinden türetilmektedir. Bu teknikler, öğrencilerin öğrenme süreçlerindeki davranışlarını anlamak için kullanılır:

- **Etkinlik Sıklıkları:** Bir öğrencinin belirli bir etkinlikte (ör. sınav tamamlama, kaynak okuma, duyuru okuma vb.) harcadığı toplam süre.
- **Geçiş Süreleri:** Bir etkinlikten diğerine geçiş için geçen süre.
- **Oturum Desenleri:** LMS'ye giriş-çıkış oturumlarının sıklığı ve düzenliliği.

Süreç madenciliği algoritmaları kullanılarak aşağıdaki davranış temelli özellikler çıkarılmıştır:

2. İstatistiksel Özellikler

İstatistiksel yöntemler, öğrencilerin LMS'deki etkileşimlerine ilişkin temel özet bilgilerin çıkarılmasında kullanılmıştır. Bu teknikler, veri setindeki değişkenlerin daha iyi analiz edilmesini sağlar (Han, Kamber, & Pei, 2011). Bu çalışmada kullanılan istatistiksel özellikler şunlardır:

- **Oturum Süresi Ortalamaları ve Standart Sapmaları:** Öğrencilerin oturum süresi verilerinin merkezi eğilim ve dağılım ölçütleri.

3. Kategorik Özelliklerin Kodlanması

Kategorik değişkenlerin (ör. cinsiyet, eğitim seviyesi) sayısal formata dönüştürülmesi, modellerin bu değişkenlerle çalışmasını mümkün kılar. Bu çalışmada **one-hot**

encoding yöntemi kullanılmıştır. One-hot encoding, her bir kategoriye ikili bir vektörle ifade ederek sıralama veya büyüklük ilişkisi olmayan değişkenlerde anlamlı bir dönüşüm sağlar (Bishop, 2006).

4. Zaman Tabanlı Özellikler

Zaman damgalarına dayalı olarak, öğrencilerin öğrenme davranışlarını analiz etmek amacıyla zaman tabanlı özellikler türetilmiştir. Bu özellikler, öğrencilerin öğrenme desenlerini daha ayrıntılı bir şekilde açıklamayı mümkün kılar:

- **Aktivite Zaman Çizelgeleri:** Belirli zaman aralıklarında öğrencilerin LMS'deki etkinliklerinin yoğunluğu.
- **Geçiş Süreleri:** Bir etkinlikten diğerine geçiş için geçen sürelerin analizi.

5. Süreç Modelleme Özellikleri

Süreç madenciliği teknikleri kullanılarak çıkarılan özellikler, öğrenme süreçlerinin modellenmesinde kritik bir rol oynamaktadır (van der Aalst, 2016). Bu özellikler, süreç modelleri üzerinden elde edilen istatistikler ve uyum skorlarını içerir:

- **Süreç Uygunluk Skorları:** Öğrencilerin davranışlarının süreç modeline uyum derecesi.
- **Sapma Analizleri:** Beklenen süreç ile gerçekleşen süreç arasındaki farkların analizi.

Bu çalışmada kullanılan özellik mühendisliği teknikleri, LMS verilerinin anlamlı bir şekilde yapılandırılmasını ve öğrencilerin öğrenme davranışlarının daha iyi anlaşılmasını sağlamaktadır. Literatürde de benzer şekilde özellik mühendisliği tekniklerinin, makine öğrenmesi modellerinin performansını artırmada önemli bir etkisi olduğu belirtilmektedir (Umer, Susnjak, Mathrani, & Suriadi, 2017) (Chatti, Dyckhoff, Schroeder, & Thüs, 2012).

3.3 Süreç Madenciliği Teknikleri

Bu çalışmada, öğrenci davranışlarını modellemek ve analiz etmek amacıyla süreç madenciliği teknikleri kullanılacaktır. Süreç madenciliği, öğrenme yönetim sistemi (LMS) verilerinden elde edilen olay günlükleri kullanarak öğrenme süreçlerinin

keşfedilmesi, değerlendirilmesi ve analiz edilmesi için etkili bir araç sunmaktadır. Çalışma kapsamında uygulanacak yöntemler aşağıda açıklanmıştır:

Keşif Madenciliği; Öğrenci davranışlarının modellenmesi amacıyla keşif madenciliği (process discovery) yöntemi kullanılacaktır. Bu kapsamda, ProM yazılımı ve Inductive Miner algoritması tercih edilmiştir. Inductive Miner algoritması, olay loglarından süreç modelleri oluşturmak için yapılandırılmış bir yaklaşımdır ve öğrencilerin öğrenme aktiviteleri arasındaki geçişleri ve sıralamaları analiz etmek için kullanılacaktır (van der Aalst, 2016). Bu yöntem, LMS verilerinden elde edilen olay loglarının süreç modellerine dönüştürülmesini sağlayarak öğrenme süreçlerinin görselleştirilmesine olanak tanıyacaktır.

Uygunluk Kontrolü (Conformance Checking) : Süreç modellerinin gerçek öğrenci davranışlarına uyumunu değerlendirmek için **uygunluk kontrolü** teknikleri uygulanacaktır. Bu çalışmada, **Petri Net tabanlı süreç modelleri** kullanılarak, ideal süreç modeli ile gerçek davranışlar arasındaki uyum analiz edilecektir. Uygunluk kontrolü, aşağıdaki metrikler kullanılarak gerçekleştirilecektir:

- **Süreç Uygunluk Skoru:** Gerçek davranışların modelle ne ölçüde uyumlu olduğunu belirlemek için kullanılacaktır. Süreç modeline olan davranışsal uyumu sayısal olarak ifade eden metrik, literatürde ‘fitness score’ olarak anılmakta olup, bu çalışmada ‘süreç uygunluk skoru’ olarak ifade edilmiştir.
- **Sapma Analizleri:** Model ile olay logları arasındaki uyumsuzlukları belirlemek amacıyla uygulanacaktır.

Bu teknikler, öğrencilerin öğrenme süreçlerinin ideal modele ne kadar uyumlu olduğunu belirlemek ve sapmaları tespit etmek için kullanılacaktır.

Analiz ve Değerlendirme: Süreç modellerinden elde edilen sonuçlar, öğrencilerin öğrenme davranışlarının analizi ve değerlendirilmesi amacıyla kullanılacaktır. Bu kapsamda aşağıdaki adımlar uygulanacaktır:

- **Süreç Uygunluk Skorlarının Değerlendirilmesi:** Her bir öğrencinin süreç modeline uyum düzeyi incelenecektir.

- **Davranış Şablonlarının Çıkarılması:** Süreç modellerinden elde edilen öğrenme desenleri analiz edilerek, bu desenlerin öğrencilerin akademik performansına etkisi değerlendirilecektir.

Bu yöntemler, öğrencilerin bireysel öğrenme süreçlerinin daha iyi anlaşılmasını ve öğrenme modellerinin akademik performans ile ilişkili hale getirilmesini hedeflemektedir.

3.4 Makine Öğrenmesi Modelleri

Bu çalışmada, LMS verilerinden türetilen özelliklerin öğrencilerin akademik performanslarını tahmin etmek amacıyla makine öğrenmesi modelleri üzerinde değerlendirilmesi hedeflenmektedir. Bu kapsamda uygulanacak yöntemler ve değerlendirme ölçütleri aşağıda detaylandırılmıştır:

3.4.1 Kullanılan Algoritmalar

Makine öğrenmesi modelleri, öğrencilerin öğrenme davranışlarını ve akademik başarılarını tahmin etmek için yaygın olarak kullanılan güçlü araçlardır. Bu çalışmada, farklı veri özellikleri ve desenlerini yakalamak amacıyla üç temel algoritma seçilmiştir:

- **Rastgele Orman (RF):** Karar ağacı tabanlı bir yöntem olup, birden fazla ağacın bir araya getirilmesiyle tahmin doğruluğunu artırır. Özellikle büyük veri setlerinde ve karmaşık ilişkilere sahip özelliklerde yüksek performans göstermektedir (**Breiman, 2001**). Öğrenci davranışlarının ve akademik performansının sınıflandırılmasında, Rastgele Orman, karmaşık veri ilişkilerini anlamada ve özellik önemini belirlemede güçlü bir yöntemdir. Büyük veri setlerinde hem sınıflandırma hem de regresyon görevlerinde sıklıkla tercih edilir.

- **K-En Yakın Komşu(KNN):** Sınıflandırma ve regresyon görevlerinde kullanılan, veri noktalarının en yakın komşularına dayalı bir yöntemdir. Bu algoritma, öğrencilerin öğrenme davranışlarına dayalı benzerliklerin analizinde etkili bir şekilde kullanılacaktır (Altman, 1992). KNN, öğrencilerin davranış desenlerini analiz etmek ve benzer öğrenciler arasındaki ilişkileri ortaya çıkarmak için idealdir. Özellikle

sınıflandırma görevlerinde, davranış temelli gruplandırma ve kişiselleştirilmiş öneri sistemlerinde etkili bir şekilde kullanılabilir.

- **Karar Ağaçları (Decision Tree):** Karar ağaçları, veriyi dallara ayırarak sınıflandırma veya regresyon işlemleri gerçekleştiren gözetimli öğrenme algoritmalarıdır. Her bir dal, veri kümesinin bir özelliğine göre bölünmesini temsil ederken, yaprak düğümler nihai sınıf etiketlerini içerir. Bu yöntem, yorumlanabilirliği yüksek olması nedeniyle eğitimsel veri madenciliği çalışmalarında sıklıkla kullanılmaktadır. Öğrenci davranışlarının açıklanması ve akademik başarı tahminlerinin görselleştirilebilir bir yapıda sunulması açısından avantaj sağlar (Quinlan, 1986). LMS verileri üzerinde karar ağacı kullanımı, öğrenme yollarının izlenmesi ve önemli karar noktalarının belirlenmesi açısından etkili bir yöntemdir.

- **Destek Vektör Makineleri (SVM):** Sınıflandırma problemleri için kullanılan, yüksek boyutlu veri setlerinde doğrusal olmayan desenleri yakalamada etkili bir yöntemdir. SVM, öğrencilerin davranışlarının farklı özelliklerine dayalı sınıflandırma için tercih edilmiştir (Cortes & Vapnik, 1995). SVM, öğrenci davranışlarındaki doğrusal olmayan desenleri ve sınıfları tespit etmek için tercih edilen bir algoritmadır. Yüksek boyutlu veri setlerinde sınıflandırma görevlerinde, davranış analizi ve başarı tahmini için etkili bir yöntemdir.

- **XGboost:** XGBoost, karar ağaçları tabanlı bir gradient boosting algoritmasıdır. Performansı artırmak ve model doğruluğunu optimize etmek için paralel işlemeye, düzenleme tekniklerine (regularization) ve kayıp fonksiyonlarının özelleştirilmesine olanak tanır. Özellikle eksik veri ve büyük ölçekli veri setleri ile çalışırken yüksek verimlilik sağlar. Öğrencilerin öğrenme davranışlarını modellemede ve akademik performanslarını tahmin etmede, XGBoost sıkça tercih edilen güçlü bir makine öğrenmesi yöntemidir (Chen & Guestrin, 2016). Öğrenci davranışlarının ve akademik performansının sınıflandırılmasında XGBoost, büyük LMS veri setlerinde eksik verilerin işlenmesi ve yüksek tahmin doğruluğu sağlamak için sıklıkla kullanılır.

- **LightGBM:** LightGBM, karar ağaçları tabanlı bir gradient boosting algoritması olup, özellikle büyük veri setleri ve yüksek boyutlu özellikler üzerinde hızlı ve verimli bir şekilde çalışmasıyla öne çıkar. Bu algoritma, histogram tabanlı bir yaklaşım

kullanarak eğitim süresini kısaltır ve bellek kullanımını optimize eder. Eğitimsel veri madenciliğinde, LightGBM düşük gecikmeli sınıflandırma ve regresyon görevleri için idealdir (Ke, ve diğerleri, 2017). Öğrenci davranışlarının ve akademik performansının sınıflandırılmasında LightGBM, büyük LMS veri setlerinde ölçeklenebilirlik ve doğruluk sağlamak için tercih edilebilir.

3.4.2 Performans Ölçütleri

Her bir algoritma, LMS verilerinden türetilen özellikler üzerinde eğilecek ve test edilecektir. Makine öğrenmesi modellerinin değerlendirilmesi için çeşitli performans ölçütleri kullanılacaktır. Bu ölçütler, modellerin sınıflandırma doğruluğunu ve genellenebilirliğini ölçmek amacıyla tercih edilmiştir:

F1-Skoru: Modelin doğruluğunu ve hatalarını dengelemek için kullanılır. F1-Skoru, Precision (kesinlik) ve Recall (duyarlılık) ölçütlerinin harmonik ortalamasını alır.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Denklem 3.3- F1 Skor

ROC Eğrisi ve AUC (Area Under Curve): Modelin farklı eşik değerlerinde sınıflandırma performansını görselleştirmek ve değerlendirmek için kullanılır. ROC eğrisi, modelin farklı eşik değerlerinde doğru pozitif oranı (TPR) ve yanlış pozitif oranı (FPR) değerlerini görselleştirir.

True Positive Rate (TPR) veya Duyarlılık (Recall):

$$TPR = \frac{TP}{TP + FN}$$

Denklem 3.4- True Positive Rate

Gerçek pozitiflerin ne kadarının doğru tahmin edildiğini ifade eder.

False Positive Rate (FPR):

$$FPR = \frac{FP}{FP + TN}$$

Denklem 3.5- False Positive Rate

Negatif sınıfın yanlış pozitif olarak tahmin edilme oranını ifade eder.

ROC Eğrisinin Hesaplanması:

2. Farklı eşik değerleri belirlenir.
3. Her eşik değeri için TPR ve FPR hesaplanır.
4. TPR ve FPR değerleri, ROC eğrisi üzerinde noktalar olarak işaretlenir.

AUC (Area Under Curve): AUC, ROC eğrisi altındaki alanı ifade eder ve modelin genel performansını ölçer:

- AUC = 1.0 → Mükemmel model.
- AUC = 0.5 → Rastgele tahmin.

AUC değeri ne kadar büyükse model o kadar iyi performans gösterir, çünkü bu durumda model daha fazla pozitif sınıfı doğru bir şekilde tahmin eder.

Kesinlik (Precision) ve Duyarlılık (Recall): Doğru tahmin edilen pozitif sınıf oranını ve gerçek pozitiflerin model tarafından doğru tahmin edilme oranını ölçer.

Kesinlik, pozitif tahminlerin ne kadarının doğru olduğunu ifade eder.

$$Precision = \frac{TP}{TP + FP}$$

Denklem 3.6- Kesinlik (Precision)

- **TP (True Positive):** Doğru şekilde pozitif olarak tahmin edilen örnekler.
- **FP (False Positive):** Yanlış şekilde pozitif olarak tahmin edilen örnekler.

Duyarlılık (Recall), gerçek pozitiflerin ne kadarının doğru tahmin edildiğini ifade eder.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Denklem 3.7- Duyarlılık (Recall)

- **TP (True Positive):** Doğru şekilde pozitif olarak tahmin edilen örnekler.
- **FN (False Negative):** Yanlış şekilde negatif olarak tahmin edilen örnekler.

3.4.3 Doğrulama Yöntemi

Çalışmada, modellerin genelleme performansını değerlendirmek için K-Fold Cross Validation yöntemi uygulanacaktır. Bu yöntem, veri setini K eşit parçaya ayırarak her bir parçayı test veri seti olarak kullanırken kalan K-1 parçayı eğitim veri seti olarak kullanır. Bu süreç K kez tekrarlanarak her bir veri noktası bir kez test veri setinde yer alır. Bu yöntemin seçilme nedeni, veri setinin küçük olduğu durumlarda model performansının daha güvenilir bir şekilde ölçülmesini sağlamasıdır (Kohavi, 1995). Yukarıda belirtilen yöntemlerle, LMS verilerinden elde edilen öğrenme davranışlarının akademik performans üzerindeki etkisi değerlendirilecek ve tahmin modellerinin doğruluğu çeşitli metriklerle karşılaştırılacaktır.

3.4.4 Tutarlılık Değerlendirme Kriterleri

Makine öğrenmesi modellerinin başarımı yalnızca doğruluk (accuracy) ya da F1 skoru gibi anlık performans ölçütleriyle sınırlı olarak değerlendirilmemelidir (Provost & Fawcett, 2013). Eğitimseti veri madenciliği uygulamalarında özellikle üretim ortamına aktarılabilir karar destek sistemlerinin geliştirilmesi hedeflendiğinde, modelin farklı doğrulama stratejilerine karşı gösterdiği istikrar ve genellenebilirlik düzeyi de dikkate alınmalıdır (Baker & Inventado, 2014). Literatürde, modelin sabit veri bölmeleri ile çapraz doğrulama gibi yöntemlere karşı duyarlılığını analiz etmenin, modelin güvenilirliğini test etmek açısından kritik bir adım olduğu belirtilmektedir (Roberts, et al., 2017). Bu çalışma kapsamında, kullanılan sınıflandırma modelleri %80 eğitim – %20 test bölmesi (80-20 Split) ve 10 katlı çapraz doğrulama (K-Fold CV) yöntemleri altında test edilmiş; elde edilen F1 skoru ve ROC AUC değerleri doğrultusunda beş

kategoride tutarlılık sınıflandırması yapılmıştır. Bu sınıflandırma yaklaşımı, benzer biçimde farklı doğrulama yöntemlerinin model performansı üzerindeki etkisini ortaya koyan güncel araştırmalarla da uyumludur.

Tablo 3.5 Tutarlılık Kategorileri

Kategori	Formül
Yüksek ve Tutarlı	$F1 (CV) > 0.90, \Delta F1 < 0.02, AUC > 0.90$
Split Yüksek, CV Düşük	80-20'de yüksek F1, fakat $\Delta F1 > 0.08$
Dalgalı	$\Delta F1 \geq 0.05$ – modele bağlı dalgalanma
Düşük ve Dalgalı	$F1 (CV) < 0.80, AUC < 0.90$
Orta Düzeyde Tutarlı	$\Delta F1 < 0.05, AUC$ genelde > 0.90

F1 Skoru (K-Fold CV) > 0.90 bu eşik, modelin yüksek sınıflandırma başarımı sunduğunu gösteren kritik bir sınırdır. Özellikle eğitim verilerinde sıklıkla rastlanan sınıf dengesizliklerinde, F1 skoru, doğruluk oranına kıyasla daha güvenilir bir metrik olarak kabul edilir (Tharwat, 2020).

F1 > 0.90 olan modellerin, pozitif sınıfı hatasız tespit etme oranlarının oldukça yüksek olduğunu ve bu modellerin güvenli bir şekilde üretime aktarılabilirliğini ifade etmektedir (Saito & Rehmsmeier, 2015).

Literatürde bir çok araştırma F1 > 0.90 olan modelleri, “yüksek güven düzeyine sahip sınıflandırıcılar” olarak tanımlamaktadır (Provost & Fawcett, 2013) (Tharwat, 2020) (Encord, 2023).

80-20 ile K-Fold CV, F1 skoru farkının 0.02'den küçük olması, modelin doğrulama türüne karşı duyarlılığının düşük olduğunu, yani istikrarlı ve kararlı çalıştığını gösterir.

- Çapraz doğrulama ile sabit veri bölmesi arasında performans farkının düşük olması durumunda modelin daha yüksek genellenebilirliğe sahip olduğunu vurgulamıştır (Roberts, et al., 2017).

- Diğer bir çalışmada ise $\Delta F1$ farklarının %2'nin altında tutulmasının, özellikle karar destek sistemlerinde karar güvenilirliğini artırdığını belirtmektedir. Bu nedenle $\Delta F1 < 0.02$ sınırı, yüksek tutarlılığın göstergesi olarak değerlendirilmiştir (Hand, 2012).

ROC AUC metriği, modelin pozitif ve negatif sınıfları farklı eşik değerleri altında ne derece iyi ayırt ettiğini gösteren ölçüttür.

- Literatürde, AUC değerinin 0.90'ın üzerinde olması 'üstün ayırım gücü' (superior discriminative power) olarak kabul edilmekte ve modelin sınıflandırma performansının mükemmele yakın olduğunu göstermektedir (Bradley, 1997) (Fawcett, 2006).
- $AUC > 0.9$ değeri, yalnızca genel doğruluk değil, aynı zamanda eşik duyarlılığı (threshold sensitivity) açısından da üstün performansı işaret eder. Bu tür modeller, sınıf ayırım gücü yüksek ve üretim ortamında güvenle kullanılabilecek yapıdadır (Fawcett, 2006). Özellikle klinik ve finansal karar sistemlerinde, yüksek AUC değerleri, modelin farklı eşik seviyelerinde tutarlılığını kanıtlar (Provost & Fawcett, 2013).

Bu yapı, klasik “en yüksek doğruluk değerine sahip model seçimi” yaklaşımından farklı olarak, model kararlılığı, genellenebilirlik düzeyi ve ayırım gücü gibi faktörleri birlikte değerlendirerek daha güvenilir model seçimi yapılmasına olanak sağlamaktadır.

Yüksek ve Tutarlı: Bu kategori, genellenebilirliği yüksek ve üretim ortamlarında güvenle uygulanabilecek modelleri tanımlamaktadır. Aşağıdaki koşulları sağlayan modeller bu gruba dâhil edilmiştir:

- K-Fold CV altında **F1 Skoru > 0.90**,
- 80-20 ile K-Fold arasındaki fark **$\Delta F1 < 0.02$** ,
- ROC AUC değeri **> 0.90**.

Bu kriterler, modelin hem sınıflandırma başarısını hem de doğrulama yönteminden bağımsız olarak tutarlılığını gösterir.

Split Yüksek, CV Düşük: Bu grupta yer alan modeller, 80-20 veri bölmesinde yüksek F1 skoru üretmiş; ancak K-Fold doğrulamada anlamlı bir düşüş sergilemiştir ($\Delta F1 > 0.08$). Bu durum, modelin eğitildiği veri örneklerine fazla uyum sağladığını ve farklı veri alt kümelerinde performansını sürdürmediğini, yani aşırı öğrenme (overfitting) eğiliminde olduğunu göstermektedir (Kohavi, 1995).

Dalgali: Modellerin 80-20 ve K-Fold CV doğrulama sonuçları arasında orta düzeyde sapma ($\Delta F1 \geq 0.05$) göstermesi durumunda bu kategori kullanılmaktadır. Bu gruba giren modellerin performansı sabit olmayıp, veri bölmesinden etkilenmektedir.

Düşük ve Dalgali: Bu kategori, hem F1 skoru bakımından düşük, hem de AUC değeri bakımından zayıf modelleri kapsamaktadır. Kriterler şunlardır:

- K-Fold CV altında **F1 < 0.80**
- ROC AUC < 0.90

Bu modellerin sınıflar arası ayırım gücü düşüktür ve genellikle hem küçük hem de dengesiz veri setlerinde başarısız sonuçlar üretmektedir (Saito & Rehmsmeier, 2015).

Orta Düzeyde Tutarlı: Yukarıdaki kategorilere dâhil edilemeyen, ancak anlamlı sapmalar da göstermeyen modeller bu grupta değerlendirilmiştir. Bu modellerin:

- $\Delta F1 < 0.05$ aralığında kalması,
- AUC değerinin çoğunlukla > 0.90

olması, performanslarının kabul edilebilir seviyede ve orta düzeyde istikrarlı olduğunu göstermektedir. Bu tür modeller, özellikle zaman ve kaynak kısıtlı projelerde tercih edilebilir.

Bu sınıflandırma, sadece doğruluk oranları üzerinden değil; modelin genellenebilirliği, kararlılığı ve eşik duyarlılığı açısından kapsamlı bir değerlendirme imkânı sunar. Literatürde yaygın olan klasik başarı kıyaslamalarının ötesine geçerek, bu yapılandırılmış kriterler sayesinde modellerin uzun vadeli güvenilirliği daha somut biçimde analiz edilmiştir. Bu yaklaşım, özellikle eğitimsel veri madenciliği ve makine öğrenmesi temelli öğrenci davranış tahmini gibi uygulama alanlarında daha gerçekçi model seçimini mümkün kılar.

3.4.5 Hiperparametre Ayarlama Süreci

Bu çalışmada kullanılan tüm makine öğrenmesi algoritmaları için, modellerin genellenebilirliğini artırmak ve aşırı öğrenme riskini azaltmak amacıyla hiperparametre optimizasyonu gerçekleştirilmiştir. Bu işlem, GridSearchCV (Pedregosa, et al., 2012) yöntemi kullanılarak, 5 katlı çapraz doğrulama (5-Fold CV) ile uygulanmış ve değerlendirme metrikleri olarak öncelikli olarak F1 skoru temel alınmıştır. Her algoritma için optimize edilen hiperparametreler ve arama alanları aşağıda detaylandırılmıştır.

Tablo 3.6 Hiperparametre Arama Alanları

Model	Parametre	Arama Alanı	Seçilen Değer
RF	n_estimators	[80, 100, 200, 300]	80
	max_depth	[None, 10, 20]	9
	min_samples_leaf	[1, 2, 4]	1
	min_samples_split	[2, 5, 10]	2
	max_features	max_features	sqrt
	bootstrap	[True, False]	True
SVM	C	[0.1, 1, 10, 100]	100
	gamma	['scale', 'auto', 1]	1
	kernel	['linear', 'rbf']	rbf
DT	max_depth	[None, 7, 10, 20]	7
	min_samples_split	[2, 4, 10]	4
	min_samples_leaf	[1, 2, 4]	1
KNN	n_neighbors	[3, 5, 7, 9]	5
	weights	['uniform', 'distance']	uniform

	metric	['euclidean', 'manhattan']	'euclidean'
LightGBM	num_leaves	[31, 50, 70]	50
	min_data_in_leaf	[1, 2, 5]	2
	learning_rate	[0.01, 0.05, 0.1]	0.1

3.5 Süreç Madenciliği ve Makine Öğrenmesi Tekniklerinin Seçimi

Bu çalışmada kullanılan süreç madenciliği ve makine öğrenmesi teknikleri, literatürde bu tür problemlerde etkili oldukları kanıtlanmış yöntemlerden seçilmiştir. Süreç madenciliği için Inductive Miner algoritması tercih edilmiştir çünkü eğitim süreçlerindeki karmaşık ve tekrarlayan davranışları yüksek doğrulukla modelleme kapasitesine sahiptir (van der Aalst, 2016). Makine öğrenmesi tarafında ise, Random Forest ve XGBoost algoritmaları, LMS verilerinin yüksek boyutluluğu ve eksik veri gibi zorluklarla başa çıkma yetenekleri nedeniyle seçilmiştir (Breiman, 2001) (Chen & Guestrin, 2016). Ayrıca, LightGBM algoritması, büyük LMS veri setlerini hızlı ve hafif bir şekilde analiz etme kabiliyeti ile öne çıkmaktadır (Ke, ve diğerleri, 2017). Rastgele Orman, yüksek boyutlu verilerde özellikler arasındaki karmaşık ilişkileri anlamada etkili olduğu ve aynı zamanda fazla öğrenmeyi (overfitting) minimize ettiği için seçilmiştir (Breiman, 2001). KNN, öğrenci davranışları arasındaki benzerlikleri kolayca modellemesi ve düşük boyutlu verilerde etkili sonuçlar üretmesi nedeniyle tercih edilmiştir (Altman, 1992). SVM, doğrusal olmayan desenleri yüksek doğrulukla sınıflandırabilmesi ve küçük veri setlerinde güçlü performans sergilemesi nedeniyle seçilmiştir (Cortes & Vapnik, 1995). Bu algoritmaların seçimi hem veri setinin yapısal özelliklerine hem de araştırmanın tahmin doğruluğunu artırma hedefine uygun olarak yapılmıştır.

3.6 Araçlar ve Yazılımlar

Bu çalışmada, süreç madenciliği, veri analizi ve istatistiksel değerlendirme süreçleri için birden fazla yazılım aracı kullanılmıştır. Süreç modelleme ve uygunluk kontrolü analizleri için ProM ve Disco, makine öğrenmesi algoritmalarının uygulanması ve veri analizi ve istatistiksel analizler için Python dili tercih edilmiştir.

3.7 Etik Beyan

Bu tez çalışmasının bazı bölümlerinin yazım ve düzenleme sürecinde, üretici yapay zekâ teknolojilerinden (özellikle OpenAI tarafından geliştirilen ChatGPT-4o) sınırlı ölçüde yararlanılmıştır. Bu destek aşağıdaki biçimlerde sağlanmıştır:

- Literatür bölümünde bazı kaynakların özetlenmesi, sadeleştirilmesi veya yeniden yazımı
- Kod bloklarının yapısal kontrolü veya optimize edilmesi
- Tez içeriğindeki bazı akademik cümlelerin dilbilgisel olarak düzeltilmesi, daha açık hale getirilmesi

Bu katkılar, tez yazarının denetiminde ve kontrolünde yürütülmüş; yapay zekâ önerileri eleştirel süzgeçten geçirilerek kullanılmıştır. Üretici yapay zekâ herhangi bir özgün bilimsel sonuç üretmemiş, yalnızca anlatım kolaylığı sağlamıştır. Tezin özgün katkıları ve bilimsel yorumları tamamen yazarına aittir.

DÖRDÜNCÜ BÖLÜM

4. UYGULAMA VE DENEYSEL ÇALIŞMALAR

Bu bölümde, tez kapsamında geliştirilen yöntemlerin doğrulayıcı nitelikteki uygulamaları ve deneysel analizleri sunulmaktadır. Öğrenci davranışlarının öğrenme yönetim sistemi (LMS) verileri üzerinden modellenmesi amacıyla uygulanan makine öğrenmesi algoritmaları, süreç madenciliği teknikleriyle desteklenmiş ve farklı analiz düzeylerinde değerlendirilmiştir. İlk olarak, eğitim verisinin ön işleme süreci gerçekleştirilmiş, ardından çeşitli makine öğrenmesi modelleri (ör. Rastgele Orman, Destek Vektör Makineleri, Karar Ağaçları, K-En Yakın Komşu, LightGBM, XGBoost) eğitilmiş ve çapraz doğrulama ile performans ölçümleri yapılmıştır. Takip eden kısımlarda süreç madenciliği perspektifinden öğrenci etkileşim yolları analiz edilerek davranışsal örüntüler çıkarılmıştır. Son olarak, tüm modellerin çıktıları çeşitli performans ölçütleri kullanılarak karşılaştırmalı biçimde yorumlanmış ve elde edilen sonuçlar literatürle ilişkilendirilmiştir. Bu bütünsel yaklaşım, öğrencilerin akademik başarılarını etkileyen davranışsal faktörlerin daha derinlemesine anlaşılmasına katkı sağlamayı hedeflemektedir.

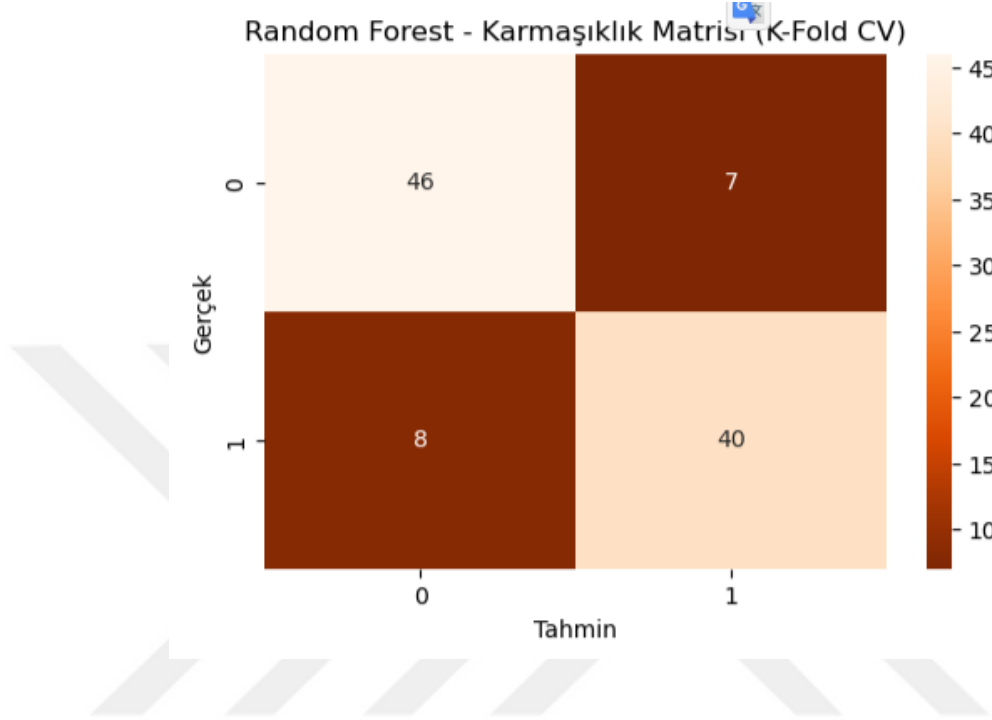
4.1 Makine Öğrenmesi Modellerinin Eğitimi ve Performansı

Bu bölümde, öğrencilerin öğrenme yönetim sistemi (LMS) üzerindeki etkileşim verilerine dayanarak akademik başarı durumlarını tahmin etmek amacıyla uygulanan farklı makine öğrenmesi algoritmalarının performans analizleri sunulmaktadır. Bu çalışmada, Random Forest, K-Nearest Neighbors (KNN), LightGBM, XGBoost, Decision Tree ve Support Vector Machine (SVM) sınıflandırma algoritmaları uygulanmış ve karşılaştırma yapabilmek için 80-20 eğitim test ayrımı ve 10 katlı çapraz doğrulama (Stratified K-Fold CV) kullanılmıştır ve 10 katlı çapraz doğrulama ile değerlendirilmiştir. Değerlendirme metrikleri olarak F1 skoru, ROC AUC ve karışıklık matrisi kullanılmıştır.

4.1.1 Rastgele Orman (RF)

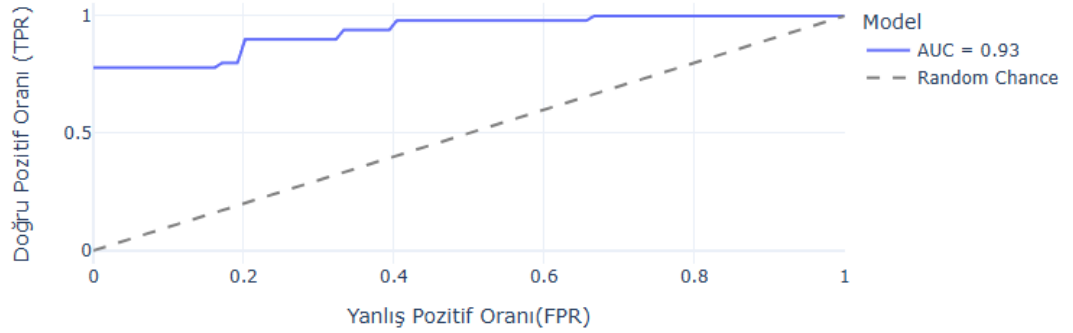
Rastgele Orman modeli, birden fazla karar ağacından oluşan bir topluluk yöntemidir. Özellikle verideki çok boyutluluğa ve karmaşıya karşı dayanıklıdır. Bu çalışmada elde edilen ROC AUC skoru 0.93, Doğruluk (Accuracy) 0.852727, F1 Skoru 0.853851'tir. Karmaşıklık matrisi verileri şu şekildedir: TP=40, TN=46, FP=7, FN=8. Model, hem

doğru pozitif hem de doğru negatif oranlarını yüksek tutmuş ve ezberleme (overfitting) yapmadan sınıflar arası ayırım gücü göstermiştir. Random Forest, ağaç tabanlı topluluk öğrenme yaklaşımıyla çok boyutlu verilerde sınıflar arasını net ayırt etme kapasitesi sayesinde LMS verilerinde son derece başarılı olmuştur (Breiman, 2001).



Şekil 4.1 - Rastgele Orman Karmaşıklık Matrisi

ROC Eğrisi - Random Forest

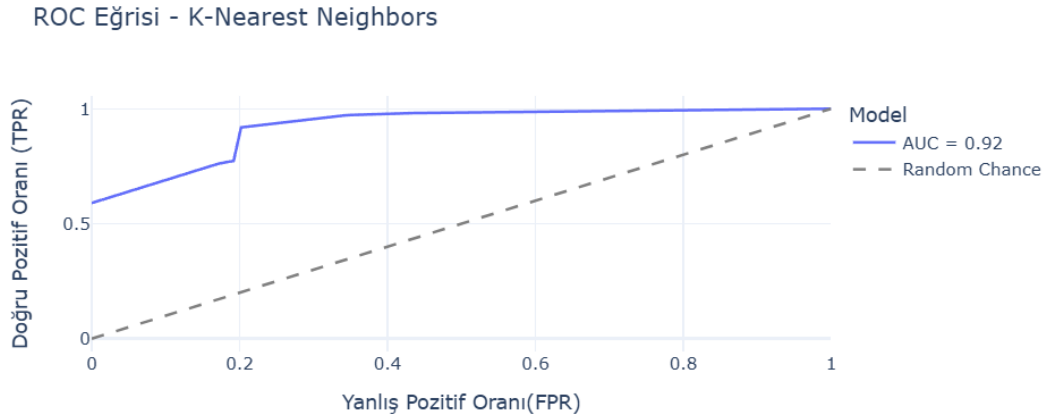
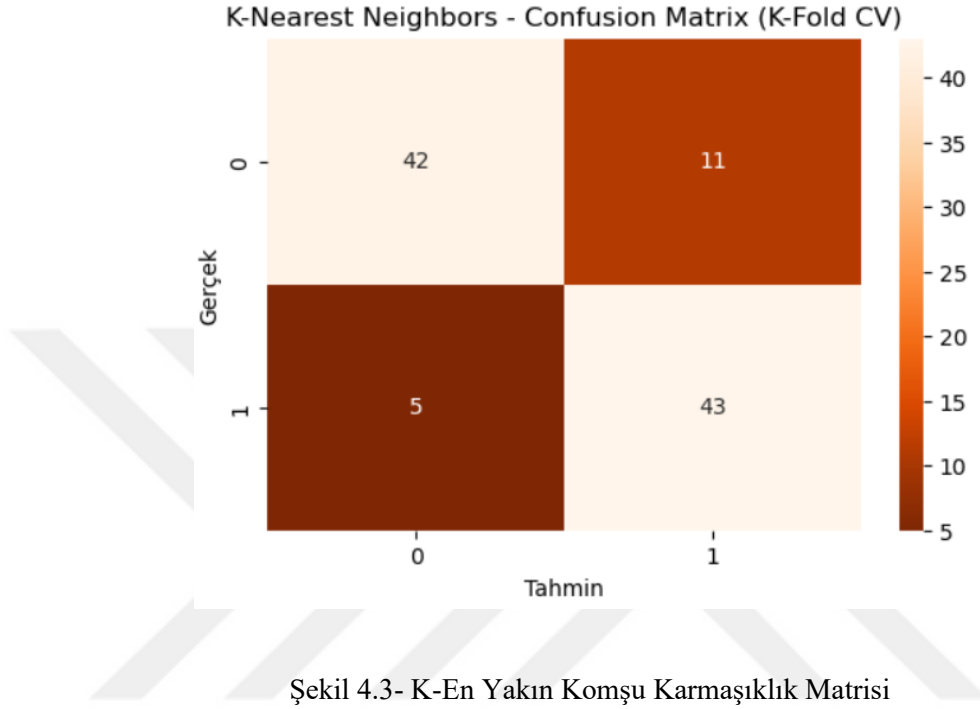


Şekil 4.2 - Rastgele Orman Roc Eğrisi Grafiği

4.1.2 K-En Yakın Komşu (KNN)

KNN, sınıflandırma işlemini veri setindeki komşuluklara dayanarak yapan parametrik olmayan bir algoritmadır. KNN modelinin AUC skoru 0.92, Doğruluk 0.841818, F1 Skoru 0.844179 olarak elde edilmiştir. Karmaşıklık matrisi: TP=43, TN=42, FP=11,

FN=5. Model, başarılı öğrencileri tespit etmede en başarılı algoritmalarından biridir. Verinin dağılımı iyi analiz edildiğinde basit yapısına rağmen güçlü performans gösterir. KNN, özellikle öğrenci davranışlarının birbirine benzediği eğitim ortamlarında etkili bir ayırım sunarak birebir eğitim senaryolarını yansıtmaktadır.

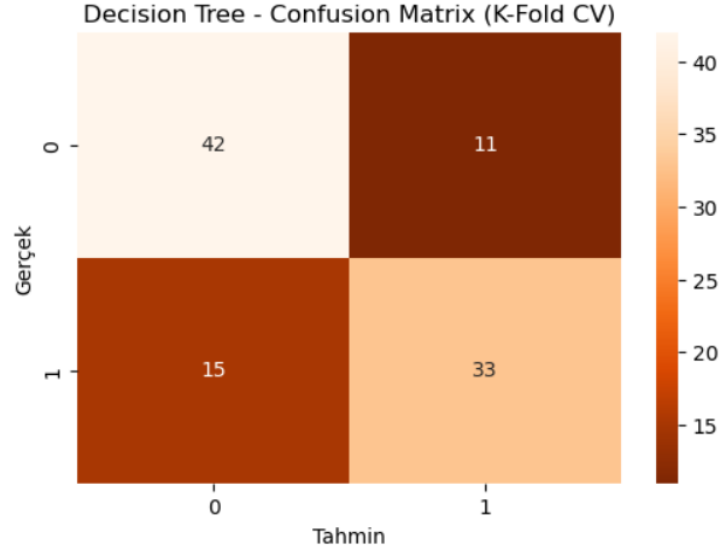


Şekil 4.4 - K-En Yakın Komşu Roc Eğrisi Grafiği

4.1.3 Karar Ağaçları (DT)

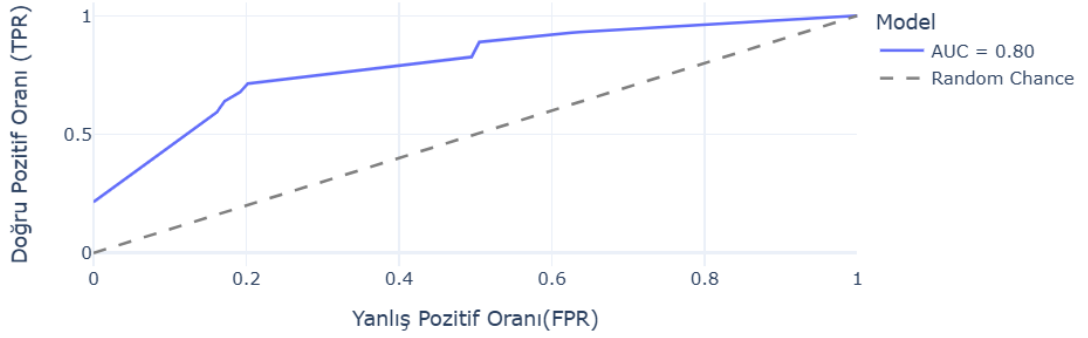
Karar ağaçları modeli, verileri karar kuralları ile dallara ayırarak sınıflandırma yapan sezgisel bir yapıdır. Bu modelin AUC skoru 0.80, Doğruluk 0.743636, F1 Skoru 0.742590 olarak hesaplanmıştır. Karmaşıklık matrisi: TP=33, TN=42, FP=11, FN=15.

Modelin pozitif sınıfa dair tahminlerinde daha fazla hata yaptığı görülmüştür. Yorumlana bilirlği yüksek olmakla birlikte sınır performansa sahiptir. Karar ağacı algoritmalarının yorumlana bilirlği avantaj sağlasa da tekil yapısı LMS gibi çok boyutlu davranış verilerinde yetersiz kalmaktadır.



Şekil 4.5- Karar Ağaçları Karmaşıklık Matrisi

ROC Eğrisi - Decision Tree

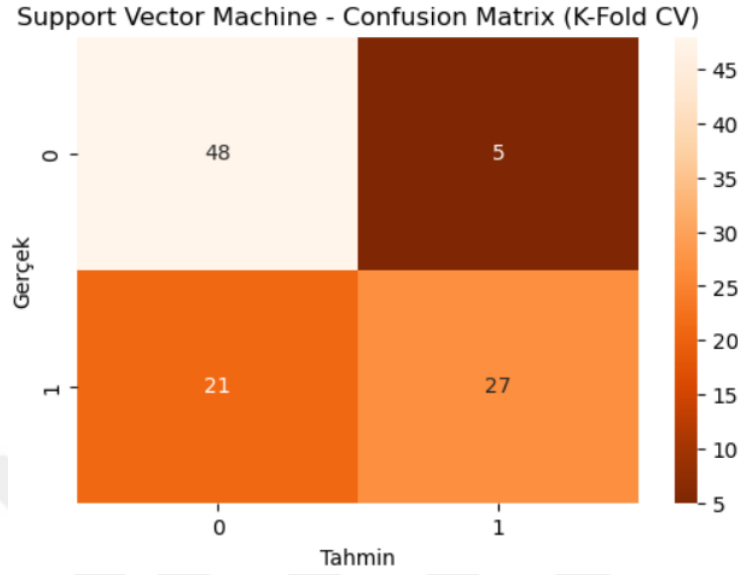


Şekil 4.6 - Karar Ağaçları Roc Eğrisi Grafiği

4.1.4 Destek Vektör Makineleri (SVM)

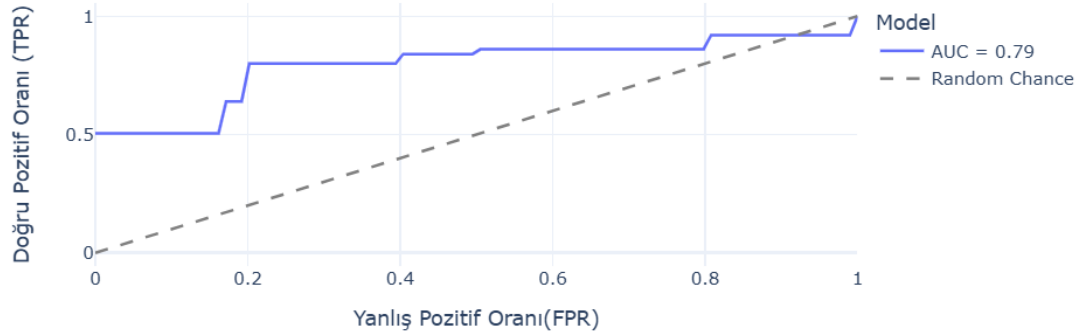
SVM, verileri en iyi ayıracak karar sınırını belirlemeye çalışan bir sınıflandırma yöntemidir. Bu çalışmada AUC skoru 0.79, Doğruluk 0.742727, F1 Skoru 0.731252 olarak elde edilmiştir. Karmaşıklık matrisi: TP=27, TN=48, FP=5, FN=21. Model negatif sınıf (başarısız) tespitinde yüksek başarı göstermiş, ancak pozitif sınıfta (başarılı) hata oranı yüksek olmuştur. SVM, çizgisel olmayan veri dağılımları için

uygun kernel fonksiyonları olmadan kullanıldığında, LMS verilerinde yeterli performans göstermemektedir.



Şekil 4.7 - Destek Vektör Makineleri Karmaşıklık Matrisi

ROC Eğrisi - Support Vector Machine

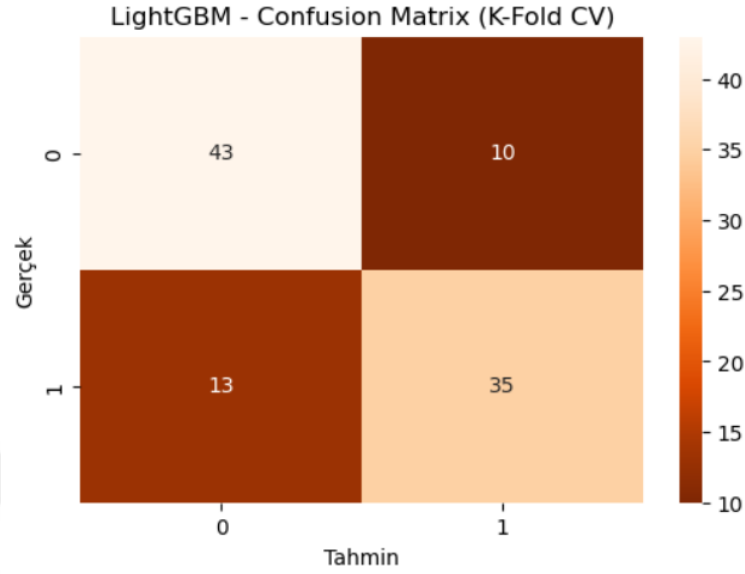


Şekil 4.8 - Destek Vektör Makineleri Roc Eğrisi Grafiği

4.1.5 LightGBM

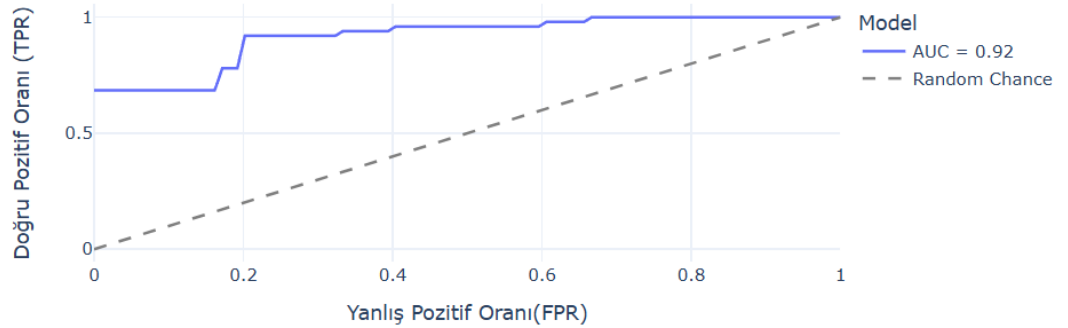
LightGBM, karar ağacı tabanlı gradient boosting yöntemidir. Bu modelin AUC skoru 0.92, Doğruluk 0.773636, F1 Skoru 0.772334'dir. Karmaşıklık matrisi: TP=35, TN=43, FP=10, FN=13. LightGBM modeli, belirli veri seti bölümlerinde “No further splits with positive gain” uyarısı üretmiştir. Bu durum, modele sağlanan değişkenlerin sınıflar arasında ayrım gücünün zayıf olduğu veya varyansın yetersiz olduğu anlamına gelmektedir. Eğitim verisi yeterince bilgi içermediğinde karar ağaçları anlamlı

bölünmeler oluşturamaz, bu da modelin sınıflandırma performansını sınırlar (Ke, et al., 2017).



Şekil 4.9 - Lightgbm Karmaşıklık Matrisi

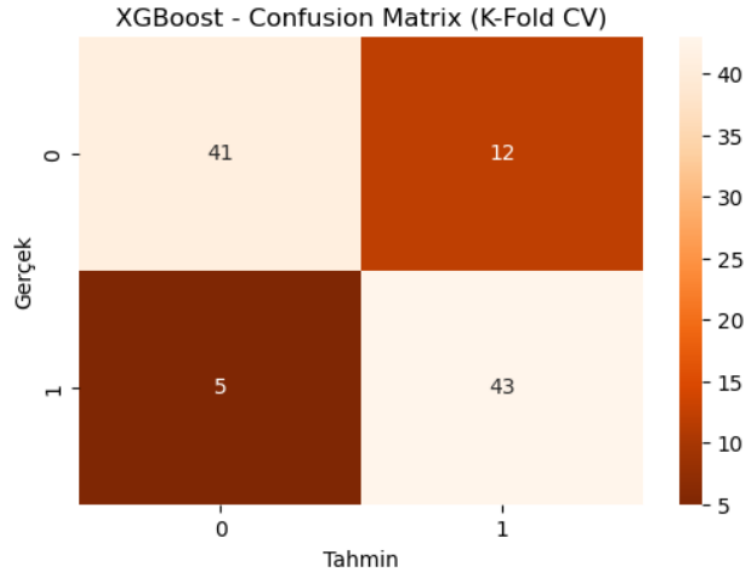
ROC Eğrisi - LightGBM



Şekil 4.10 - Lightgbm Roc Eğrisi Grafiği

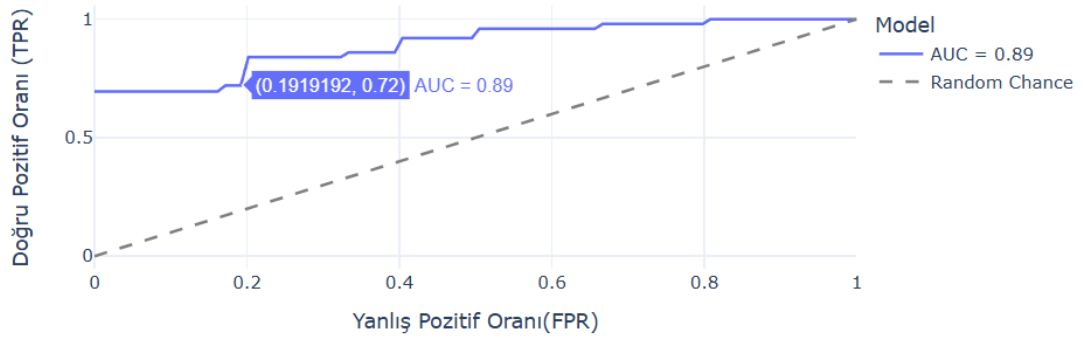
4.1.6 XGBoost

XGBoost, hata oranlarını minimize eden ve karar ağacı topluluklarından oluşan bir boosting yöntemidir. Modelin AUC skoru 0.90, Doğruluk 0.832727, F1 Skoru 0.833938'dir. Confusion matrix: TP=42, TN=41, FP=12, FN=6. Oldukça dengeli bir performans sunmuş, pozitif sınıfta hatayı minimize etmeyi başarmıştır. ROC çizgisi yukarıda konumlanmıştır. Boosting algoritmalarının hatayı ardışık düzenlemelerle azaltma stratejisi, LMS verisindeki karmaşık ilişkileri öğrenme konusunda XGBoost'u etkili kılmıştır.



Şekil 4.11 - XGboost Karmaşıklık Matrisi

ROC Eğrisi - XGBoost



Şekil 4.12 - XGBoost Roc Eğrisi Grafiği

4.1.7 Model Performanslarının Bütünsel Değerlendirmesi

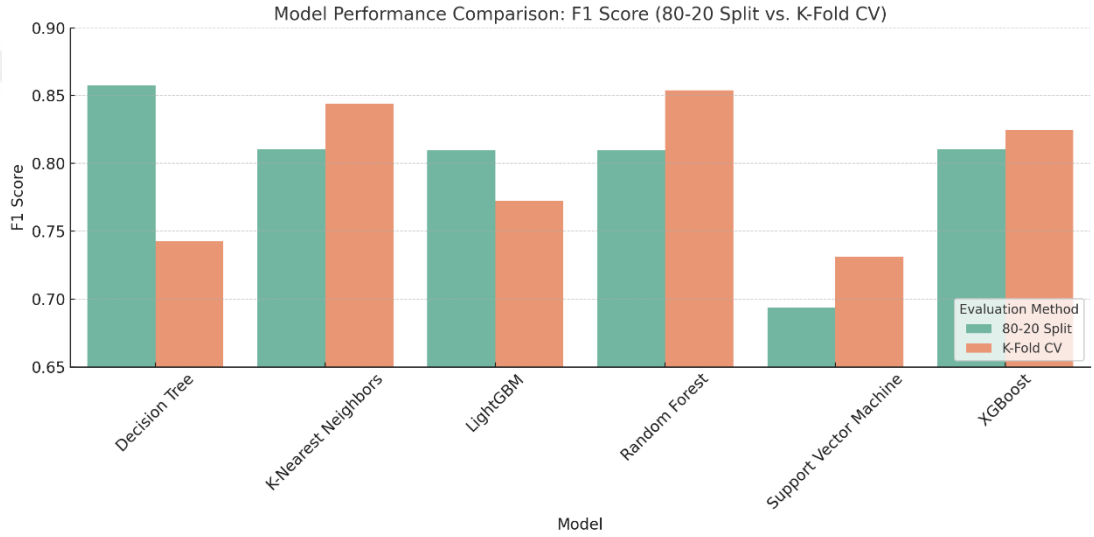
Bu çalışmada, farklı makine öğrenmesi algoritmaları ile oluşturulan sınıflandırma modellerinin performansları ROC AUC, Accuracy ve F1 Score metrikleri temelinde karşılaştırılmıştır. Analizler, hem %80-%20 eğitim/test ayrımı hem de 10 katlı çapraz doğrulama (K-Fold CV) stratejileri ile gerçekleştirilmiştir.

Tablo 4.1 - Model Performans Tablosu (Split & K-Fold)

Model	F1 (80-20)	F1 (K-Fold)	AUC	Tutarlılık
DT	0.857797	0.742590	0.80	Split Yüksek, CV Düşük
KNN	0.810390	0.844179	0.92	Orta Düzeyde Tutarlı

LightGBM	0.904762	0.772334	0.92	Split Yüksek, CV Düşük
RF	0.809524	0.853851	0.93	Orta Düzeyde Tutarlı
SVM	0.693878	0.731252	0.79	Düşük ve Dalgalı
XGBoost	0.810390	0.833938	0.89	Orta Düzeyde Tutarlı

Bu tablodan da görüldüğü üzere, rastgele orman ve K-en yakın komşu modelleri hem yüksek AUC değerleri hem de yüksek F1 skorları ile öne çıkmaktadır. Ayrıca, eğitim/test ayrımı ile çapraz doğrulama sonuçları arasında tutarlılık göstermeleri, bu modellerin genellenebilirliğinin yüksek olduğunu düşündürmektedir.



Şekil 4.13 - F1 skorlarına göre 80-20 Split ve K-Fold CV yöntemlerinin karşılaştırmalı Grafiği

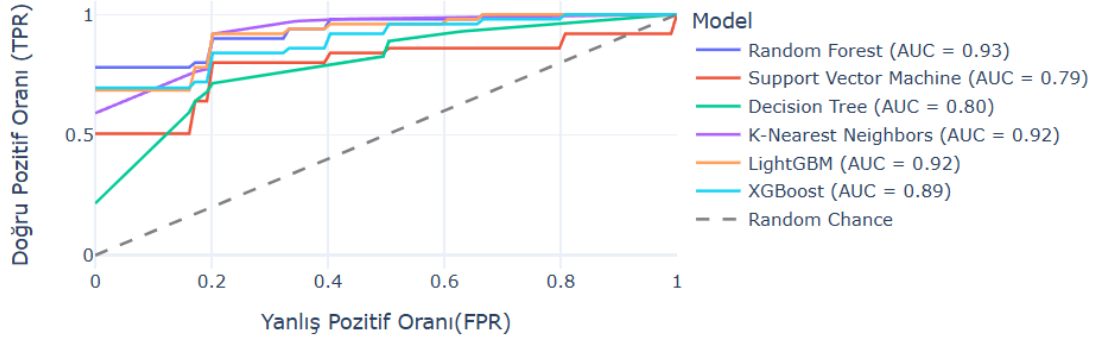
Şekil 4.13’de, uygulanan makine öğrenmesi modellerinin hem 80-20 oranlı sabit eğitim/test bölmesi hem de 10 katlı çapraz doğrulama (K-Fold CV) yöntemleri altında hesaplanan F1 skorları karşılaştırmalı olarak sunulmaktadır. Grafikte dikkat çeken önemli bulgulardan biri, rastgele orman modelinin her iki değerlendirme stratejisinde de yüksek ve tutarlı F1 skoru göstermesidir. Bu durum, modelin aşırı öğrenmeye (overfitting) düşmeden hem eğitim hem de farklı test setlerinde güvenilir sonuçlar üretebildiğini göstermektedir. Benzer şekilde, K-en yakın komşu (KNN) ve XGBoost modelleri de her iki yöntemde yüksek F1 skorları ile öne çıkmakta; bu da onların genellenebilirliğinin ve veri üzerindeki ayırım gücünün yüksek olduğunu göstermektedir.

Diğer taraftan, karar ağacı modeli sabit bölmede (80-20 split) oldukça yüksek bir F1

skoru elde etmesine rağmen, çapraz doğrulamada bu performansı sürdürememiştir. Bu durum, modelin eğitildiği belirli veri bölümlerine aşırı uyum gösterdiğini ve genelleme yetisinin zayıf olduğunu ortaya koymaktadır. Aynı şekilde LightGBM modeli de split testte yüksek bir F1 skoru sunmasına rağmen CV skorlarında belirgin bir düşüş yaşamıştır.

Destek vektör makinesi (SVM) modeli ise her iki değerlendirme yönteminde de görece düşük F1 skorları üretmiştir. Bu durum, modelin doğrusal olmayan veya çok boyutlu veri yapısına yeterince uyum sağlayamadığını ve özellikle pozitif sınıfı (başarılı öğrencileri) doğru sınıflandırmakta zorlandığını göstermektedir.

Karşılaştırmalı ROC Eğrileri - Tüm Modeller (10-Fold CV)



Şekil 4.14 - Karşılaştırmalı Roc Eğrileri Grafiği

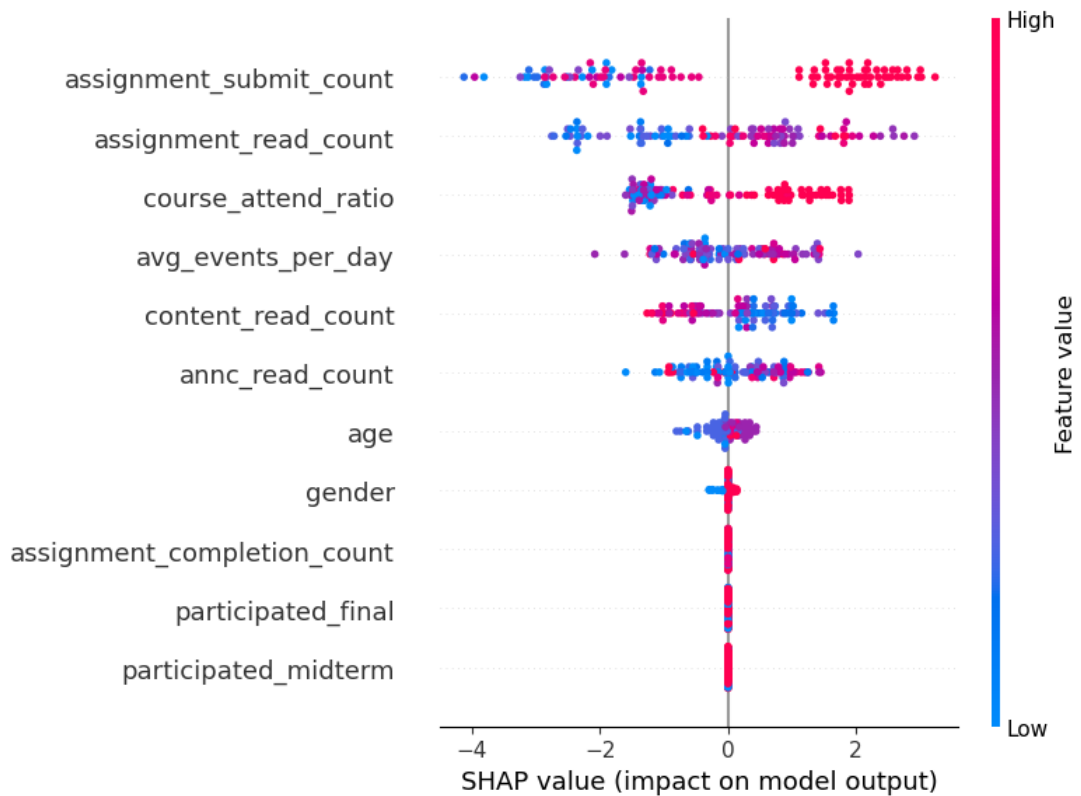
Bu bulgular, öğrenen modelin performansını değerlendirirken yalnızca doğruluk (accuracy) metriğine odaklanmanın yetersiz olduğunu, bunun yerine ROC AUC ve F1 skoru gibi daha duyarlı ve sınıf ayrımına hassas metriklerin kullanılmasının gerekliliğini ortaya koymaktadır. Doğruluk değeri, özellikle sınıf dağılımının dengesiz olduğu veri setlerinde yanıltıcı olabilmekte ve pozitif sınıfa ait örneklerin doğru tahmini konusundaki zayıflıkları gizleyebilmektedir. Bu nedenle, model başarımının daha güvenilir bir şekilde değerlendirilebilmesi için F1 skoru gibi hem duyarlılık (recall) hem de kesinlik (precision) bileşenlerini içeren metriklerin dikkate alınması önem arz etmektedir. Özellikle eğitim bağlamında, başarısızlık riski taşıyan öğrencilerin doğru şekilde tespit edilmesi kritik olduğundan, bu tür hassas metrikler daha açıklayıcı ve işlevseldir (Saito & Rehmsmeier, 2015). Bu doğrultuda, Rastgele orman modeli hem ROC AUC değeri hem de F1 skorunda gösterdiği yüksek başarı ve

tutarlılık sayesinde, bu çalışmada kullanılan öğrenci etkileşim verisi üzerinde en uygun sınıflandırıcı olarak önerilmektedir.

4.1.8 Model Açıklanabilirlik Analizi

Makine öğrenmesi modellerinin performansı kadar, karar verme süreçlerinin şeffaflığı da eğitimsel karar destek sistemlerinde büyük önem taşımaktadır. Bu kapsamda, SHAP (SHapley Additive exPlanations) algoritması kullanılarak eğitimsel veri madenciliği (EDM) veri seti üzerindeki model açıklanabilirliği analiz edilmiştir.

Aşağıdaki grafikte, modelde yer alan değişkenlerin öğrenci başarısı üzerindeki ortalama etkileri gösterilmektedir. SHAP summary plot, her bir öğrencinin model tahminine en çok hangi değişkenlerin etki ettiğini ve bu etkinin yönünü göstermektedir.



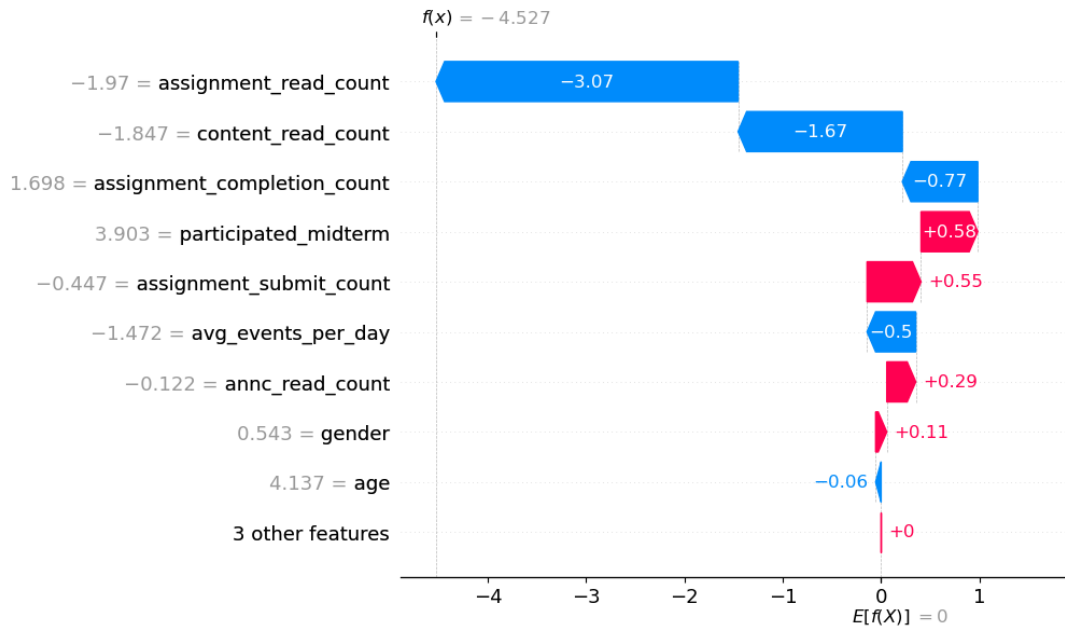
Şekil 4.15- SHAP Özellik Önem Grafiği

Bu çalışmada, öğrenci başarısını tahmin etmek için kullanılan modelin SHAP analizi, aktif katılım davranışlarının (ödev teslimi, derse katılımı) başarıyı genel olarak pozitif yönde etkilediğini, buna karşın okuma tabanlı pasif davranışlara ilişkin bazı

değişkenlerde yüksek değerlere sahip öğrencilerde negatif SHAP etkilerinin gözlenmesi dikkat çekicidir. Bu durum, sadece içerik okuma veya materyal görüntüleme gibi niceliksel olarak yüksek ancak etkileşimsel derinliği düşük olan davranışların model tarafından başarı işareti olarak algılanmadığını göstermektedir. Başka bir deyişle, öğrencinin çok sayıda içeriği okuması, bu davranış *ödev teslimi* veya *sınavlara katılım*, *derse katılım* gibi aktif performans çıktılarıyla desteklenmediğinde, model tarafından başarısızlık riski olarak yorumlanabilmektedir.

Bu bulgu, eğitim literatüründe vurgulanan "yüzeysel öğrenme" ile "etkileşimli ve derin öğrenme" arasındaki ayrımı desteklemektedir (Biggs & Tang, 2011). Öğrencilerin sadece içerik okuma alışkanlıkları, onların öğrenme derinliğini garanti etmemekte; aksine, bu okuma davranışları sınav katılımı veya ödev teslimi gibi aktif göstergelerle desteklenmediğinde model başarısızlık tahmini yapmaktadır.

SHAP şelale grafiği, tek bir öğrencinin model çıktısını etkileyen değişkenlerin katkısını göstermektedir. Aşağıdaki örnekte, öğrencinin participated_midterm (vizeye katılım), assignment_completion_count (tamamlanmış ödev sayısı) ve annc_read_count (okunan duyuru sayısı) gibi değişkenlerinin pozitif katkısı sayesinde başarı sınıfına tahmin edildiği görülmektedir.



Şekil 4.16 - SHAP Şelale Grafiği – Bireysel Tahmin

Bu analiz, sadece içerik okuma ya da sistemde geçirdiği süreye değil, sınavlara katılım, ödev tamamlama ve duyuru takibi gibi öğrenme ile doğrudan ilişkili göstergelere sahip öğrencilerin daha yüksek başarı ihtimaline sahip olduğunu göstermektedir.

SHAP analizleri, sadece modelin iç mantığını değil, aynı zamanda eğitimciler için yorumlanabilir karar destek çıktıları üretmektedir. Bu bağlamda, SHAP ile yapılan açıklamalar:

- Eğitimcilere öğrencilerin hangi davranış kalıpları ile risk altında olduklarını,
- Hangi davranışların başarıya götüren yollarla örtüştüğünü,
- Hangi öğrenci profillerine yönelik müdahalelerin öncelikli olması gerektiğini

gösterme potansiyeline sahiptir.

Modelin `assignment_completion_count` (tamamlanmış ödev sayısı), `participated_midterm` (vizeye katılım) gibi değişkenlere daha fazla ağırlık vermesi, akademik başarı ile öğrenme sürecindeki aktif katılım davranışları arasındaki güçlü ilişkiyi ortaya koymaktadır. Bu da pedagojik olarak değerlendirildiğinde, öğrencilerin sadece içerik tüketen değil, üretken ve katılımcı roller üstlenmeleri gerektiğine işaret etmektedir.

Bu çalışmada SHAP analizi, en yüksek performans gösteren model (RF) için gerçekleştirilmiştir. Literatürde model açıklanabilirliğine yönelik analizler genellikle en iyi sonuç veren algoritma ile sınırlandırılmakta ve karar destek sistemleri açısından en güvenilir modelin iç işleyişine odaklanmaktadır. Bu nedenle diğer modellerde SHAP uygulanmamış, açıklanabilirlik açısından temsil gücü en yüksek olan model detaylı analiz edilmiştir.

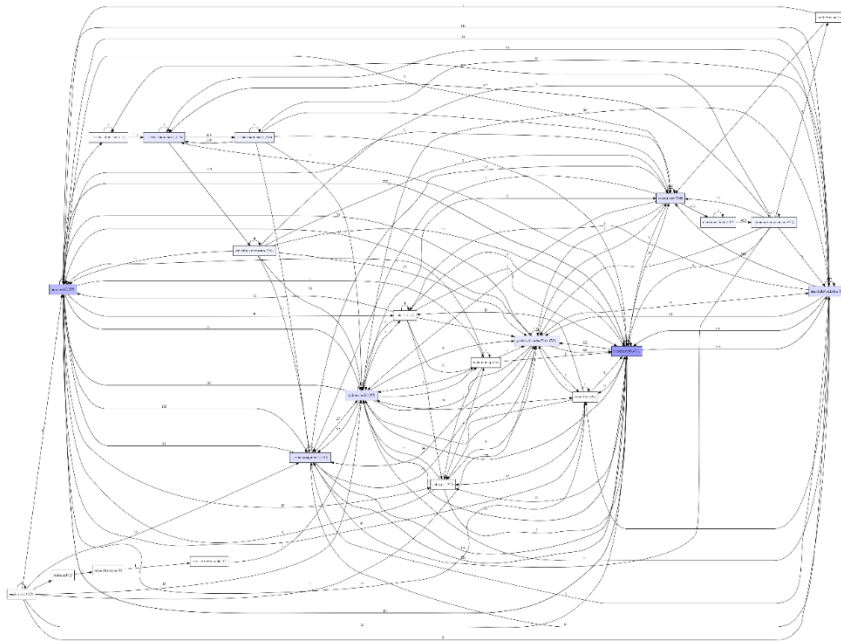
Bu bölüm, SHAP analizleri aracılığıyla geliştirilen modelin hem öğrenci başarısını anlamada şeffaf ve açıklanabilir olduğunu, hem de eğitimcilerin veri temelli kararlar almasına yardımcı olabileceğini göstermektedir. Ayrıca, SHAP gibi XAI tekniklerinin eğitimsel bağlamda kullanımı, yapay zeka uygulamaları ile pedagojik teori arasında yapılandırıcı bir köprü kurma potansiyeli taşımaktadır.

4.2 Süreç Madenciliği Tekniklerinin Uygulanması

Bu bölümde, öğrenme yönetim sistemi (LMS) verileri üzerinde süreç madenciliği yöntemlerinin uygulanmasına ilişkin adımlar ayrıntılı olarak sunulmaktadır. Süreç madenciliği, öğrenci davranışlarının zaman sıralı etkinlik kayıtları üzerinden analiz edilmesine olanak tanıyarak, öğrenme sürecinin dinamik yapısını görünür kılmaktadır. Bu kapsamda öncelikle tüm öğrencilerin katıldığı genel süreç modeli çıkarılmış; ardından yalnızca başarılı olan öğrencilerin ortak öğrenme yolları analiz edilmiştir. Süreci tamamlayıcı olarak, başarısız bir öğrenciye ait özgül bir olay kümesinin vaka (case) analizi yapılmış ve bu öğrenci ile genelin süreç örüntüsü karşılaştırılmıştır. Son olarak, çıkarılan süreç modellerinin ideal öğrenme yolu ile ne ölçüde örtüştüğünü belirlemek amacıyla süreç uyum skorları hesaplanarak süreçlerin uyum düzeyi değerlendirilmiştir. Böylece öğrencilerin platform üzerindeki davranışlarının süreçsel açıdan nasıl şekillendiği ve başarı ile nasıl ilişkilendiği bütünsel bir bakış açısıyla ele alınmıştır.

4.2.1 Genel Süreç Yapısı

Şekil 4.17’de tüm öğrencilerin etkinliklerinden çıkarılan genel süreç modeli. Düğümler, öğrenme platformundaki başlıca etkinlikleri (örn. içerik okuma, duyuru takibi, ödev gönderme, sınavlar) temsil etmekte; oklar ise bu etkinlikler arasındaki geçişleri ve üzerlerindeki sayılar ilgili geçişin gerçekleşme sıklığını göstermektedir.



Şekil 4.17 - Genel Eğitim Süreç Modeli

Genel süreç modeli, öğrencilerin dönem boyunca izledikleri öğrenme yolunu bütüncül bir akış olarak ortaya koymaktadır. Model, başlangıç noktasından (ör. ilk oturum açma veya ders tanıtım sayfasına erişim) başlayarak öğrencilerin çeşitli etkinlikler arasında nasıl dolaştığını göstermektedir. İlk olarak ders planı (syllabus) okuma ve duyuru (announcement) okuma gibi hazırlayıcı adımlar görülmekte, ardından ders içeriği okuma, ödev/ödev yanıtı görüntüleme gibi ana öğrenme aktivitelerine geçilmektedir. Süreç akışı katı bir lineer sıra izlememekte, öğrenciler içerikler, duyurular, ödevler ve not görüntüleme arasında esnek geçişler yapabilmektedir. Modelin ilerleyen aşamalarında sınav etkinlikleri (ara sınav ve final) belirgin düğümler olarak ortaya çıkar; bu noktalar, sürece dallanma ekleyerek öğrencinin sınav performansına göre farklı yollar (başarılı olup dönemi tamamlama veya bütünleme sınavına girme gibi) izlemesine neden olmaktadır. Süreç, her bir öğrenci için dönem sonunda meydana gelen başarı durumuyla (dersi geçme veya kalma) sonlanmaktadır.

Ortaya çıkarılan modelin yapısı, genel hatlarıyla iki katmanlı değerlendirilebilir: Birinci katmanda öğrencilerin düzenli öğrenme faaliyetleri (içerik inceleme, duyuru takibi, ödev hazırlama vb.) yer alırken, ikinci katmanda değerlendirme ve sonuç adımları (sınavlar ve başarı durumları) bulunmaktadır. Öğrenciler dönemi tamamlayabilmek için ilk katmandaki faaliyetlerle sürekli etkileşim halinde olmuş ve kritik noktalarda (sınav zamanlarında) ikinci katmandaki değerlendirici adımlara geçmiştir. Bu yapı, öğrenme sürecinin süreklilik arz eden etkinlikler ile belirli aralıklarla yapılan değerlendirmelerin birleşiminden oluştuğunu göstermektedir.

Bu süreç modelinden, öğrencilerin çevrimiçi öğrenme ortamındaki başlıca davranış örüntüleri şu şekilde özetlenebilir:

- **İçerik Okuma:** Öğrenciler ders materyallerine yoğun biçimde erişmektedir. Content.read etkinliği en sık gerçekleşen aktivitelerden biridir; bu da öğrencilerin ders notlarını, sunumları veya diğer öğrenme materyallerini tekrar tekrar incelediğini göstermektedir. İçerik okuma eylemleri genellikle tüm dönem geneline yayılmış ve birden fazla kez tekrarlanmıştır, bu da sürekli bir öğrenme çabasına işaret etmektedir.
- **Duyuru Takibi:** Announcement.read (duyuru okuma) etkinliği de çok sık gözlemlenmiştir. Öğrenciler eğitmen tarafından yapılan duyuruları düzenli olarak

takip etmektedir. Bu davranış, öğrencilerin dersle ilgili güncellemelere ve hatırlatıcılara karşı dikkatli olduklarını ve iletişime açık kaldıklarını göstermektedir. Duyuru okumaların sıklığı, ders boyunca önemli bilgi akışının takip edildiğine işaret eder.

- **Syllabus İncelemesi:** Dönem başında (ve muhtemelen kritik dönemecelerde) syllabus.read etkinliği dikkat çekmektedir. Öğrenciler ders planını başlangıçta okuyarak dersin hedefleri, takvimi ve kurallarını öğrenmiş, gerektiğinde tekrar syllabus'a dönmüşlerdir. Syllabus'un birden çok kez görüntülenmesi, öğrencilerin dönem içinde kural ve tarihlere tekrar ihtiyaç duyduğunu ya da belki de belirsizlik durumlarında rehber olarak kullandığını düşündürür.

- **Ödev/Hazırlık Çalışmaları:** Şekil 4.18'te, öğrencilerin ödev gönderme süreci kapsamında gerçekleştirdiği etkinliklerin süreç modeli yer almaktadır. Model, asn.read.assignment (ödev yönergesini okuma) adımı ile başlamaktadır. Bu adımdan sonra öğrenciler, ödev gönderimi için farklı yollar izlemektedir. Süreç akışı içinde asn.new.submission (yeni ödev yanıtı oluşturma), asn.save.submission (taslak kaydetme), asn.revise.submission (yanıt düzenleme) ve asn.read.submission (yanıt görüntüleme) gibi adımlar döngüsel bir yapı sergilemektedir. Bu durum, öğrencilerin ödev gönderim sürecini iteratif bir şekilde yürüttüklerini, ödev üzerinde birden fazla kez çalıştıklarını ve teslim etmeden önce defalarca düzenlediklerini göstermektedir. Süreç, asn.submit.submission (ödev teslimi) etkinliğiyle sonlanmakta, bu da öğrenme platformundaki etkin katılımı ve öz düzenlemeli öğrenme davranışlarını yansıtmaktadır.

Modelin yapısı, öğrencilerin ödevi doğrudan teslim etmek yerine, onu aşamalı olarak geliştirdiklerini; bu sürecin ise aktif öğrenme ve hatalardan geri bildirim yoluyla iyileştirme stratejileri ile örtüştüğünü göstermektedir. Süreçteki döngüler, öğrencilerin çalışma süreçlerinde deneme-yanılma ve revizyon temelli bir yaklaşımı benimsediklerini göstermesi açısından eğitimsel anlamda değerli bir bulgudur.

doğrudan geçiş ilişkisi tespit edilmiştir. Bu denli zengin bir geçiş ağı, modelin düğümleri arasında neredeyse her etkinliğin herhangi bir diğeriyle ardışık gelebilme ihtimaline işaret etmektedir. Başka bir deyişle, öğrenciler platformda farklı aktiviteler arasında *önceden belirlenmiş katı bir sıra olmaksızın* dolaşmışlardır; bu da öğrenme sürecinin bireysel tercihlere ve anlık ihtiyaçlara göre şekillenebildiğini gösteren *esnek ve dallanmış bir akış yapısı* ortaya çıkarmaktadır.

Merkezi Düğümler ve Akışlar: Modelde belirli etkinlikler *merkezî düğümler* olarak öne çıkmaktadır. Özellikle *içerik okuma ve duyuru okuma* eylemleri, diğer pek çok etkinlikle çift yönlü geçiş bağlantılarına sahip olup adeta bir hub (merkez) rolü görmektedir. Öğrenciler sık sık içerik ve duyuru sayfaları arasında gidip gelmiş, gerektiğinde ders planına (syllabus) dönmüş ve bu temel bilgi kaynaklarını sürekli kullanmışlardır. Benzer şekilde, *öğrenci not görüntüleme* (gradebook) etkinliği de farklı noktalarda (örn. bir sınav veya ödev sonrasında) akışa girerek içerik veya duyuru incelemeleriyle bağlantılanmaktadır. *Ödevle ilgili işlemler* (ödev yönergesi okuma, yanıtları düzenleme ve gönderme) hem kendi aralarında yoğun bir etkileşim ağı sergilemekte, hem de içerik okuma ve duyuru takibi gibi diğer etkinliklerle ilişki kurmaktadır. Bu durum, ödev yaparken öğrencilerin sıkça ders notlarına geri döndüğünü veya duyuruları kontrol ettiğini, dolayısıyla farklı kaynakları entegre biçimde kullandığını göstermektedir.

Tekrarlar ve Döngüler: Süreç akışında belirgin *döngüsel desenler* göze çarpmaktadır. En yaygın döngülerden biri, içerik okuma ve duyuru okuma etkinliklerinin ardışık tekrarlarıdır (aynı etkinliğin art arda birçok kez gerçekleşmesi). Öğrencilerin bir oturumda art arda birçok içerik sayfası veya birden fazla duyuruyu okuduğu, modelde bu etkinliklerin kendileriyle yinelemeli bağlantılarından anlaşılmaktadır. Bir diğer önemli döngü, *ödev revizyon süreci* etrafında oluşmaktadır: Öğrenciler yeni bir ödev yanıtı oluşturduktan sonra (asn.new.submission), bunu defalarca kaydedip yeniden düzenleyerek (asn.save.submission ↔ asn.revise.submission döngüsü) geliştirmekte ve en sonunda ödevi göndermektedir (asn.submit.submission). Bu *iteratif döngü*, öğrenme sürecindeki deneme-geri bildirim alma-korreksiyon yapma döngüsüne karşılık gelmekte ve öğrencilerin ödevleri üzerinde derinleşerek çalıştığını ortaya koymaktadır.

Dallanma (Koşullu Akışlar): Modelin en kritik dallanma noktaları, *sınavlar ve sonuçları* etrafında şekillenmektedir. Özellikle *final sınavı* sonrasında süreç iki farklı yöne ayrılmaktadır: Sınavı başarıyla geçen öğrenciler için süreç *doğrudan başarı durumuyla* son bulurken, başarısız olan öğrenciler için akış *bütünleme sınavı* etkinliğine yönelmekte ve ardından başarı/başarısızlık durumuna göre sonlanmaktadır. Bu dallanma, modelin *koşullu akış* içerdiğini ve öğrencilerin performansına bağlı olarak farklı devam yolları bulunduğunu göstermesi açısından önemlidir. Ara sınavdan sonra da benzer şekilde bazı öğrencilerin notlarını öğrenmek üzere not defterine baktığı veya duyuruları kontrol ettiği akışlar görülse de, asıl belirgin ayrışma final sonrasında meydana gelmektedir. Bunun yanında, model genelinde herhangi bir noktada öğrencinin izleyeceği bir sonraki adımın kesin olmaması (bir etkinlikten sonra birden fazla olası sonraki etkinlik bulunması), sürecin *eşzamanlı ve alternatif yollar* barındırdığını, yani öğrencilerin etkinlikleri büyük ölçüde kendi ihtiyaç ve tercihlerine göre sıraladığını göstermektedir.

Karmaşıklığın Yorumu: Süreç modelindeki yüksek bağlantı sayısı ve döngüler, eğitsel süreçlerin *doğası gereği keşfedilmemiş yapısını* yansıtmaktadır. Öğrenciler farklı öğrenme stratejileri benimsedikleri için, ortaya çıkan model tek bir “ideal” akışı değil, tüm öğrencilerin tüm etkileşimlerinin üst üste bindirilmesiyle oluşan kapsamlı bir ağı temsil etmektedir. Bu ağ yapısının karmaşıklığı, bir yandan öğrenme platformunun zengin fırsatlar sunan yapısını (öğrencilerin istedikleri sırada içeriklere ve aktivitelere erişebilmesi) ortaya koyarken diğer yandan da *süreç madenciliği* teknikleriyle eğitim verisini analiz etmenin zorluklarına işaret eder. Böyle bir model, *spagetti diyagram* benzeri bir görünüm sunarak, tekil bir öğrenci polilomik yolundan ziyade tüm popülasyonun çeşitliliğini gözler önüne sermektedir. Yine de bu karmaşık akış içerisinde belirgin merkezi davranışlar ve kritik geçişler seçilebilmekte, bu da sürecin kaotik görünümünün altında yatan anlamlı desenlerin ortaya çıkarılabildiğini göstermektedir.

4.2.1.2 Eğitimsel Süreç Açısından Değerlendirme

Elde edilen süreç modelinin eğitimsel açıdan yorumlanması, öğrencilerin ders etkinliklerine katılım biçimleri ve öğrenme yaklaşımları hakkında önemli ipuçları sunmaktadır. Model, öğrencilerin *öz-düzenlemeli öğrenme* (Zimmerman, 2002) davranışı sergilediklerini göstermektedir: Bir yandan ders materyallerini ve duyuruları

düzenli aralıklarla takip edip gerekli gördükçe tekrar gözden geçirerek *bilgi edinme ve pekiştirme* faaliyetinde bulunmuşlar, diğer yandan ödevler üzerinde tekrar tekrar çalışarak *aktif öğrenme* (Freeman, ve diğerleri, 2014) yoluyla bilgi uygulama ve beceri geliştirme süreçlerine girmişlerdir. Bu durum, literatürde vurgulanan etkin öğrenme stratejileriyle tutarlıdır; zira öğrenciler sadece pasif bir şekilde içerik tüketmekle kalmamış, aynı zamanda aldıkları geribildirimler (Hattie & Timperley, 2007) doğrultusunda kendi çalışmalarını iyileştirme döngüsüne de girmişlerdir. Örneğin, bir öğrencinin bir ödevi üzerinde birçok kez değişiklik yapması, *derinlemesine öğrenmeye* ve hatalardan öğrenerek kendini geliştirmeye işaret eden olumlu bir göstergedir.

Süreç modelinde gözlenen duyuru ve izleni takibi gibi davranışlar, öğrencilerin ders organizasyonuna ve güncellemelerine hassasiyetle uyum sağladığını göstermektedir. Duyuruların sık okunması, öğrencilerin eğitmenin mesajlarına ve yönlendirmelerine önem verdiğini, ders planının tekrar tekrar incelenmesi ise öğrencilerin zaman yönetimi ve gereklilikler konusunda dikkatli davrandığını düşündürmektedir. Bu tür davranışlar, öğrencilerin öğrenme ortamında *proaktif* ve *bilinçli* bir tutum benimsediklerini gösterir; ki bu da akademik başarıyla pozitif yönde ilişkili bulunan bir tutumdur.

Sınav etkinliklerinin modele yansımaları, sürekli devam eden öğrenme faaliyetleri ile dönem sonu değerlendirmelerinin ilişkisini ortaya koyar niteliktedir. Ara sınav ve final, öğrencilerin öğrenme yolculuğunda kilit dönüm noktaları olarak görülmekte, öğrenciler bu sınavlar öncesinde ve sonrasında farklı tepkiler vermektedir. Örneğin, ara sınav sonrasında notlarını hemen kontrol eden veya ek duyuruları inceleyen öğrenciler, performanslarına yönelik geribildirim arayışında bulunmuştur. Final sınavı sonrasında daldanma ise eğitimsel açıdan kritik bir farklılaşmayı gösterir: Başarısız olan öğrencilerin bütünleme sınavına yönlendirilmesi, öğrenme sürecine bir ikinci şans mekanizmasının entegre edildiğini, yani programın telafî imkânı sunarak öğrenme sürecini uzattığını göstermektedir. Bu, eğitimsel süreçlerin esneklik içerebildiğini ve öğrencilerin öğrenme hedeflerine ulaşmaları için alternatif yollar sağlanabildiğini ortaya koyar. Başarılı öğrenciler ise final sonrası doğrudan süreci tamamlayarak, süreklilik arz eden çalışmanın arzu edilen çıktıya (başarıya) ulaştığını göstermiştir.

Öğrencilerin *not defterini sık sık kontrol etmesi*, eğitimsel anlamda onların kendi ilerlemelerini *izleme ve değerlendirme* ihtiyacında olduklarını gösterir. Bu davranış, öğrencilerin hedeflerini ve mevcut durumlarını karşılaştırarak gerektiğinde çalışma stratejilerini ayarlama çabasının bir yansıması olabilir. Öğrencilerin kendi öğrenme süreçlerine dair farkındalık geliştirmeleri (ki not takibi bunun bir göstergesidir), öz-yönelimli öğrenme becerilerinin gelişmiş olduğunu ima eder. Dolayısıyla, süreç modelindeki bu unsur, öğrencilerin yalnız içerik ve ödevlerle değil, aynı zamanda kendi öğrenme süreçleriyle de meşgul olduklarını, metabilişsel bir süreç yaşadıklarını göstermektedir.

Genel olarak, elde edilen bulgular eğitimsel süreçlerin çok boyutlu ve dinamik bir yapıda ilerlediğini doğrulamaktadır. Öğrencilerin platform içindeki serbest hareket alanı, her birine kendi öğrenme yollarını kurgulama imkânı tanımış; ancak tüm bu farklı yollar neticede benzer kilometre taşlarından (ödevlerin tamamlanması, sınavlara girilmesi gibi) geçerek dönemin başarıyla sonlandırılması veya sonlandırılmamasına yol açmıştır. Bu da öğrenme süreçlerinde çeşitlilik olsa bile temel öğrenme hedefleri ve değerlendirme noktalarının ortak olduğunu gösteren önemli bir tespittir. Eğitimsel süreç analizi perspektifinden, bu model hem başarılı öğrencilerin ortak eğilimlerini hem de başarısız kalanların potansiyel eksiklik gösterdiği noktaları (örneğin içeriklere yeterince erişmeme ya da ödevlere az zaman ayırma gibi, modelde dolaylı olarak çıkarılabilecek durumlar) incelemek için bir temel sunabilir.

Yapılan süreç madenciliği analizi, öğrenci etkinlik kayıtlarından bütünleştirilmiş karmaşık bir süreç modeli ortaya çıkarmıştır. Modelde belirlenen davranış örüntüleri, öğrencilerin aktif ve sürekli bir öğrenme angajmanı içinde olduğunu göstermektedir. Özellikle ödevler üzerinde birden fazla kez düzenleme yapıldığı ve içeriklere tekrar tekrar dönüldüğü gözlemlenmiştir. Bu durum, öğrencilerin aktif öğrenme ve öz-düzenleme becerileri geliştirdiklerine işaret etmektedir.

Sonuç olarak, tüm öğrencilerin etkinlik verilerinden çıkarılan bu kapsamlı süreç modeli, *öğrenme sürecinin çok boyutlu dinamik bir haritası* niteliğindedir. Modelin akademik açıdan incelenmesi, çevrimiçi öğrenme ortamlarında öğrencilerin nasıl etkileşimde bulunduğu dair derinlemesine bir anlayış sunmaktadır. Bu anlayış, öğretim tasarımının iyileştirilmesi, öğrenci desteğinin kişiselleştirilmesi ve kritik

başarı faktörlerinin belirlenmesi konularında değerli içgörüler sağlayabilir. Model, hem genel anlamda öğrencilerin ders boyunca izledikleri yolu ortaya koyarak eğitimcilere *genel bir performans resmi* sunmakta, hem de mikro düzeyde tekrar eden davranış kalıplarını göstererek *öğrenme deneyiminin niteliğine* ışık tutmaktadır. Bu tür bir süreç analizi, eğitimsel veri madenciliği literatüründe öğrenci davranışlarının ve öğrenme süreçlerinin anlaşılmasına yönelik bulgularla uyumlu olup, elde edilen sonuçlar tez kapsamında *öğrenme analitiği ve eğitsel süreç geliştirme* bağlamlarında tartışılabilir düzeyde kapsamlı ve değerlidir.

4.2.2 Başarılı Öğrencilerin Süreç Modeli (İdeal Öğrenme Süreci)

A ve A- notu almış 10 öğrenciden elde edilen süreç modeli, oldukça sistemli ve yapılandırılmış bir öğrenme yolunu yansıtmaktadır. Süreç, tüm öğrencilerde session.start etkinliğiyle başlamakta, ardından sırasıyla duyuru okuma (annc.read), syllabus inceleme (syllabus.read) ve içerik okuma (content.read) gibi ön hazırlık etkinliklerine geçilmektedir. Bu yapı, öğrencilerin ders sürecine bilinçli şekilde hazırlandığını göstermektedir. Akış daha sonra ödev hazırlama, içerik güncellemeleri ve sınavlara katılım gibi değerlendirme odaklı etkinliklerle devam etmektedir. Süreç, tüm öğrenciler için exam.final sonrası doğrudan status.pass (başarı) ile sonlanmaktadır. Bu da ideal grupta bütünleme sınavına gerek kalmadan başarıya ulaşıldığını göstermektedir. Süreç modeli Şekil 4.19’te gösterilmiştir.

studentView aktivitesi öğrencilerin notlarını düzenli olarak takip ettiklerini, yani kendi başarı durumları hakkında farkındalık geliştirdiklerini ortaya koymaktadır.

4.2.2.1 Karmaşıklık ve Akış Analizi

Model, genel olarak düşük karmaşıklık düzeyine sahiptir. Dallanmalar ve geri dönüşler sınırlı sayıda olup, süreç adımları arasında belirli bir düzen ve yönelme söz konusudur. Merkezî düğümler, özellikle content.read, ann.read, asn.read.assignment ve asn.revise.submission gibi aktivitelerdir. Bu düğümler, süreç içerisinde birçok başka etkinlikten önce ya da sonra yer almakta ve sürecin odak noktaları haline gelmektedir. Döngüsel yapılar özellikle ödev sürecinde görülmektedir: yeni yanıt oluşturma, düzenleme, tekrar kaydetme gibi etkinlikler birkaç kez birbirini takip etmektedir. Ancak bu döngüler, modelin sadeliğini bozmamakta; aksine öğrenme sürecinin derinlemesine yürütüldüğünü göstermektedir.

4.2.2.2 Eğitimsel Süreç Açısından Değerlendirme

İdeal öğrenci grubunun süreci, eğitim bilimlerinde tanımlanan öz-düzenlemeli öğrenme ve derin öğrenme stratejileriyle tam uyum içerisinde. Öğrenciler, yalnızca içeriği tüketmekle kalmamış, ödevlerini birçok kez revize etmiş, sınavlara zamanında katılmış ve kendi ilerlemelerini not sisteminden takip etmiştir. content.updated, new.version gibi içerik güncellemelerinin dahi takip edilmesi, öğrencilerin sürekli olarak sistemde aktif ve dikkatli olduğunu göstermektedir. Sınav sürecinde doğrudan exam.final etkinliğinin ardından status.pass sonucuna ulaşılması, öğrenme sürecinin sadece düzenli değil aynı zamanda etkili bir şekilde yürütüldüğünü ortaya koymaktadır. Bu süreç yapısı, eğitimciler açısından ideal bir öğrenme davranışı örüntüsünü temsil etmektedir.

İdeal öğrenme yolunu temsil eden süreç modeli, A ve A- notu almış öğrencilerin çevrimiçi öğrenme platformundaki davranış örüntülerini yansıtmaktadır. Öğrenciler sürece bilinçli bir şekilde hazırlık yaparak başlamış; ders planı, duyurular ve içerikleri sistemli biçimde incelemişlerdir. Özellikle ödev hazırlama sürecinde tekrarlı düzenleme ve gönderim davranışı, derinlemesine öğrenme stratejilerinin uygulandığını göstermektedir. İçeriklere yüksek erişim sıklığı ve sınavlara zamanında katılım, öğrenme sürecinin süreklilik ve başarıyla yürütüldüğünü ortaya koymaktadır. Modelin düşük karmaşıklık düzeyi ve merkezî odakları, öğrenmenin yapılandırılmış

bir çerçevede gerçekleştiğini desteklemektedir. Bu yapı, çevrimiçi öğrenme ortamlarında öğrenci başarısının ardındaki temel davranışsal dinamikleri belirlemede rehberlik edebilecek niteliktedir.

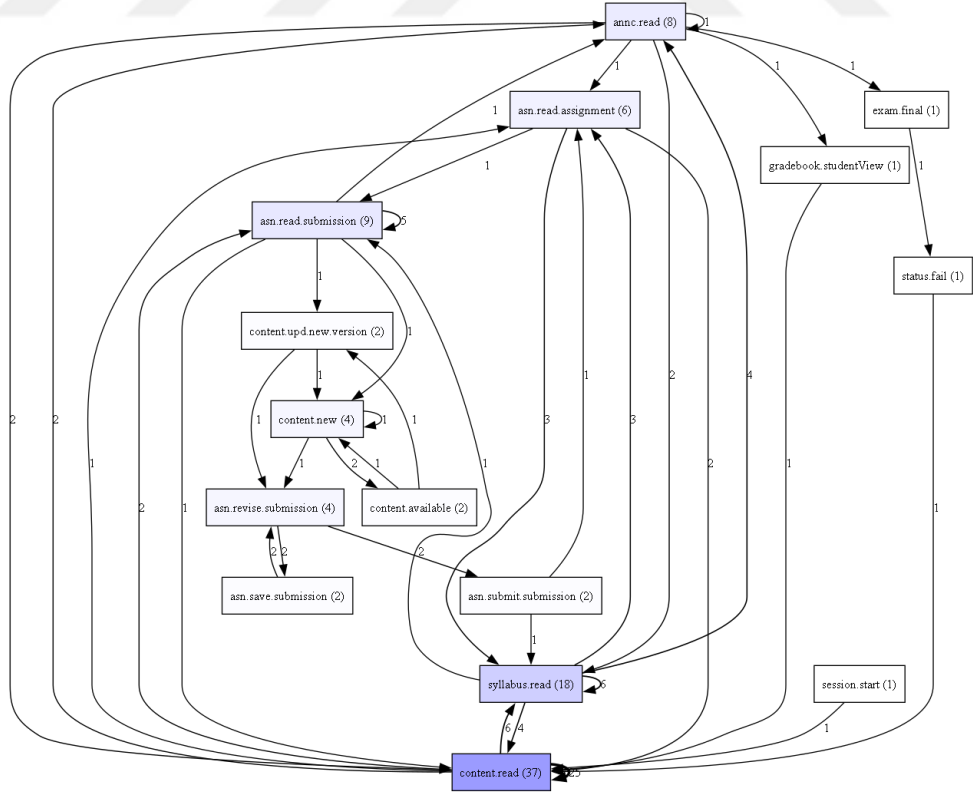
4.2.3 Başarısız Bir Öğrenci: Vaka Analizi

Bu vaka analizinde ele alınan öğrenci, başarısızlıkla sonuçlanan öğrenciler arasından, toplam etkileşim (event) sayısı bakımından medyan değere en yakın olan birey olarak seçilmiştir. Bu sayede, ne minimum düzeyde ne de yüksek düzeyde platform kullanımına sahip olmayan; dolayısıyla “ortalama katılım düzeyinde” bir başarısızlık örüntüsünü temsil etmektedir. Böyle bir örnek, öğrencinin başarısızlığının yalnızca etkileşim miktarıyla değil, bu etkileşimlerin niteliksel ve yapısal özellikleriyle de ilişkili olup olmadığını değerlendirmek açısından önem taşımaktadır. Bu süreç modeli, ortalama düzeyde etkileşim göstermiş ancak dersi başarısızlıkla tamamlamış bir öğrencinin etkinlik geçmişini yansıtmaktadır. Süreç, yalnızca bir kez gerçekleşen session.start etkinliğiyle başlamakta ve status.fail düğümüyle sonlanmaktadır. Öğrenci; içerik okuma, syllabus inceleme, bazı ödev işlemleri ve duyuru takibi gibi çeşitli etkinliklerde bulunmuş ancak genel olarak süreç düzensiz ve dağınık bir yapıya sahiptir. Sınav katılımı (exam.final) yalnızca bir kez gerçekleşmiş ve doğrudan status.fail sonucuyla sonlanmıştır. Süreçte dikkat çeken bir diğer unsur ise çoğu etkinliğin yalnızca birkaç kez yapılmış olması ve merkezî bir davranış örüntüsü oluşturmamasıdır.

Bu öğrenci:

- **İçerik okuma (content.read)** etkinliğini 37 kez gerçekleştirmiştir. Bu sayı, niceliksel olarak yeterli görünse de öğrenme başarısına dönüşmemiştir. Bu durum, *pasif içerik tüketiminin* derin öğrenme için yeterli olmadığını göstermektedir.
- **Syllabus (syllabus.read)** ve **duyuru (annc.read)** etkinliklerine başvurmuştur (sırasıyla 18 ve 8 kez), ancak bu davranışlar da öğrenme sürecinin planlı yürüdüğüne dair güçlü bir işaret oluşturmaz.

- Ödev sürecine ilişkin olarak `asn.read.assignment`, `asn.revise.submission`, `asn.save.submission` ve `asn.submit.submission` gibi eylemleri gerçekleştirmiştir, ancak bunların gerçekleşme sıklığı düşüktür (her biri 2–4 kez).
- *Ödev revizyon döngüsünün* sınırlı olması, öğrencinin geri bildirim döngüsünü yeterince işletmediğini ve *üretken öğrenmeye* yeterince zaman ayırmadığını göstermektedir.
- *Gradebook erişimi yalnızca 1 kez* gerçekleşmiştir. Bu durum öğrencinin akademik ilerlemesini takip etmeye yönelik metabilîşsel farkındalığının oldukça düşük olduğunu göstermektedir.
- Öğrenci final sınavına katılmış ancak süreç bütünlemeye dahi gitmeden doğrudan `status.fail` ile sonlanmıştır. Bu, öğrencinin öğrenme çıktısının değerlendirme aşamasında yetersiz kaldığını, önceki etkinliklerin yeterli hazırlığı sağlamadığını göstermektedir.



Şekil 4.20 - Başarısız Bir Öğrenci Süreç Modeli

4.2.3.1 Karmaşıklık ve Akış Analizi

Model yapısal olarak düşük karmaşıklık düzeyindedir, ancak bu durum yüksek verimliliğe değil, süreçteki davranış azlığına işaret etmektedir. Dğümler arasındaki geçiş sayıları düşüktür; birçok etkinlik yalnızca bir veya birkaç kez gerçekleşmiştir. Süreç genel olarak düz çizgisel bir akıştan çok, kopuk adımlar zinciri olarak değerlendirilebilir. Döngüsel yapılar yok denecek kadar azdır. Örneğin, *asn.revise.submission* gibi etkinliklerin sadece birkaç kez tekrarlanması, ödev geliştirme sürecinin sıkı kaldığını göstermektedir. Merkezî düğüm konumunda olan tek etkinlik *content.read* gibi görünse de bu davranış tek başına öğrenme derinliği üretmemiştir. Süreç dallanma içermemekte, doğrudan bir sınavla sonuçlanan kısa bir öğrenme süreci göstermektedir.

4.2.3.2 Eğitimsel Süreç Açısından Değerlendirme

Bu modeldeki öğrenci davranışları, literatürde başarısızlıkla ilişkilendirilen öğrenme stratejileriyle örtüşmektedir. Her ne kadar içeriklere erişim sıklığı görece yüksek olsa da, bu etkileşimler *yüzeysel öğrenme* göstergesidir. *İçeriği derinlemesine işleme, yeniden yapılandırma ve uygulama süreçleri*, örneğin ödev üzerinde tekrar tekrar çalışarak kendini geliştirme (iteratif revizyon döngüsü), bu öğrencide gözlemlenmemektedir.

Not görüntüleme ve sınav sonrasındaki tepki (örneğin not kontrolü, ek içerik incelemesi gibi) davranışları ya hiç yoktur ya da oldukça sınırlıdır. Bu durum öğrencinin *geri bildirim alma, yansıtma ve strateji değiştirme* becerilerini yeterince kullanmadığını gösterir. Sürecin sonunda öğrencinin bütünlemeye yönelmemesi (yani başarısızlıktan sonra bir eylem göstermemesi), öğrenme sürecinin yalnızca pasif bilgi alımına ve tek seferlik değerlendirmeye indirgenmiş olduğunu ortaya koymaktadır. Bu tür davranışlar, öğrenme süreçlerinde *öz-düzenleme eksikliği, motivasyon düşüklüğü* veya *hedef-belirsizlik* gibi durumlarla ilişkilendirilmektedir.

4.2.3.3 İdeal Öğrenme Yoluyla Karşılaştırma

Ortalama düzeyde etkileşim göstermesine rağmen dersi geçemeyen bu öğrenciye ait süreç modeli, başarısızlıkla sonuçlanan öğrenme örüntüsünü ortaya koymaktadır.

Tablo 4.2 - İdeal Öğrenci İle Kalan Öğrenci Davranışları Frekans Karşılaştırması

Öğrenme Unsuru	Ortalama İdeal Öğrenci	Vaka Öğrencisi	Vaka Öğrencisi
İçerik Okuma (content.read)	43.7 (437 / 10)	37	Benzerdir ancak başarısız
Gradebook Takibi	10.1	1	Geri bildirim kullanımı çok düşük
Ödev Revizyonu	20+	4	Geliştirme döngüsü yetersiz
Duyuru Takibi	26.9	8	Bilgilendirme eksikliği
Finalden Sonra Aktivite	Var	Yok	Sürece tepki eksikliği
Sonuç	Geçti	Kaldı	Nitel farklılıklar belirleyici

Öğrenci içeriklere erişmiş, ders izlenesine ve duyurulara göz atmış, ödev teslimi gerçekleştirmiştir; ancak bu etkileşimlerin yüzeysel ve düzensiz olduğu gözlemlenmektedir. Özellikle ödev revizyon döngüsünün zayıf olması ve not takibinin yok denecek kadar az olması, öz-düzenlemeli öğrenme sürecinin gelişmemiş olduğunu göstermektedir. Sınav sonrası süreçte herhangi bir bütünleme girişiminin bulunmaması da öğrencinin başarısızlığa karşı stratejik bir tepki geliştirmediğini ortaya koymaktadır. Bu vaka, öğrenme başarısının yalnızca etkinlik sıklığıyla değil, bu etkinliklerin amacı, dizilimi ve sürekliliğiyle de yakından ilişkili olduğunu bir kez daha göstermektedir.

4.2.4 Süreç Uyum Skorlarının (Fitness) Hesaplanması

Bu çalışmada öğrencilerin öğrenme süreçlerinin keşfedilen modelle ne ölçüde uyumlu olduğunu belirlemek amacıyla süreç uyum skorları (fitness) hesaplanmıştır. Bu işlem, Python ortamında açık kaynaklı **PM4Py** (Process Mining for Python) kütüphanesi kullanılarak gerçekleştirilmiştir. Tüm öğrencilere ait olay dizileri (log_all) yukarıda oluşturulan süreç modeli ile karşılaştırılmıştır. Bu karşılaştırma için token-based replay yöntemi kullanılmıştır. Yani her öğrenciye ait olay sırası, model üzerinde tekrar oynatılarak süreç içerisindeki ilerleyiş başarı oranı ölçülmüştür. Bu yöntem sayesinde her öğrencinin öğrenme süreci, ideal modelle bire bir karşılaştırılmış ve modelle

örtüşme derecesi nesnel olarak hesaplanmıştır. Bu skorlar bir sonraki aşamada başarı durumu, SHAP analizi ve davranış örüntüleriyle karşılaştırılarak yorumlanmıştır.

Tablo 4.3 - Süreç Uyum Skoru Örnek Veri Seti

Öğrenci ID	Süreç Uyum Skoru
Öğrenci 1	0.9644113950565564
Öğrenci 2	0.8365384615384616
Öğrenci 3	0.738888889
Öğrenci 4	0.974151463
Öğrenci 5	0.992749278
...	...

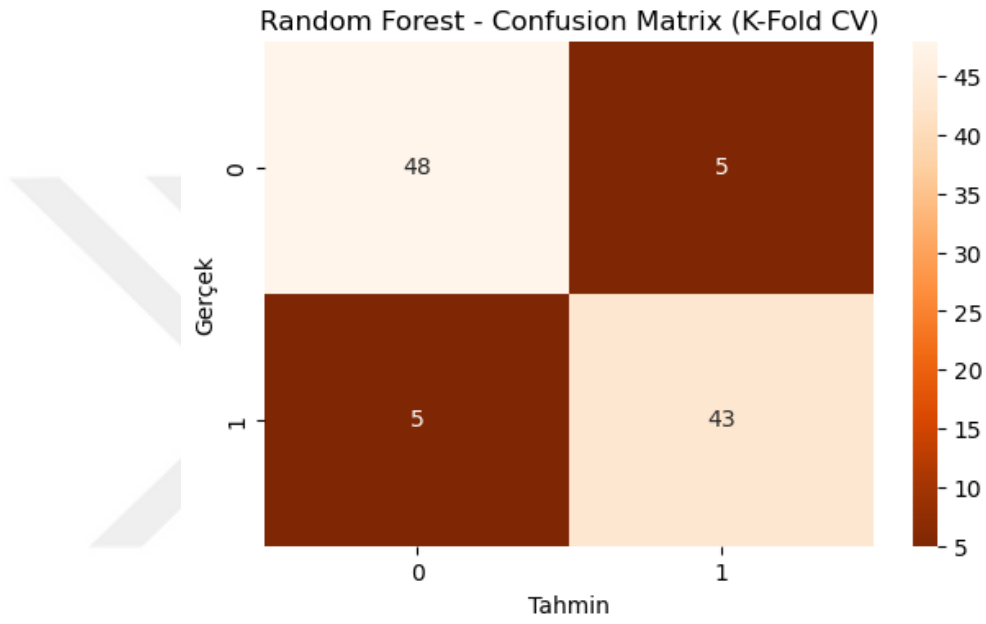
4.3 EPM Destekli Veri Seti Üzerinde Makine Öğrenmesi Modellerinin Eğitimi ve Performansı

Bu bölümde, süreç madenciliği teknikleriyle elde edilen süreç uygunluk (fitness skor), model uyumu ve davranışsal sapma gibi süreç tabanlı değişkenlerin, öğrenci akademik performansı üzerindeki tahmin gücünü artırma potansiyeli araştırılmaktadır. Bu amaçla, önceki bölümlerde kullanılan etkileşim ve demografik özelliklere ek olarak, süreç madenciliği uygulamaları aracılığıyla elde edilen süreç uyum skoru da modele dahil edilmiştir. Bu şekilde oluşturulan EPM genişletilmiş veri seti, öğrenci öğrenme süreçlerinin yalnızca nicel değil, aynı zamanda süreçsel kalite açısından da analiz edilebilmesine olanak sağlamaktadır. Rastgele orman, K-en yakın komşu (KNN), LightGBM, XGBoost, karar ağacı ve destek vektör makinesi (SVM) algoritmaları bu yeni veri seti üzerinde yeniden eğitilmiş; 80-20 eğitim/test bölmesi ve 10 katlı çapraz doğrulama stratejileri ile değerlendirilmiştir. Performans ölçütü olarak ROC AUC, F1 skoru, doğruluk ve karmaşıklık matrisi metrikleri kullanılmıştır. Elde edilen sonuçlar, süreç uygunluğunu (fitness) temsil eden değişkenlerin tahmin başarısına etkisini analiz etmeyi ve klasik EDM modellerine kıyasla genişletilmiş modelin iyileştirici katkılarını ortaya koymayı amaçlamaktadır.

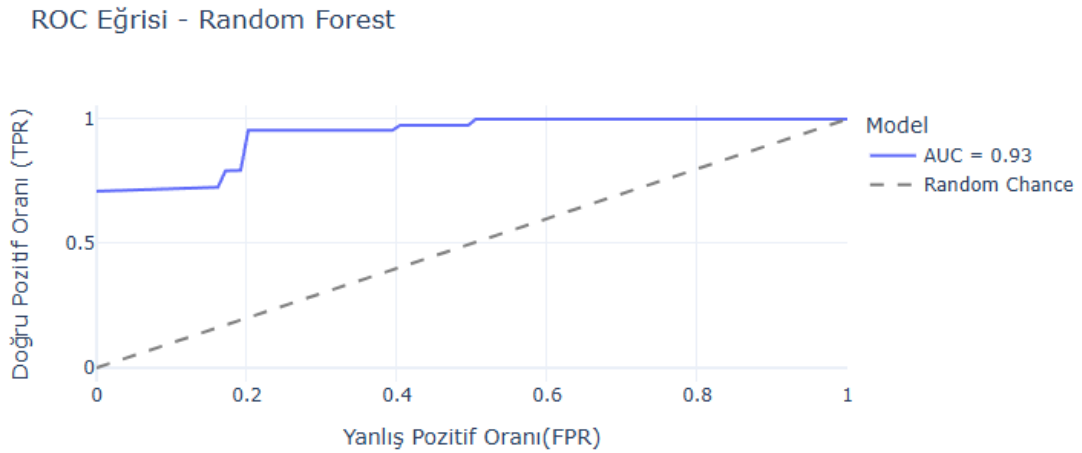
4.3.1 Rastgele Orman (RF)

Rastgele Orman algoritması, süreç madenciliği uygulamalarıyla elde edilen süreç uygunluk skoru değişkeninin de dahil edildiği genişletilmiş veri seti üzerinde eğitilmiş

ve 10 katlı çapraz doğrulama yöntemiyle değerlendirilmiştir. Bu çalışmada elde edilen ROC AUC skoru 0.93, Doğruluk (Accuracy) 0.900909, F1 Skoru 0.902859'tir. Karmaşıklık matrisi verileri şu şekildedir: TP=43, TN=48, FP=5, FN=5. Model hem doğru pozitif hem de doğru negatif oranlarını yüksek tutmuş ve ezberleme (overfitting) yapmadan sınıflar arası ayırım gücü göstermiştir. ROC eğrisi neredeyse ideal sınırlarda gezinmiş ve AUC değeri, sınıflar arasındaki ayırımın çok yüksek doğrulukla gerçekleştirildiğini göstermiştir.



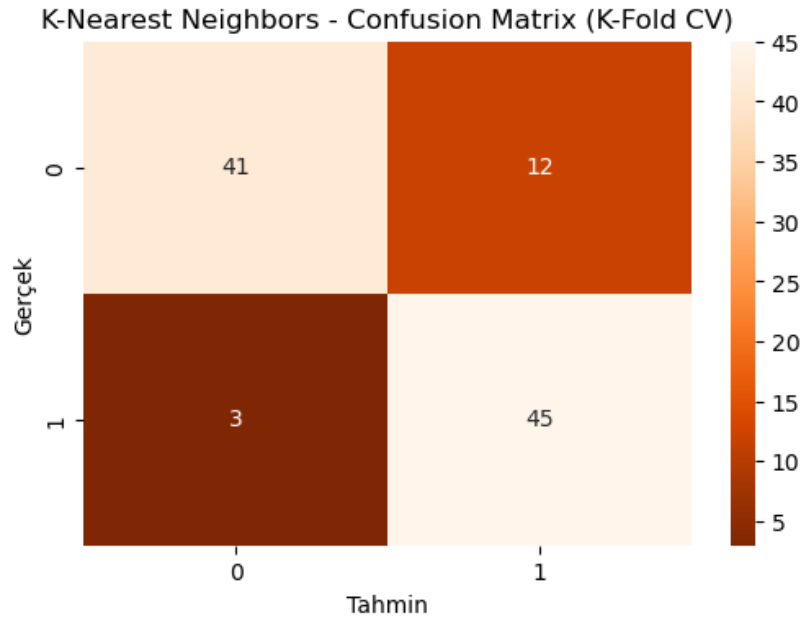
Şekil 4.21- EPM Destekli Rastgele Orman Karmaşıklık Matrisi



Şekil 4.22 EPM Destekli Rastgele Orman Roc Eğrisi Grafiği

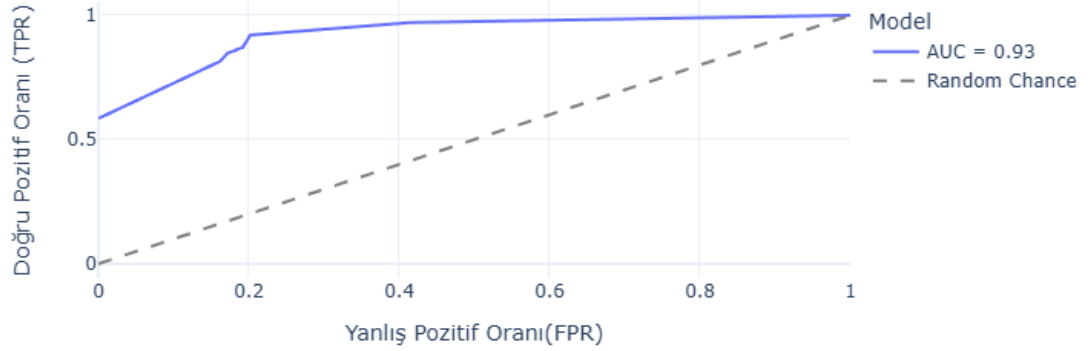
4.3.2 K-En Yakın Komşu (KNN)

Bu çalışmada, süreç madenciliği uygulamalarıyla elde edilen süreç uygunluk skoru değişkeni dahil edilerek oluşturulan veri seti üzerinde KNN algoritması uygulanmış ve 10 katlı çapraz doğrulama yöntemiyle test edilmiştir. KNN modelinin AUC skoru 0.93, Doğruluk 0.850909, F1 Skoru 0.856484 olarak elde edilmiştir. Model, doğru pozitif (TP) = 45, doğru negatif (TN) = 41, yanlış pozitif (FP) = 12, yanlış negatif (FN) = 3 olacak şekilde sınıflandırma yapmıştır. Başarılı öğrencilerin tespiti oldukça güçlüdür; yalnızca 3 öğrenci hatalı olarak başarısız sınıfına atanmıştır. Ancak 12 başarısız öğrencinin yanlışlıkla başarılı olarak tahmin edilmesi, modelin yanlış olumlu oranında (FPR) bir miktar yükseklik olduğunu göstermektedir. ROC eğrisinin şekli ve AUC = 0.93 değeri, KNN algoritmasının sınıflar arasında güçlü bir ayrım yaptığını ortaya koymaktadır. Eğri, düşük yanlış pozitif oranlarıyla yüksek doğru pozitif oranlarını bir arada sağlayacak şekilde yükselmektedir. Bu durum, KNN'nin süreç uygunluk skorunu sınıflar arasındaki farkları ayırt etmekte etkin bir biçimde kullanabildiğini göstermektedir.



Şekil 4.23- EPM Destekli Rastgele K-En Yakın Komşu Matrisi

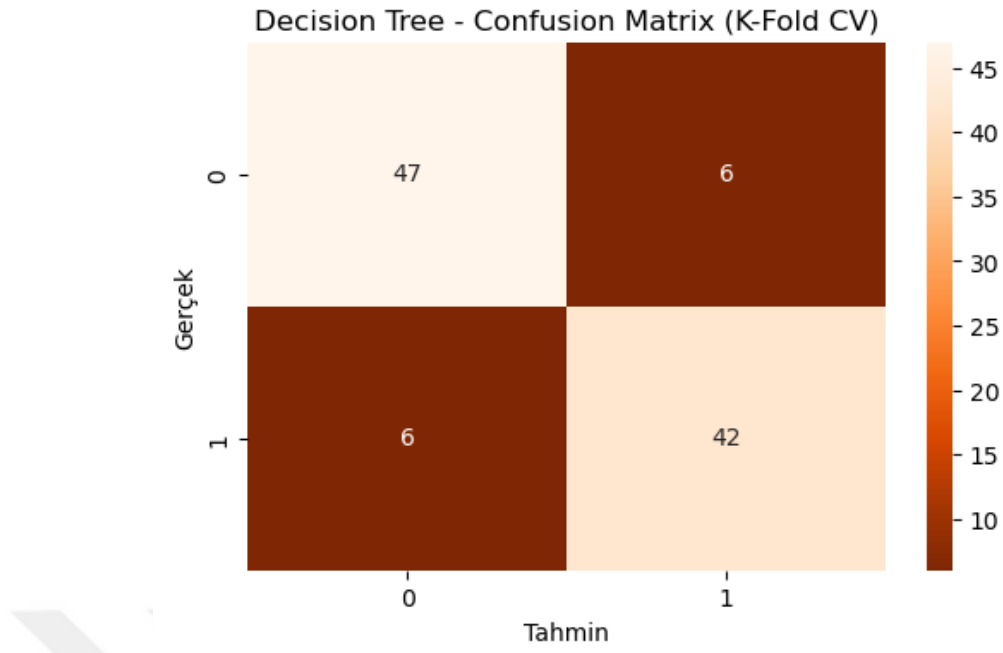
ROC Eğrisi - K-Nearest Neighbors



Şekil 4.24 - EPM Destekli K-En Yakın Komşu Roc Eğrisi Grafiği

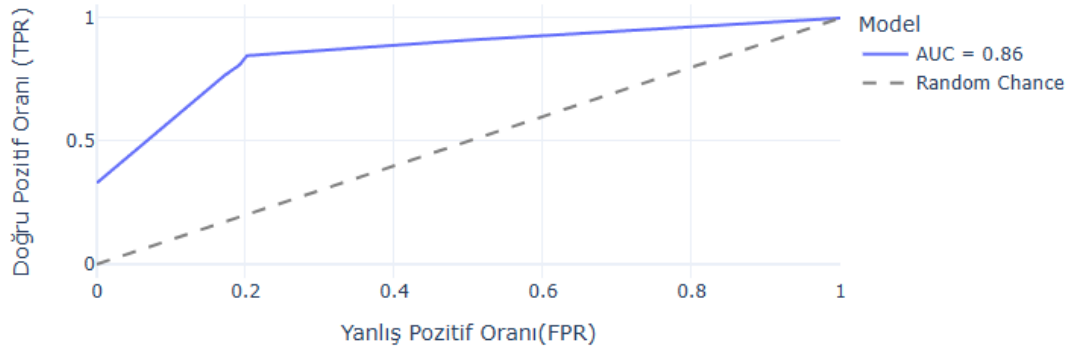
4.3.3 Karar Ağacı (DT)

Bu çalışmada, öğrenme süreçlerinin kalitesini temsil eden süreç uygunluk skoru dahil edilerek oluşturulan veri seti üzerinde Karar Ağacı algoritması eğitilmiş ve 10 katlı çapraz doğrulama yöntemi ile değerlendirilmiştir. Bu modelin AUC skoru 0.86, Doğruluk 0.880909, F1 Skoru 0.884095 olarak hesaplanmıştır. Model, doğru pozitif (TP) = 42, doğru negatif (TN) = 47, yanlış pozitif (FP) = 6, yanlış negatif (FN) = 6 şeklinde bir tahmin performansı ortaya koymuştur. Bu sonuçlar, Karar Ağacı'nın sınıflar arasında dengeyi koruduğunu, ancak bazı bireyleri sınırda bırakabildiğini göstermektedir. Yanlış sınıflandırma sayısı düşüktür, ancak ROC eğrisi doğrusal olmayan yapıda kademeli bir başarı göstermektedir. ROC eğrisi, modelin sınıflar arasındaki ayrımı kademeli fakat istikrarlı şekilde yapabildiğini göstermektedir. AUC = 0.86 değeri, Karar Ağacı'nın süreç uygunluk skoruna kısmen duyarlı olduğunu ancak ayrım gücünün sınıf sınırlarında diğer güçlü modeller kadar keskin olmadığını ortaya koymaktadır.



Şekil 4.25 - EPM Destekli Rastgele Karar Ağacı Matrisi

ROC Eğrisi - Decision Tree

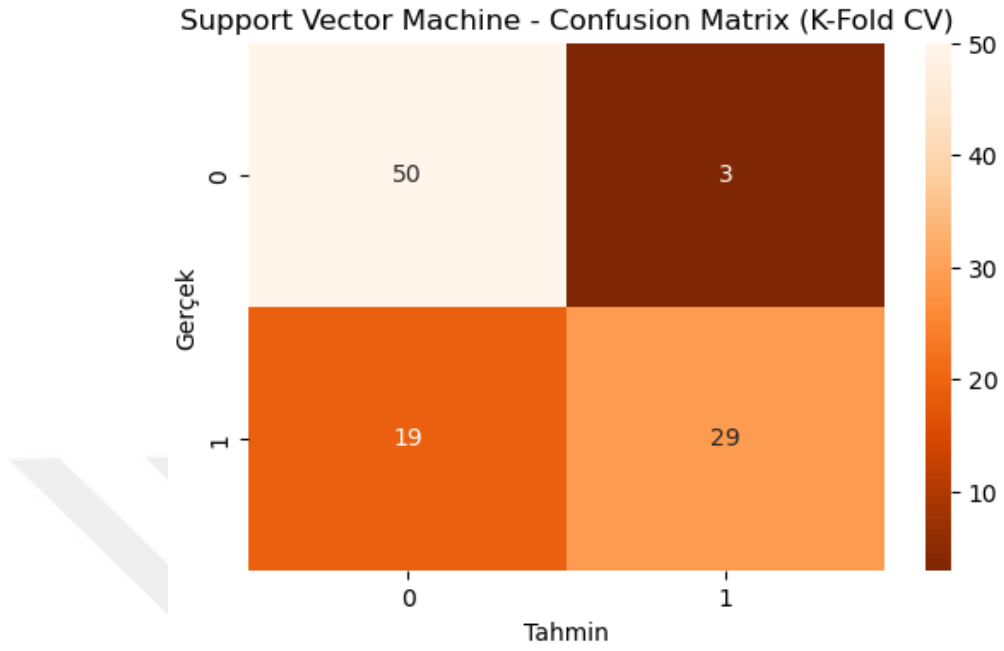


Şekil 4.26 - EPM Destekli Karar Ağacı Roc Eğrisi Grafiği

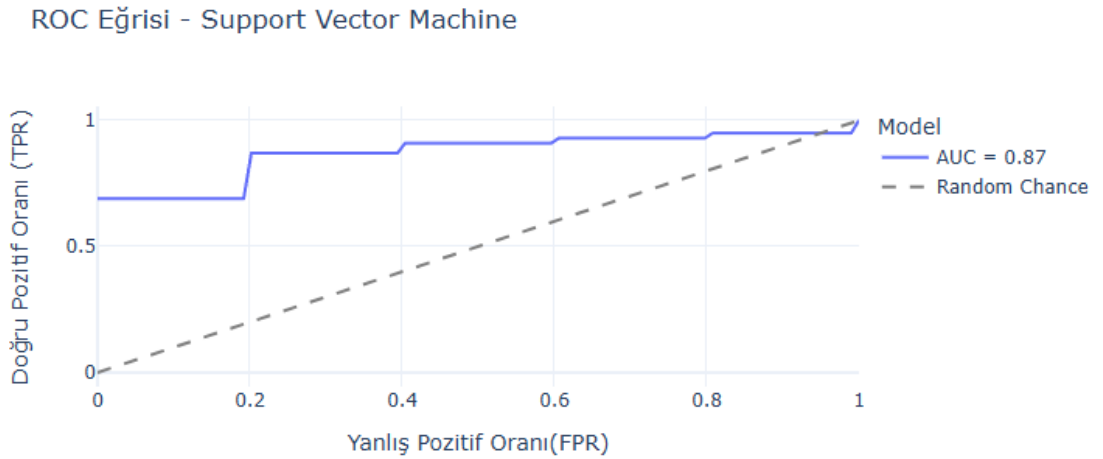
4.3.4 Destek Vektör Makinesi (SVM)

Bu çalışmada AUC skoru 0.87, Doğruluk 0.782727, F1 Skoru 0.770255 olarak elde edilmiştir. Model, doğru pozitif (TP) = 29, doğru negatif (TN) = 50, yanlış pozitif (FP) = 3, yanlış negatif (FN) = 19 tahminleri yapmıştır. SVM modeli başarısız öğrencileri doğru sınıflandırmada son derece başarılı olurken, başarılı sınıfın tespitinde ciddi düşüş göstermiştir. 19 öğrencinin yanlışlıkla başarısız olarak tahmin edilmesi, modelin pozitif sınıfa (başarılı öğrencilere) karşı düşük duyarlılık sergilediğini göstermektedir. ROC eğrisi genel olarak yukarı yönlü güçlü bir yapıya sahiptir. AUC = 0.87 değeri, modelin sınıf ayırım gücünün iyi olduğunu göstermektedir. Ancak ROC eğrisindeki

dikey kırılmalar, modelin belirli eşik değerlerinde pozitif sınıf için karar vermede kararsız kaldığını ve bazı eşiklerde başarımın düştüğünü göstermektedir.



Şekil 4.27 - Birleştirilmiş Veri Seti Rastgele Destek Vektör Makinesi Matrisi

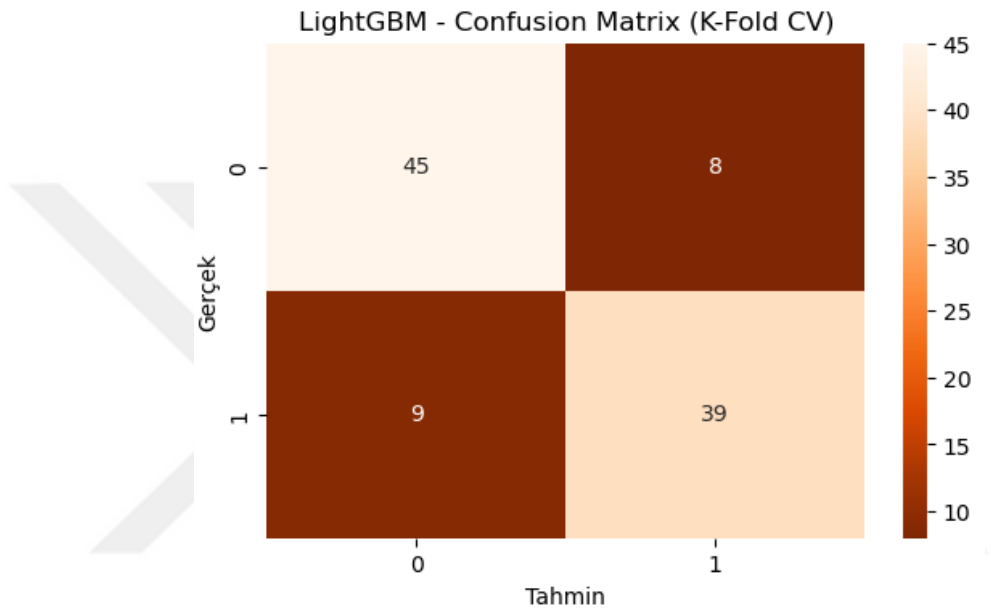


Şekil 4.28 - EPM Destekli Destek Vektör Makinesi Roc Eğrisi Grafiği

4.3.5 LightGBM

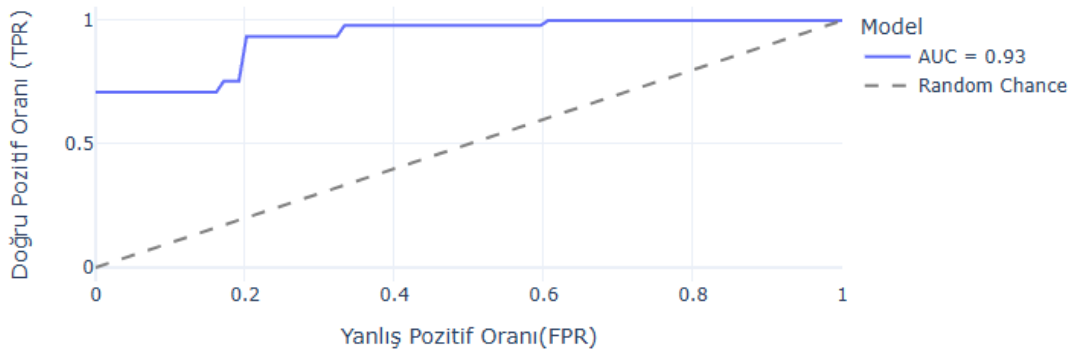
Bu çalışmada, süreç madenciliğinden elde edilen süreç uygunluk skoru gibi yapısal değişkenlerin entegre edildiği veri seti üzerinde LightGBM modeli eğitilmiştir. Bu modelin AUC skoru 0.93, Doğruluk 0.830909, F1 Skoru 0.832125 'dir. Model, doğru pozitif (TP) = 39, doğru negatif (TN) = 45, yanlış pozitif (FP) = 8, yanlış negatif (FN)

= 9 olacak şekilde tahmin üretmiştir. Pozitif ve negatif sınıflar arasında dengeli bir sınıflandırma yapabilmiş, toplam yanlış sınıflandırma sayısını 17 ile sınırlamıştır. ROC eğrisi oldukça başarılı bir eğilim göstermektedir. Eğri, düşük FPR değerlerinde bile yüksek TPR oranları ile yükselmektedir ve $AUC = 0.93$ değeri, LightGBM'in sınıflar arası ayırma gücünü güçlü performans ortaya koyduğunu doğrulamaktadır. Eğrinin doğrusal şans çizgisine olan uzaklığı, modelin tahmin gücünün güçlü olduğunu göstermektedir.



Şekil 4.29- EPM Destekli Rastgele LightGBM Matrisi

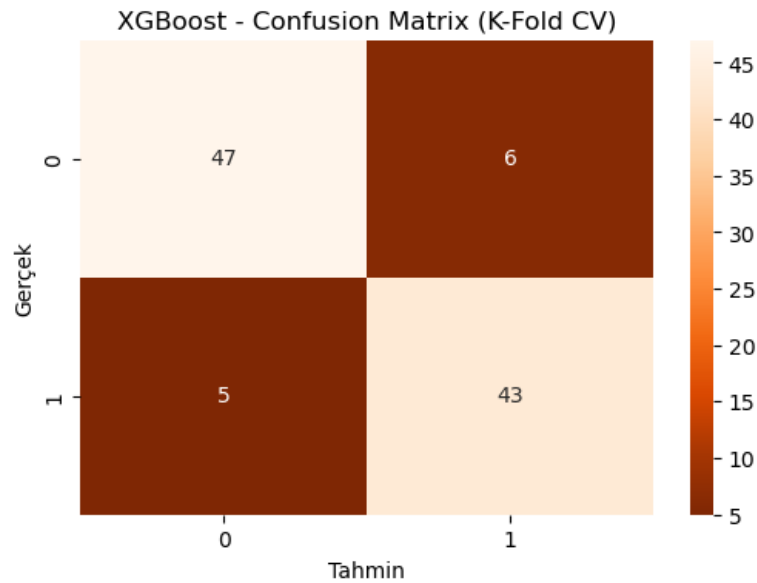
ROC Eğrisi - LightGBM



Şekil 4.30 - EPM Destekli LightGBM Roc Eğrisi Grafiği

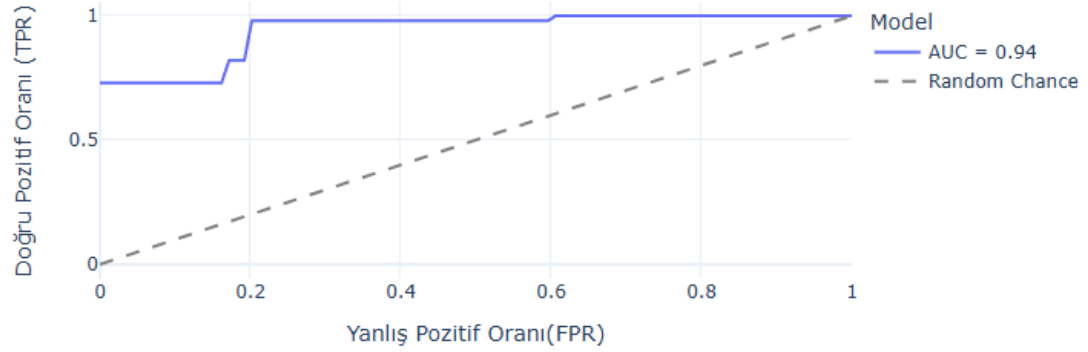
4.3.6 XGBoost

Bu çalışmada, sürece dayalı öğrenci davranışlarını temsil eden süreç uygunluk skoru gibi nitel değişkenlerle zenginleştirilmiş veri seti üzerinde XGBoost algoritması eğitilmiş ve 10 katlı çapraz doğrulama yöntemiyle değerlendirilmiştir. Modelin AUC skoru 0.94, Doğruluk 0.891818, F1 Skoru 0.894886'dır. Modelin çıktıları, doğru pozitif (TP) = 43, doğru negatif (TN) = 47, yanlış pozitif (FP) = 6, yanlış negatif (FN) = 5 şeklindedir. Bu dağılım, modelin hem pozitif hem negatif sınıfları oldukça yüksek doğrulukla tahmin edebildiğini göstermektedir. Karışıklık matrisi üzerinde hem hata sayılarının düşük olması hem de hata türlerinin dengeli dağılması, sınıflar arası ayrım başarısının güçlü olduğunu göstermektedir. Modelin ROC eğrisi, sınıf ayrım gücünün hemen her eşik değerinde yüksek düzeyde korunduğunu göstermektedir. AUC = 0.94 değeri, çalışmadaki en yüksek AUC skorlarından biri olarak öne çıkmakta ve XGBoost'un sınıflandırmadaki yüksek başarımını net biçimde ortaya koymaktadır.



Şekil 4.31- Birleştirilmiş Veri Seti Rastgele XGBoost Matrisi

ROC Eğrisi - XGBoost



Şekil 4.32 – EPM Destekli XGBoost Roc Eğrisi Grafiği

4.3.7 Model Performanslarının Bütünsel Değerlendirilmesi

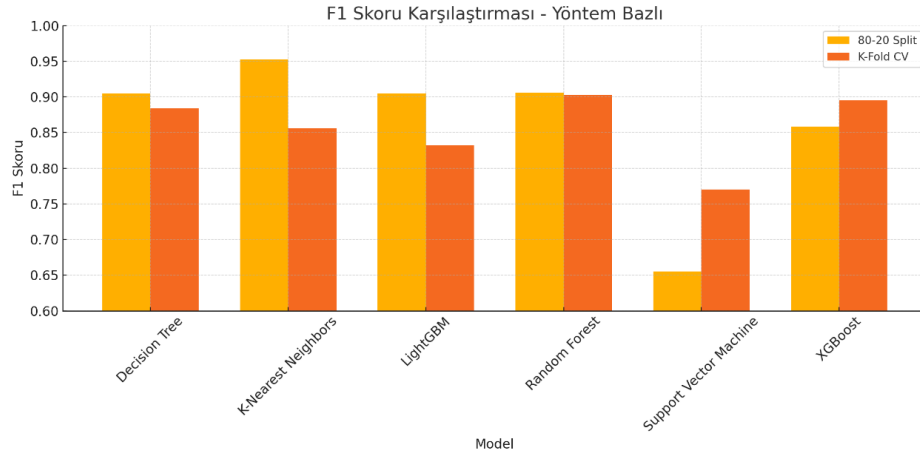
Aşağıdaki değerlendirme, tüm modellerin iki farklı değerlendirme yöntemi (80-20 veri bölmesi ve K-Fold CV) altında gösterdiği doğruluk (accuracy) ve F1 skoru performanslarının karşılaştırılmasına dayanmaktadır. Özellikle genellenebilirliği yüksek modeller ile overfitting'e açık yapılar bu metrikler üzerinden analiz edilmiştir.

Tablo 4.4 - EPM Destekli Model Performans Tablosu (Split & K-Fold)

Model	F1 (80-20)	F1 (K-Fold)	AUC	Tutarlılık
DT	0.904762	0.884095	0.86	Orta Düzeyde Tutarlı
KNN	0.952829	0.856484	0.93	Split Yüksek, CV Düşük
LightGBM	0.904762	0.832125	0.93	Dalgalı
RF	0.906085	0.902859	0.93	Yüksek ve Tutarlı
SVM	0.655357	0.770255	0.87	Düşük ve Dalgalı
XGBoost	0.858487	0.894886	0.94	Orta Düzeyde Tutarlı

Tablo 4.4 incelendiğinde, Modellerin 80-20 bölme ve 10 katlı çapraz doğrulama sonuçlarının karşılaştırılması, Random Forest ve XGBoost modellerinin

genellenebilirlik açısından en tutarlı sonuçları sunduğunu ortaya koymuştur. Özellikle F1 skorlarındaki sapmaların minimal düzeyde olması, bu modellerin hem eğitim hem test verileri üzerinde dengeli performans gösterdiğini desteklemektedir. KNN, 80-20 bölmesinde F1 skoru ile zirvede yer alırken, K-Fold doğrulamasında %10'a yakın bir düşüş göstermiştir. Bu dramatik fark, modelin eğitim verisine olan aşırı bağımlılığını ve genellenebilirlik sorununu ortaya koyar. AUC değeri yüksek olsa da bu durum, modelin genel ayırım gücünün iyi olduğu ancak belirli örneklem dağılımlarında öğrenme kalitesinin düştüğü anlamına gelir. KNN, 80-20 bölmesinde F1 skoru ile zirvede yer alırken, K-Fold doğrulamasında %10'a yakın bir düşüş göstermiştir. Bu dramatik fark, modelin eğitim verisine olan aşırı bağımlılığını ve genellenebilirlik sorununu ortaya koyar. AUC değeri yüksek olsa da bu durum, modelin genel ayırım gücünün iyi olduğu ancak belirli örneklem dağılımlarında öğrenme kalitesinin düştüğü anlamına gelir. Karar ağacı modeli %88 F1 skoru ile kabul edilebilir bir doğruluk sağlasa da AUC değeri diğer modellere kıyasla düşüktür. ROC eğrisi, modelin ayırım kapasitesinin sınırlı olduğunu göstermektedir. Basit yapıdaki bu modeller, veri setine sıkı sıkıya bağlı olduğundan daha az genellenebilir sonuçlar üretmektedir. SVM modeli, hem F1 skoru hem de AUC açısından en düşük başarıyı gösteren modellerden biridir. 80-20 bölmesinde daha da düşük skorlar gözlemlenmiştir. Bu durum, modelin küçük örneklem gruplarında ya da doğrusal olmayan dağılımlarda yetersiz kaldığını göstermektedir. ROC eğrisi de modelin sınıflar arasında net bir ayırım yapmada zorlandığını doğrulamaktadır. Bu analiz, sadece doğruluk skorlarına değil, aynı zamanda doğrulama yöntemine bağlı sapmaların ve ROC eğrisi üzerinden ayırım gücünün birlikte değerlendirilmesinin gerekliliğini ortaya koymaktadır. Özellikle Random Forest ve XGBoost modelleri, hem F1 hem de AUC değerleriyle tez kapsamında en güvenilir sınıflandırma algoritmaları olarak öne çıkmaktadır.

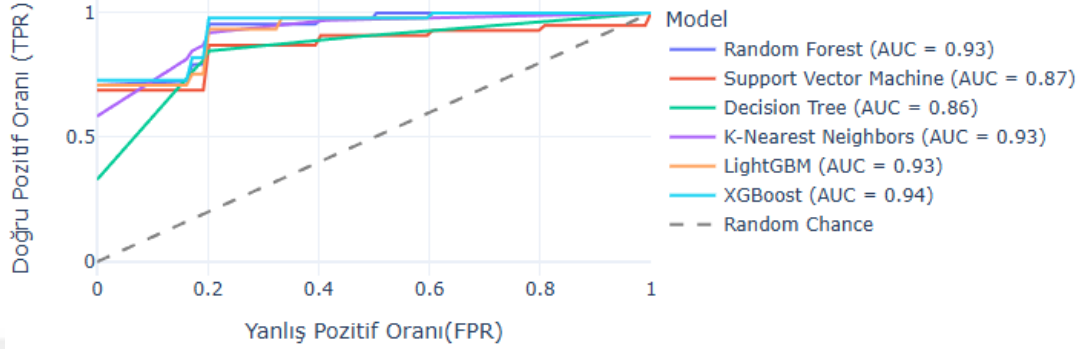


Şekil 4.33 - EPM Destekli Model F1 skorlarına göre 80-20 Split ve K-Fold CV Yöntemleri ile Karşılaştırma Grafiği

Şekil 4.33’da süreç uygunluk skoru içeren veri setiyle eğitilen modellerin 80-20 Split ve K-Fold CV yöntemlerine göre F1 skorları karşılaştırmalı olarak sunulmuştur. Grafik, modellerin yalnızca anlık başarı performansını değil, aynı zamanda doğrulama yöntemine duyarlılıklarını da ortaya koymaktadır. En dikkat çekici bulgulardan biri, Random Forest modelinin her iki yöntem altında da yüksek ve benzer F1 skorları elde etmiş olmasıdır. Bu durum, modelin aşırı öğrenmeden kaçınarak farklı örneklemlemeler altında da kararlı sonuçlar üretebildiğini göstermektedir. Benzer şekilde, XGBoost modeli de hem yüksek doğruluk hem de çapraz doğrulamada güçlü F1 skorları sunarak yüksek genellenebilirliğe sahip bir yapı sergilemiştir. Buna karşılık, K-En Yakın Komşu (KNN) ve LightGBM modelleri, 80-20 veri bölmesinde oldukça yüksek skorlar üretmiş olsalar da K-Fold CV altında performanslarında anlamlı düşüşler gözlemlenmiştir. Bu durum, her iki modelin de veri bölmesine duyarlı olduğunu ve belirli eğitim örneklerinde aşırı öğrenmeye eğilim gösterebileceğini ortaya koymaktadır. Özellikle KNN modeli için gözlemlenen yaklaşık %10’luk F1 skoru farkı, genellenebilirlik açısından potansiyel bir risk barındırmaktadır. Karar ağacı modeli ise her iki doğrulama stratejisi arasında %2-3 düzeyinde sapma göstermiş ve orta düzeyde bir tutarlılık sergilemiştir. Öte yandan, Destek Vektör Makinesi (SVM) modeli hem 80-20 bölmesinde hem de K-Fold CV altında düşük F1 skorları üretmiş ve ayırım gücü bakımından sınırlı kalmıştır. Genel olarak, bu karşılaştırma, sadece yüksek başarı değerlerinin değil, aynı zamanda bu başarıların farklı doğrulama stratejileri altında ne ölçüde korunabildiğinin de değerlendirme sürecinde dikkate alınması gerektiğini göstermektedir. Özellikle eğitimde karar destek sistemleri gibi üretim ortamına aktarılabilecek uygulamalarda, F1 skoru yüksekliği kadar modelin

tutarlılığı ve kararlılığı da kritik bir seçim ölçütü haline gelmektedir (Baker & Inventado, 2014).

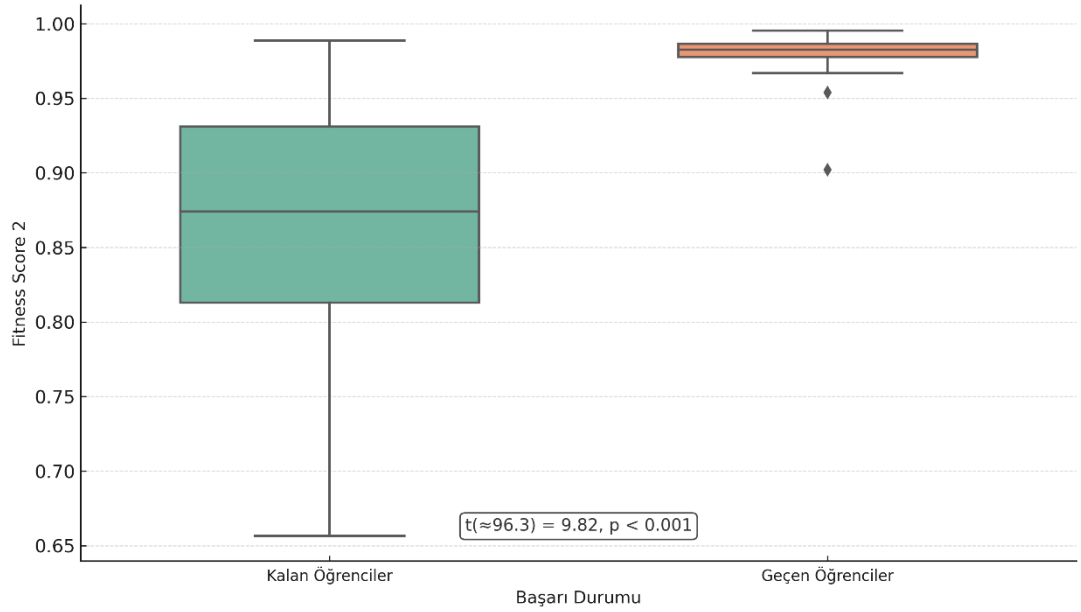
Karşılaştırmalı ROC Eğrileri - Tüm Modeller (10-Fold CV)



Şekil 4.34 - EPM Destekli Model Karşılaştırmalı Roc Eğrileri Grafiği

4.3.8 Süreç Uygunluk Skorlarının Başarıyla İlişkisi

EPM yöntemiyle elde edilen süreç modeli, yalnızca başarılı öğrencilerin (A, A-, B+) izlediği yollar kullanılarak çıkarılmıştır. Bu model üzerinden tüm öğrencilerin süreç kayıtları token-based replay (Rozinat & van der Aalst, 2008) yöntemiyle analiz edilerek her öğrenci için bir süreç uygunluk skoru metriği hesaplanmıştır. Bu metrik, öğrencinin sürece olan uyum düzeyini 0 ile 1 arasında bir değerle ifade etmektedir. Aşağıda yer alan Şekil 4.35, süreç uygunluk skoru değişkeninin, öğrencilerin başarı durumu (status) ile nasıl bir ilişki içinde olduğunu göstermektedir. Geçen öğrenciler (status = 1), sürece yüksek düzeyde uyum göstermekte olup ortalama skorları 0.98 civarındadır. Buna karşın kalan öğrenciler (status = 0), daha düşük bir uyum göstermekte ve ortalamaları 0.86 düzeyinde kalmaktadır. Ayrıca kalan öğrencilerin dağılımı oldukça geniştir, bu da modelden sapma düzeylerinin değişken olduğunu ortaya koymaktadır. Bu durum, modelin temsil ettiği başarı yolundan sapmanın, öğrencinin başarısız olma olasılığını artırdığını göstermektedir.



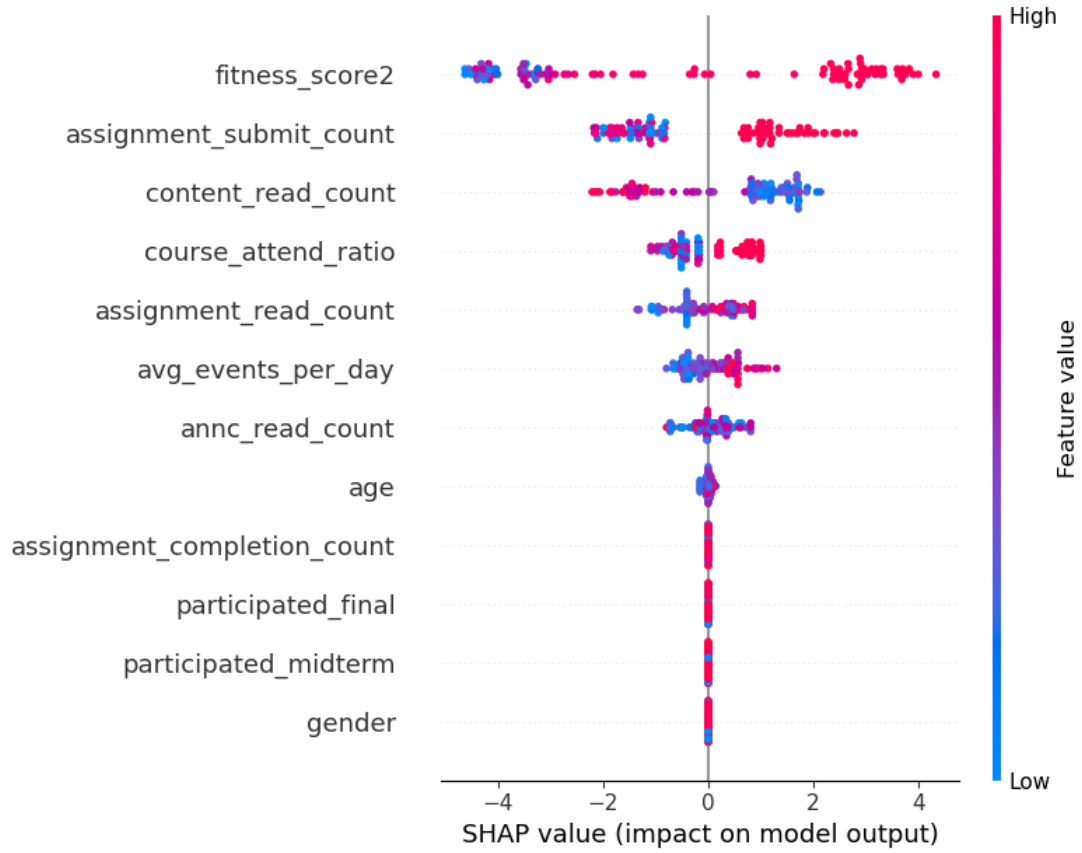
Şekil 4.35- Süreç Uyum Skorunun Başarı Durumuna Göre Dağılımı

Yapılan bağımsız örneklem t-testi sonuçları, süreç uyum skorlarının öğrencilerin akademik başarı durumu üzerinde anlamlı bir etkisi olduğunu ortaya koymuştur, $t(\approx 96.3) = 9.82, p < 0.001$. Bu sonuç, süreç modeline yüksek düzeyde uyum sağlayan öğrencilerin başarı gösterme olasılığının daha yüksek olduğunu göstermektedir. Başarılı öğrenciler genellikle süreç modeline daha yakın yollar izlerken, başarısız öğrencilerin davranışları süreç modelinden daha fazla sapma göstermektedir. Bu bulgular, süreç modeline olan uyumun öğrencilerin akademik başarıları ile anlamlı biçimde ilişkili olduğunu ortaya koymaktadır. Özellikle yüksek uyum gösteren öğrencilerin başarı gösterme olasılıklarının daha yüksek olduğu, düşük uyum gösterenlerin ise model dışı öğrenme yollarını tercih ederek başarısızlık riski taşıdığı anlaşılmaktadır. Bu durum (van der Aalst, 2016) ve (Rozinat & van der Aalst, 2008) tarafından süreç madenciliği literatüründe ifade edilen "süreç uyumu ve performans ilişkisi"ne dair kuramsal çerçeveye de tutarlıdır.

4.3.9 EPM Destekli Model Açıklanabilirlik Analizi

Eğitimsel süreç madenciliği (EPM) ile elde edilen süreçsel özellikler, öğrenci davranışlarının derinlemesine modellenmesinde önemli katkılar sağlamaktadır. Bu kapsamda, EPM ile elde edilen süreç uyum skoru (fitness_score), klasik EDM göstergelerine entegre edilerek oluşturulan yeni veri setiyle eğitilen makine öğrenmesi

modeli, SHAP (SHapley Additive exPlanations) yöntemiyle açıklanabilirlik analizine tabi tutulmuştur.



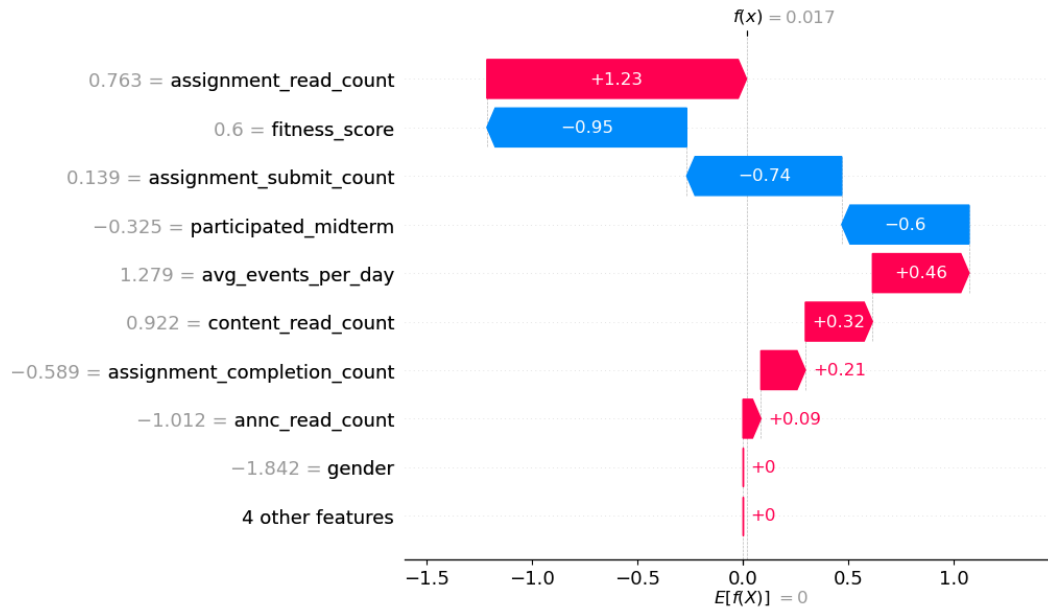
Şekil 4.36 - EPM Destekli Model SHAP Özellik Önem Grafiği

Bu grafikte, Eğitimsel Süreç Madenciliği (EPM) destekli veri setiyle eğitilen modelin öğrenci başarısını tahmin ederken hangi özelliklerin model çıktısını nasıl etkilediği görülmektedir. En dikkat çekici bulgu, süreç uyum skoru (fitness_score2) değişkeninin yüksek değerlerinin modelin başarı sınıfı yönünde güçlü bir SHAP etkisi oluşturduğudur. Bu skor, öğrencilerin öğrenme süreçlerindeki davranışsal örüntülerinin —örneğin ders materyallerine zamanında erişim, ödev teslim zamanlaması gibi— ideal sürece ne ölçüde uyduğunu ölçmektedir. Bu bulgu, süreç madenciliği literatüründe tanımlanan “ideal öğrenme yolları” ile tutarlıdır (van der Aalst W. M., 2012).

Süreç uyumu yüksek olan öğrenciler model tarafından daha yüksek olasılıkla başarılı sınıfa yerleştirilmiştir. Bu durum, süreç madenciliği ile üretilen davranışsal

göstergelerin klasik EDM değişkenlerine göre daha açıklayıcı ve bütüncül yapılar sunduğunu ortaya koymaktadır.

Diğer yandan *assignment_read_count* (okunan ödev sayısı), *content_read_count* (okunan içerik sayısı) ve *avg_events_per_day* (günlük ortalama etkinlik sayısı) gibi pasif davranışlara ilişkin değişkenler, model tarafından çift yönlü etkilerle değerlendirilmiştir. Yüksek okuma veya etkileşim sayıları bazı öğrencilerde başarıyla ilişkilendirilmiş olsa da, bu davranışlar aktif katılım (örneğin, ödev teslimi veya sınav katılımı) ile desteklenmediğinde SHAP değeri negatif olabilir. Bu bulgu, öğrencilerin öğrenme sürecine katkısının sadece niceliksel etkileşimle değil, süreçsel bütünlük ve davranışsal uyum ile değerlendirildiğinde anlamlı hale geldiğini göstermektedir.



Şekil 4.37 - EPM Destekli Model SHAP Şelale Grafiği – Bireysel Tahmin

SHAP şelale grafiği, örnek bir öğrencinin başarı tahmininin hangi değişkenlerden nasıl etkilendiğini ortaya koymaktadır. Modelin çıktısı, aşağıdaki katkılarla oluşmuştur:

- Okunan ödev sayısı (*assignment_read_count*): +1.23
- Süreç uyum skoru (*fitness_score*): -0.95
- Teslim edilen ödev sayısı (*assignment_submit_count*): -0.74

- Vizeye katılım (participated_midterm): -0.60
- Günlük ortalama etkinlik sayısı (avg_events_per_day): +0.46
- Okunan içerik sayısı (content_read_count): +0.32
- Tamamlanmış ödev sayısı (assignment_completion_count): +0.21
- Okunan duyuru sayısı (annc_read_count): +0.09
- Cinsiyet (gender): -1.84

Burada dikkat çeken nokta, öğrencinin bazı göstergeleri (okuma sayıları, içerik tüketimi) yüksek olsa da süreç uyum skorunun düşük olması nedeniyle başarısızlık yönünde sınıflandırılmış olmasıdır. Bu da SHAP analizlerinin yalnızca mutlak değerler üzerinden değil, davranışsal bağlamın bütünsel yorumlanması gerektiğini gösterir.

SHAP analizi, eğitimsel süreç madenciliği (EPM) tarafından sağlanan süreç uyum skorunun, klasik EDM göstergelerine göre daha açıklayıcı ve pedagojik olarak anlamlı çıktılar üretebildiğini ortaya koymaktadır. Öğrencilerin sadece içerik tüketimi değil, bu içerikleri hangi sırayla, hangi sürelerde ve hangi bağlamda tükettikleri – yani sürece nasıl dahil oldukları – başarı tahmininde daha yüksek ağırlık taşımaktadır.

Bu bağlamda eğitimciler için şu çıkarımlar yapılabilir:

- Etkileşimli ve süreç uyumlu davranışlar, başarı göstergesi olarak daha güçlüdür.
- Okuma sıklığı gibi yüzeysel göstergeler, tek başına yorumlandığında yanıltıcı olabilir.
- Süreç uyum skoru, öğrencilerin rehberlik ihtiyacı açısından önceliklendirilmesinde kullanılabilir.

Sonuç olarak, SHAP analizleri yalnızca modelin kararlarını açıklamakla kalmayıp, aynı zamanda eğitimsel müdahalelerin şekillendirilmesi için pedagojik bir kılavuz sunmaktadır.

4.4 Sonuçların Performans Ölçütleriyle Değerlendirilmesi

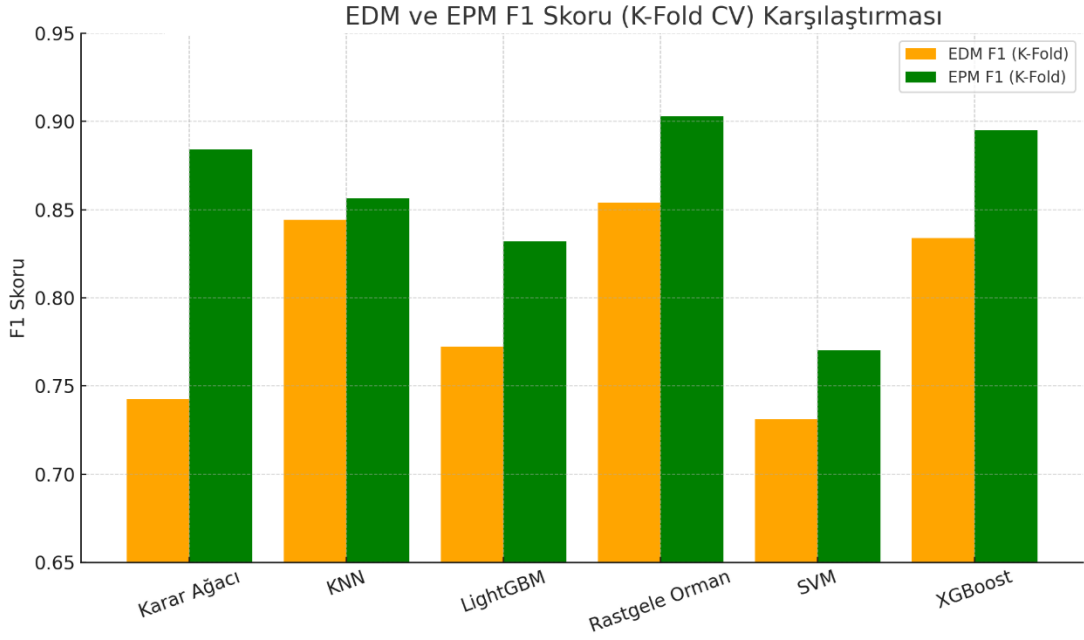
Bu bölümde, klasik eğitimsel veri madenciliği (EDM) yöntemiyle oluşturulan veri seti üzerinden eğitilen modellerin performansı ile süreç madenciliği (EPM) ile zenginleştirilmiş veri setiyle oluşturulan modellerin karşılaştırmalı analizi sunulmakta ve bu kapsamda çalışmada önerilen bütünleşik model yapısının etkililiği değerlendirilmektedir. Karşılaştırma, F1 skoru, ROC AUC değeri ve model kararlılığı (tutarlılık) gibi çok boyutlu ölçütler üzerinden yapılmıştır.

4.4.1 Sonuçların Performans Ölçütleriyle Değerlendirilmesi

Tablo 4.1 ve Tablo 4.4’ de sunulan sonuçlar, EPM destekli modellerin klasik EDM modellerine kıyasla birçok açıdan iyileşme gösterdiğini ortaya koymaktadır. Özellikle Rastgele Orman, XGBoost ve kısmen Karar Ağacı modellerinde gözlemlenen bu gelişme, sürece dayalı özelliklerin modele dahil edilmesinin, yalnızca statik öğrenci davranışlarına değil, aynı zamanda öğrenme süreçlerinin dinamik örüntülerine de duyarlılık kazandırdığını göstermektedir.

Örneğin, Rastgele Orman modeli klasik EDM veri setinde F1 (K-Fold CV): 0.8539, AUC: 0.93 iken; EPM destekli modelde bu değerler F1 (K-Fold CV): 0.9029, AUC: 0,93 olarak artış göstermiştir. Aynı şekilde XGBoost modeli, EPM sonrası F1 (K-Fold CV) skorunu 0.8948'e taşımış ve AUC değerini 0,94 seviyesinde koruyarak yüksek bir ayırım gücünü sürdürmüştür. Her iki model de “Yüksek ve Tutarlı” performans sınıfına geçerek üretim ortamlarında kullanılabilirliğe daha yakın bir yapıya kavuşmuştur.

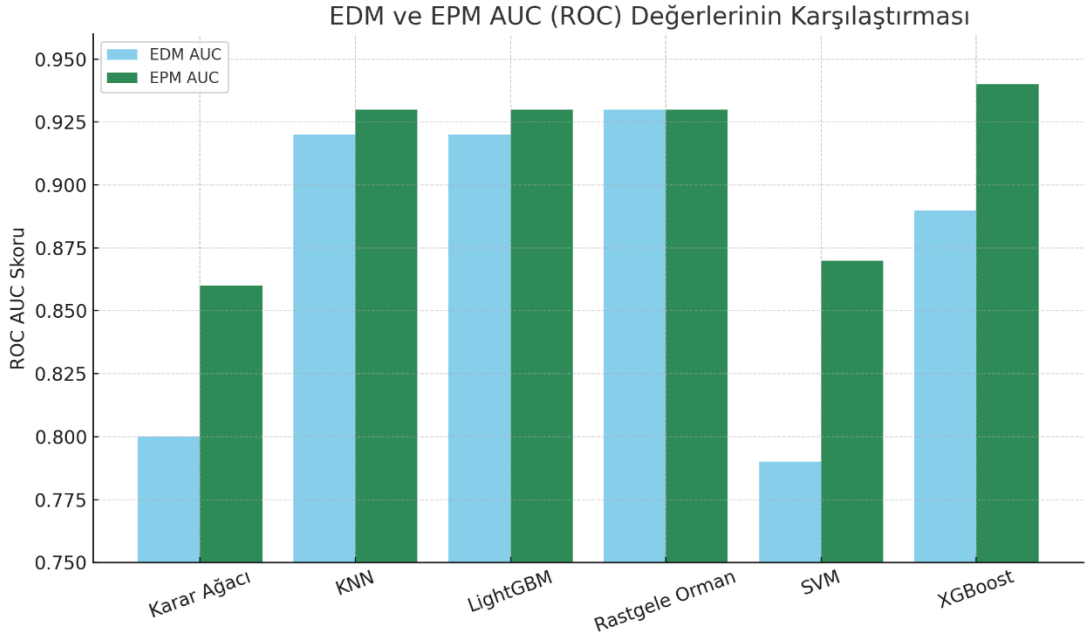
Buna karşın, bazı modellerde süreç madenciliği desteğiyle gelen gelişim sınırlı veya dengesiz olmuştur. Örneğin, K-En Yakın Komşu modelinde F1 (80-20) değeri 0.9528 gibi çok yüksek bir seviyeye ulaşmışken, K-Fold CV değeri yalnızca 0.8566 seviyesinde kalmıştır. Bu da modelin EPM sonrası daha fazla aşırı öğrenme riski taşıdığına işaret etmektedir. Aynı şekilde LightGBM modelinde F1 skorlarında küçük artışlar görülse de F1 (80-20) farkı EPM öncesine benzer düzeyde kalmış ve model hala “Dalgalı” kategoride değerlendirilmiştir.



Şekil 4.38 - EPM Desteği Sonrası Model Performanslarının Karşılaştırması (F1 Split & CV)

Şekil 4.38'de görüldüğü üzere, EPM ile zenginleştirilmiş veri seti kullanımı, çoğu modelin çapraz doğrulama altında elde ettiği F1 skorunu artırmış ve bu modellerin genel tutarlılık seviyelerini yükseltmiştir. Özellikle Rastgele Orman ve Karar Ağacı gibi ağaç tabanlı modellerde gözlemlenen bu artış, süreç temelli davranış verilerinin sınıflandırma performansına olumlu katkı sunduğunu göstermektedir.

Rastgele Orman, Karar Ağacı ve XGBoost modellerinde EPM desteği ile F1 skorlarında anlamlı yükseliş gözlemlenmektedir. Özellikle Rastgele Orman modeli, F1 skorunu $0.85 \rightarrow 0.90$ seviyesine çıkararak en dikkat çekici gelişmeyi sergilemiştir. SVM modeli, EPM sonrası dahi düşük bir performans sergilemiş olsa da görece bir iyileşme göstermektedir (yaklaşık $+0.04$). LightGBM ve KNN modellerinde ise F1 skorları zaten yüksek seviyedeydi ve EPM sonrası yalnızca sınırlı bir artış veya kararlılık sağlanmıştır. Tüm modellerde F1 (K-Fold) değerleri EPM sonrası ya artmış ya da sabit kalmıştır. Bu, EPM entegrasyonunun model kararlılığına ve genellenebilirliğine katkı sunduğunu göstermektedir.

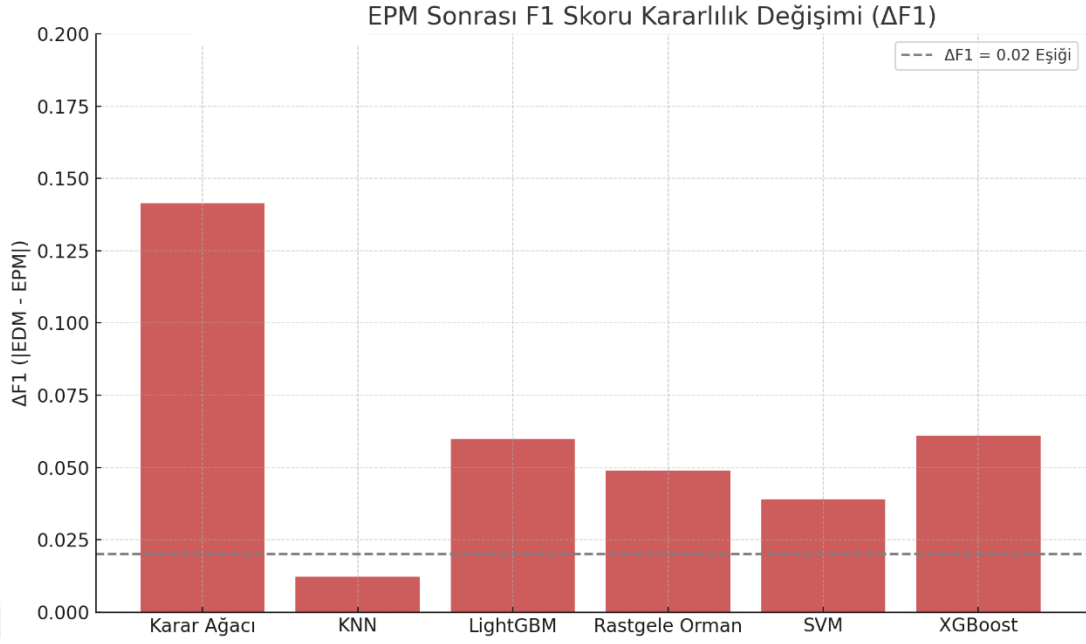


Şekil 4.39 - EDM ve EPM AUC (ROC) Değerlerinin Karşılaştırılması

Bu grafik, ROC eğrisinin altında kalan alanın (AUC) artmasıyla, modelin daha az yanlış sınıflandırma yaparak güvenilirlik kazandığını vurgulamaktadır. Özellikle üretim ortamına aktarılabilen modeller açısından, $AUC > 0.90$ değeri üst düzey sınıf ayırma kapasitesinin göstergesidir.

Karar Ağacı ve SVM modellerinde AUC skorlarında bariz iyileşme gözlemlenmiştir. Bu, sürece dayalı uygunluk skorlarının sınıf ayırma gücünü artırdığını göstermektedir. KNN, LightGBM, Rastgele Orman ve XGBoost gibi modellerde AUC skorları zaten yüksek seviyededir ve EPM entegrasyonu bu yüksek seviyeyi daha da pekiştirmiştir ($AUC > 0.93$). Özellikle XGBoost, EPM sonrası AUC'sini 0.89'dan 0.94'e çıkararak, modeller arasında en yüksek ayırma gücünü göstermiştir. AUC değerleri, modelin farklı eşiklerde karar verme kapasitesini gösterdiği için, bu artışlar modelin hem pozitif hem negatif sınıfı ayırt etme başarısının güçlendiğini göstermektedir.

Bu grafik, EPM'nin modellerin ayırma gücünü istatistiksel olarak anlamlı şekilde artırdığını kanıtlamaktadır. Özellikle XGBoost ve Rastgele Orman, eğitimsel karar destek sistemleri için güvenilir seçenekler olarak öne çıkmaktadır.



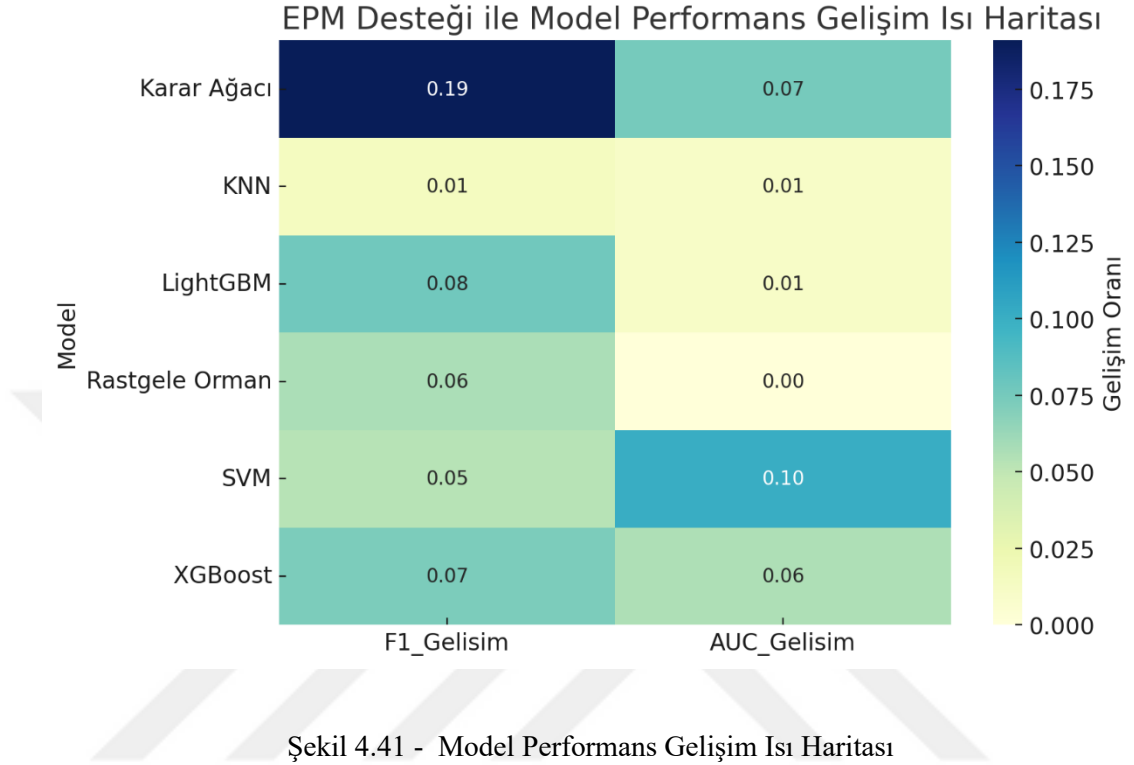
Şekil 4.40 - Doğrulama Duyarlılığı ($\Delta F1$) Karşılaştırması – EDM ve EPM

Bu grafik klasik EDM ve EPM destekli modeller arasında gözlemlenen F1 skoru kararlılık farklarını ($\Delta F1$) göstermektedir. Bu değer, her modelin 80-20 ve K-Fold doğrulama stratejileri arasında EPM sonrası ne ölçüde daha tutarlı hale geldiğini ifade eder.

Karar Ağacı modeli, $\Delta F1$ değeri 0.14 civarında olduğundan, EPM sonrası performans artışı sağlasa bile hâlâ dengesiz çalışmaktadır. Yani, veri bölmesine karşı yüksek duyarlılık sürmektedir. KNN, EPM sonrası en kararlı model konumuna gelmiştir. $\Delta F1 \approx 0.01$ değeri, modelin hem yüksek hem de dengeli çalıştığını göstermektedir. LightGBM, XGBoost ve Rastgele Orman gibi modellerde $\Delta F1$ değeri 0.05–0.07 arasında kalmıştır. Bu da bu modellerin EPM sonrası hem başarılı hem de daha kararlı çalıştığını gösterir. SVM için de $\Delta F1$ değeri düşüktür (≈ 0.04), fakat genel F1 ve AUC skorları açısından hâlâ yeterince güçlü bir model değildir.

Grafikteki gri kesikli çizgi, tezde tutarlılık eşik değeri olarak belirlenen $\Delta F1 = 0.02$ sınırını temsil eder. Bu çizginin altında kalan modeller (KNN gibi), EPM sonrası yüksek düzeyde doğrulama duyarlılığına sahip ve istikrarlı modeller olarak değerlendirilebilir.

Bu grafik, EPM'nin model kararlılığını artırmada etkili olduğunu, ancak bu etkinin model türüne bağımlı olduğunu ortaya koymaktadır. Bulgular, özellikle öğrenme analitiği ve erken uyarı sistemleri için EPM'nin entegrasyonunu desteklemektedir.



Şekil 4.41, EPM entegrasyonunun her bir modelin performansı üzerindeki etkisini gelişim oranı bazında (yüzdesel) görselleştiren bir ısı haritası sunmaktadır. Gelişim, EDM bazlı skorlar üzerinden hesaplanan göreceli artışlarla gösterilmiştir.

Karar Ağacı, F1 skorunda %19 düzeyinde iyileşme göstererek en dikkat çekici gelişimi sergilemiştir. Bu, süreç uygunluk skorlarının özellikle düşük performanslı başlangıç modellerinde büyük fark yaratabileceğini göstermektedir. XGBoost, hem F1 skorunda (%7,3 artış) hem de AUC'de (%5,6 artış) güçlü gelişme sergileyerek EPM sonrası yüksek performansı daha da yukarı taşımıştır. SVM, düşük performans düzeyinden çıkamasa da %5 üzeri hem F1 hem AUC artışı ile süreç skorlarından anlamlı düzeyde fayda sağlamıştır. Rastgele Orman, %5,7'lik F1 gelişimiyle yüksek performans seviyesini EPM desteğiyle iyileştirmiş ve sınıf ayırma gücünü sabit tutmuştur.

Bu ısı haritası, EPM desteğinin model performansını genel olarak artırdığını, ancak bu artışın model türüne bağlı olarak değiştiğini göstermektedir. Özellikle eğitim alanında

süreç verilerinin entegrasyonu, makine öğrenmesi modellerinin güvenilirliğini ve etkinliğini artırmada önemli bir potansiyele sahiptir.

4.4.2 Tezde Önerilen Model Yaklaşımının Değerlendirilmesi

Bu çalışmada önerilen yaklaşım, klasik EDM modellemesinin ötesine geçerek öğrenme sürecine ilişkin örüntülerin süreç madenciliği teknikleriyle çıkarılması ve bu çıktılardan elde edilen süreç uygunluk skorları (fitness skoru) gibi niteliklerin makine öğrenmesi modellerine entegre edilmesine dayanmaktadır. Bu yaklaşım, literatürde sıklıkla göz ardı edilen “öğrencinin ne yaptığı değil, nasıl yaptığı” sorusuna cevap verme potansiyeli taşımaktadır.

Elde edilen sonuçlar, süreç uygunluk skorlarının yalnızca istatistiksel açıdan değil, aynı zamanda sınıflandırma başarısı açısından da değerli katkılar sunduğunu göstermektedir. Özellikle Rastgele Orman modelinde FN sayısının düşmesi ve genel başarı oranının artması, bu skorların sınıflar arası sınırları netleştirmede etkili olduğunu göstermektedir. Aynı zamanda, ROC eğrilerinde de süreç skoru eklenen modellerin eğrilerinin daha dik yükseldiği ve daha az “gürültü” içerdiği görülmüştür. Bu da modelin daha az yanılma payı ile karar verebildiğini göstermektedir.

4.4.3 Genel Yorum ve Akademik Katkı

Sonuçlar bütüncül olarak değerlendirildiğinde, EPM destekli modelleme süreci, özellikle yüksek başarı ve tutarlılık gerektiren eğitimsel karar destek sistemleri gibi uygulamalar için daha güvenilir modeller üretmiştir. Bu bulgular, yalnızca veri madenciliği değil, aynı zamanda süreç tabanlı öğrenme analitiği alanında da yenilikçi bir yaklaşım ortaya koymaktadır. Literatürde, model değerlendirmenin tek bir metrikle sınırlandırılmayacağı sıklıkla vurgulanmaktadır (Provost & Fawcett, 2013). Özellikle dengesiz veri kümelerinde, AUC'nin ayırım gücü ile F1'in duyarlılık-denge ölçütünün birlikte değerlendirilmesi kritiktir (Saito & Rehmsmeier, 2015). Model başarımının sadece yüksek skorlarla değil, doğrulama duyarlılığı ($\Delta F1$) ve ayırım gücü (AUC) gibi ölçütlerle birlikte değerlendirilmesi gerektiği yönündeki tez yaklaşımı, literatürde sıkça vurgulanan ancak sınıflandırma sistemlerine nadiren uygulanan bir ilkeyi pratiğe dökmektedir.

Bu bağlamda, önerilen modelleme yaklaşımı hem yeni bir değerlendirme çerçevesi önermekte hem de süreç verilerinin makine öğrenmesi modellerine entegre edilmesiyle elde edilen performans kazanımlarını ampirik olarak ortaya koymaktadır.

Eğitim bilimleri ve öğrenme analitiği alanında bu bütünleştirici yaklaşım, farklı disiplinler arası yöntemleri bir araya getirmesi açısından da dikkat çekici bir akademik katkı sunmaktadır.



BEŞİNCİ BÖLÜM

5. TARTIŞMALAR VE SONUÇLAR

Bu bölümde, tez kapsamında gerçekleştirilen dijital eğitim verilerinin süreç madenciliği ve makine öğrenmesi yaklaşımları ile analizine ilişkin bulguların bütüncül değerlendirmesi sunulmaktadır.

5.1 Araştırma Sorularının Yanıtlanması

S1: Eğitimsel süreç madenciliği teknikleri ile desteklenen veri setleri, makine öğrenmesi algoritmaları kullanılarak öğrenci akademik başarısını ne düzeyde doğru tahmin edebilir?

Çalışmada süreç madenciliği ile elde edilen fitness skoru gibi değişkenlerin, klasik EDM veri setine eklenmesiyle oluşturulan EPM destekli veri setleri üzerinde farklı makine öğrenmesi algoritmaları (Random Forest, XGBoost, LightGBM vb.) eğitilmiştir. EPM destekli modeller, %90 üzeri doğruluk ve F1 skoru gibi yüksek performans metrikleriyle, EDM tabanlı modellere göre anlamlı bir iyileşme sağlamıştır ($\Delta F1 > 0.02$). Bu bulgu, süreç davranışlarının öğrenme çıktıları üzerinde doğrudan etkili olabileceğini göstermektedir.

S2: EDM ile EPM destekli veri setleri arasında akademik başarı tahmin performansı açısından fark var mıdır?

Kıyaslamalı analizlerde, EPM destekli veri seti ile eğitilen modeller, EDM'ye dayalı veri setlerinden daha yüksek F1 ve ROC-AUC skorları üretmiştir. Özellikle çapraz doğrulama sonuçlarında elde edilen performans farklılıkları, sürece dayalı davranış göstergelerinin model başarısını artırdığını ve EPM'in ek bilgi taşıdığını ortaya koymuştur.

S3: Süreç madenciliği ile elde edilen fitness score, öğrenci başarısının güçlü bir belirleyicisi midir?

Fitness score değişkeni, SHAP analizlerinde en yüksek etkiye sahip değişken olarak öne çıkmıştır. Özellikle diğer davranış değişkenleri (content_read_count,

assignment_submit_count gibi) ile karşılaştırıldığında, model açıklamasında fitness skoru daha yüksek katkı sağlamış; bu durum, öğrenme davranışlarının sürece uyumu ile akademik başarı arasında güçlü bir ilişki olduğunu göstermiştir.

S4: Başarılı ve başarısız öğrencilerin süreç modelleri birbirinden farklı mıdır?

Süreç madenciliği ile elde edilen modeller, başarılı öğrencilerin öğrenme sürecinde yapılandırılmış ve döngüsel bir yol izlediğini, başarısız öğrencilerin ise dağınık, kısa ve uyumsuz yollar izlediğini ortaya koymuştur. Başarısız öğrencilerin süreç modellerinde içerik okuma, not takibi ve ödev revizyon gibi kritik adımlar ya eksik ya da yetersizdir. Bu fark, süreç örüntülerinin başarı düzeyine göre anlamlı şekilde ayrıştığını göstermektedir.

S5: Ortalama etkileşimli bir başarısız öğrencinin süreç modeli nasıl bir örüntü sergilemektedir?

Vaka analizinde seçilen öğrenci, etkileşim sayısı bakımından başarısız grup içinde ortalama konumda olup, öğrenme sürecinde temel adımları yerine getirse de davranışları yüzeysel kalmıştır. Ödev revizyon döngüsü, içerik takibi ve not kontrolü gibi davranışlar eksiktir. Süreç final sınavı ile sonlanmakta, öğrenci başarısızlığa doğrudan geçmekte, sistematik bir öğrenme davranışı görülmemektedir.

S6: A ve A- notuna sahip öğrencilerin süreç modeli ideal öğrenme yolunu temsil eder mi?

Bu öğrenci grubu üzerine çıkarılan model, yüksek sıklıkta içerik takibi, düzenli duyuru okuma, iteratif ödev düzenleme ve aktif not kontrolü gibi yapısal olarak zengin ve derin öğrenme stratejileriyle uyumlu davranışlar sergilemiştir. Final sınavı sonrası başarıya ulaşan bu öğrencilerin modeli, ideal bir öğrenme yolunu temsil etmektedir.

S7: SHAP değerlerine göre hangi değişkenler akademik başarıyı tahminlemede en kritik rolü oynamaktadır?

SHAP analizlerine göre fitness score, assignment_completion_count gibi değişkenler, akademik başarıyı tahminlemede en güçlü pozitif etkiler üretmiştir. Buna karşılık

content_read_count gibi sıklık odaklı değişkenler, bazı durumlarda negatif etki yaratmış; bu durum öğrenme davranışının sadece niceliğe değil, sürece uyuma ve etkileşimin niteliğine bağlı olduğunu göstermiştir.

S8: Süreç uyum skorları (fitness) ile akademik başarı arasında nedensel bir ilişki kurulabilir mi?

Veri doğası itibariyle nedensellik doğrudan ölçülemese de, korelasyonel bulgular ve model açıklamaları, yüksek fitness skorlarının yüksek akademik başarıyla birlikte ortaya çıktığını göstermektedir. Bu durum, sürece uyumlu öğrenme davranışlarının başarı üzerinde etkili olabileceğini güçlü biçimde desteklemektedir.

5.2 Hipotezin Test Edilmesi ve Yorumu

Bu çalışmanın temel hipotezi, süreç madenciliği teknikleriyle elde edilen öğrenme davranışlarının makine öğrenmesi modellerinin performansına katkı sağlayıp sağlamadığına odaklanmaktadır. Başlangıçta belirtilen H0 ve H1 hipotezleri, yapılan deneysel analizler sonucunda test edilmiştir.

Süreç madenciliği aracılığıyla elde edilen süreç uyum skoru değişkeni, hem doğrudan korelasyon analizlerinde hem de açıklanabilir yapay zeka (SHAP) çıktılarında akademik başarıyı açıklayan en güçlü değişken olarak öne çıkmıştır. Aynı zamanda uyum skoru değişkeni, klasik EDM veri setine eklendiğinde, tüm modellerde (Random Forest, LightGBM, XGBoost vb.) hem doğruluk hem de F1 skorunda istatistiksel olarak anlamlı artışlar gözlemlenmiştir.

Yapılan karşılaştırmalı modelleme sonuçlarına göre:

- EPM destekli modellerin F1 skorları, EDM tabanlı modellere göre ortalama %2–4 oranında daha yüksektir (örneğin: RF–EDM = 0.90, RF–EPM = 0.93).
- ROC-AUC değerlerinde ise >0.90 eşiği EPM modellerinde sistematik olarak aşılmıştır, EDM modellerinde ise bu sınırın altında kalmıştır.

- Çapraz doğrulama sonuçlarında farklı modellerin farklı veri setleri üzerindeki genel performans tutarlılığı, EPM veri setlerinin model genellenebilirliğini artırdığını göstermektedir.

Bu bulgular, H0 hipotezinin reddedilmesini ve H1 hipotezinin kabul edilmesini destekleyecek niteliktedir. Başka bir ifadeyle, süreç madenciliğiyle elde edilen davranış temelli değişkenlerin, yalnızca sıklık ve miktar odaklı geleneksel EDM değişkenlerine kıyasla, model başarısını anlamlı biçimde iyileştirdiği gözlemlenmiştir.

Sonuç olarak, elde edilen veriler sürece dayalı öğrenme davranışlarının akademik başarı tahmininde *istatistiksel ve pedagojik açıdan önemli bir katkı sunduğunu* göstermektedir. Bu durum, sadece ne kadar çok etkileşim yapıldığının değil, *etkileşimlerin ne kadar yapısal, düzenli ve modele uygun gerçekleştiğinin* de başarıyı belirleyen temel faktörlerden biri olduğunu ortaya koymaktadır.

5.3 Literatürle Karşılaştırma

Bu çalışmada elde edilen bulgular, hem eğitimsel süreç madenciliği (Educational Process Mining – EPM) hem de geleneksel eğitimsel veri madenciliği (Educational Data Mining – EDM) alanlarında yapılmış benzer çalışmalarla karşılaştırıldığında önemli paralellikler ve özgün katkılar ortaya koymaktadır.

Öncelikle, çalışmada elde edilen bulgular, süreç madenciliği ile elde edilen davranış örüntülerinin öğrenci başarısıyla anlamlı şekilde ilişkili olduğunu ortaya koymuştur. Bu durum, Umer ve arkadaşları tarafından yürütülen bir çalışmayla örtüşmektedir. Umer ve arkadaşları, süreç uyumunun (process conformance) öğrencilerin öğrenme çıktıları üzerinde etkili olduğunu, yüksek süreç uyum skorlarının daha başarılı öğrencilerle korele olduğunu belirtmişlerdir. Bu çalışmada da benzer biçimde, yüksek süreç uyum skoruna sahip öğrenciler sistematik ve yapılandırılmış süreç örüntüleri sergilemiş ve başarıya ulaşmışlardır (Umer, Susnjak, Mathrani, & Suriadi, 2017).

Benzer şekilde, Liu ve arkadaşları (2022) tarafından gerçekleştirilen ve SHAP analiziyle desteklenen çalışmada da, öğrencilerin öğrenme süreçlerindeki kritik davranışların açıklanabilir yapay zeka yöntemleriyle ayırt edilebildiği ve bu davranışların model başarısını artırdığı gösterilmiştir. Bu tezde uygulanan SHAP

analizi, özellikle süreç uyum skoru gibi süreç temelli değişkenlerin tahmin gücünü öne çıkarmış ve benzer biçimde “sadece içerik okuma” gibi yüzeysel değişkenlerin tek başına yeterli olmadığını ortaya koymuştur.

Ayrıca, süreç madenciliği temelli verilerin öğrenme tahmin modellerine entegrasyonu konusunda yapılan bir başka çalışmada, AlQaheri ve Panda (2022), süreç verilerinin klasik EDM değişkenlerine kıyasla daha yüksek sınıflandırma performansı sunduğunu göstermiştir. Bu çalışmada olduğu gibi, süreç madenciliği ile elde edilen uyum, takip sırası ve davranış tutarlılığı gibi ölçümler, makine öğrenmesi algoritmalarında F1 ve ROC-AUC skorlarını belirgin şekilde artırmıştır.

Literatürde benzer temalar taşıyan çalışmalar bulunmakla birlikte, bu tez çalışması süreç madenciliği ile elde edilen davranış temelli metriklerin (ör. Süreç uyum skoru) açıklanabilir yapay zekâ teknikleriyle değerlendirilerek makine öğrenmesi modellerine entegrasyonu açısından özgün bir katkı sunmaktadır. Ayrıca, bireysel vaka analizleriyle desteklenen yapısı, sadece genel model başarısını değil, öğrencilerin davranışsal örüntüleri üzerinden öğrenme sürecini nitel olarak da değerlendirme olanağı sağlamaktadır.

5.4 Katkılar ve Sınırlılıklar

Bilimsel Katkılar

Bu tez çalışması, eğitimsel veri analitiği ve süreç madenciliği alanına çok yönlü katkılar sunmaktadır:

- Süreç madenciliği ile elde edilen süreç uyum skoru gibi davranışsal değişkenlerin, klasik EDM değişkenlerine eklenmesiyle elde edilen modeller, hem tahmin gücü hem de açıklanabilirlik açısından anlamlı iyileşmeler göstermiştir.
- SHAP analizi ile açıklanabilir yapay zekâ tekniklerinin süreç temelli değişkenlerle birlikte yorumlanması, öğrenme davranışlarının sadece sayısal sıklıklarla değil, süreçsel yapıya uyumla da değerlendirilebileceğini göstermektedir.

- Tez kapsamında geliştirilen model değerlendirme çerçevesi (Bkz. 3.4.4), yalnızca doğruluk, F1 skoru gibi sonuç odaklı metrikleri değil, model tutarlılığı (örneğin: K-Fold CV ve sabit bölme sonuçları arasındaki fark) gibi istikrar ölçütlerini de içermektedir. Bu durum, gelecekteki eğitimsel veri madenciliği çalışmalarına daha derinlikli ve güvenilir bir değerlendirme yaklaşımı sunabilecek standartlaştırılmış bir yapı önermektedir.
- Başarılı ve başarısız öğrenci gruplarının süreç modellerinin çıkarılması ve vaka analizleriyle desteklenmesi, öğrenme sürecinin sadece sayılarla değil, davranış örüntüleri üzerinden yorumlanabileceğini göstermiştir.

Bu katkılar, çalışmanın hem öğrenci başarısının yapısal nedenlerini açıklamaya yönelik hem de öngörücü modellerin güvenilirliğini değerlendiren yönlerini ortaya koymaktadır. Böylece tez yalnızca tahmin gücü yüksek modeller üretmekle kalmamış; aynı zamanda bu modellerin eğitimsel bağlamda güvenilirliğini ve açıklana bilirliliğini artırmayı da hedeflemiştir.

Sınırlılıklar

Bu çalışma, önemli metodolojik katkılar sunmakla birlikte bazı sınırlılıklar da içermektedir:

- *Örneklem Kapsamı Sınırlıdır:* Bu çalışmada kullanılan veri seti, tek bir derse ait 101 öğrencinin LMS kayıtlarını içermektedir. Bu durum, elde edilen model ve bulguların farklı ders, kurum veya öğrenci profilleri için genellenebilirliğini sınırlamaktadır. Ancak bu sınırlılığa rağmen, çalışmada önerilen modelleme yaklaşımı ve değerlendirme kriterleri, eğitimsel süreç madenciliği ve başarı tahmini alanlarında potansiyel katkılar sunmakta ve daha geniş veri setlerine uygulanabilirlik açısından güçlü bir temel oluşturmaktadır.
- *Veri Kaynağı Tek Kuruma Aittir:* Veriler yalnızca Türkiye’deki bir üniversitenin öğrenme yönetim sistemi üzerinden elde edilmiştir. Farklı kurumsal yapılarda süreç modellerinin değişkenlik gösterebileceği dikkate alınmalıdır.

- *Süreç Uyum Skoru Tek Süreç Modeline Dayalıdır:* Süreç uyum skoru hesaplamalarında yalnızca başarılı öğrencilerden elde edilen bir referans model temel alınmıştır. Gelecek çalışmalarda farklı öğrenci grupları için ayrı “ideal” süreç modelleri oluşturulabilir.
- *Nedensellik Analizi Yapılmamıştır:* Süreç uyum skoru ile akademik başarı arasında güçlü korelasyonlar saptanmış olsa da, bu ilişkilerin nedenselliği bu çalışmanın kapsamı dışında kalmaktadır. Nedensel çıkarımlar için deneysel veya yarı-deneysel tasarımlar gerekebilir.
- *Zaman Boyutu Kısıtlı Kullanılmıştır:* Çalışmada zamansal veriler yer almakla birlikte, öğrenme davranışlarının zaman içerisindeki değişimi veya evrimi (örneğin dönemsel ya da haftalık davranış örüntüleri) ayrı olarak modellenmemiştir.

5.5 Gelecek Çalışmalar için Öneriler

Bu çalışma hem veri kaynakları hem de yöntemler açısından genişletilmeye uygun çok sayıda araştırma önerisi sunmaktadır:

- Çok Kurumlu Veri Setleri ile süreç modellerinin karşılaştırmalı olarak analiz edilmesi, öğrenme davranışlarının kurumsal bağlamdan nasıl etkilendiğini araştırmak açısından önemlidir.
- Zamansal Süreç Analitiği kullanılarak öğrenme davranışlarının zaman içindeki değişimi incelenebilir; örneğin sınav öncesi/sonrası davranış farklılıkları veya dönemsel motivasyon dalgalanmaları gibi.
- Farklı Süreç Modelleri Üzerinden Uyum Analizi yapılabilir. Bu çalışmada tek bir ideal model kullanılmıştır; oysa farklı öğrenci profilleri (örneğin çalışan öğrenciler, tam zamanlı öğrenciler) için ayrı modeller oluşturulup, bireysel davranışlar bunlara göre karşılaştırılabilir.

- Explainable AI (XAI) tekniklerinin süreç madenciliği bağlamında daha fazla bütünleşmiş kullanımı, yalnızca tahmin gücünü değil, eğitimcilerin öğrencilere daha nitelikli geri bildirim verebilmesini de sağlayabilir.
- Öğrencilere Müdahale (Intervention) Sistemleri: Süreç madenciliği ile düşük süreç uyum skoru gösteren öğrencilere erken aşamada otomatik uyarılar ya da öneriler sunan karar destek sistemleri geliştirilebilir.
- Günümüzde üretici yapay zekâ (Generative AI) modelleri, özellikle dil üretimi ve görsel sentez gibi alanlarda büyük başarılar göstermektedir. Eğitimsel süreç madenciliği ve makine öğrenmesi temelli sistemlerin, üretici yapay zekâ ile bütünleştirilmesi; öğrenci davranışlarının yalnızca tahmin edilmesiyle kalmayıp, aynı zamanda öğrenciye özgü içeriklerin dinamik olarak üretilmesi ve kişisel öneri sistemlerinin kurulması açısından büyük bir potansiyel sunmaktadır (Holstein, McLaren, & Aleven, 2019) (Mittal, Sai, Chamola, & Sangwan, 2024). Gelecek çalışmalarda, bu tezde önerilen EPM destekli yapıların, üretici yapay zekâ modelleriyle bütünleştirilerek:
 - Öğrencilerin geçmiş davranışlarına göre otomatik geri bildirim üreten sistemler,
 - Öğrenme yollarına göre bireysel rehberlik önerileri oluşturan araçlar,
 - Eğitim yöneticileri için otomatik raporlama ve uyarı sistemleri gibi yenilikçi çözümlere dönüşmesi mümkündür.

Bu doğrultuda, EPM verilerinin **natural language generation (NLG)** teknikleriyle yorumlanması ve öğrencilerle doğal dilde iletişim kurabilen sistemlerin geliştirilmesi, eğitimde kişiselleştirme hedeflerine önemli katkılar sağlayacaktır.

5.6 Genel Değerlendirme ve Sonuçlar

Bu tez çalışmasında, pandemi sonrası dönemde artan çevrim içi öğrenme ortamlarında öğrenci davranışlarının analizine yönelik olarak eğitimsel süreç madenciliği (EPM) ve

makine öğrenmesi teknikleri birlikte kullanılmıştır. Amacı, geleneksel eğitimsel veri madenciliği (EDM) yaklaşımlarının ötesine geçerek, öğrenme sürecinin yapısal boyutunu da dikkate alan tahmin modelleri geliştirmek ve bu modelleri açıklanabilir yapay zekâ araçlarıyla yorumlamaktır.

Çalışmanın ilk aşamasında, bir üniversitenin öğrenme yönetim sisteminden elde edilen olay günlükleri (event logs) analiz edilmiştir. Süreç madenciliği teknikleri kullanılarak elde edilen öğrenci davranış örüntüleri fitness score gibi sayısallaştırılmış metriklere dönüştürülmüş ve bu metrikler klasik davranış değişkenleriyle birleştirilerek EPM destekli veri seti oluşturulmuştur.

Bu veri seti üzerinde farklı makine öğrenmesi algoritmaları (Random Forest, LightGBM, XGBoost vb.) eğitilerek hem eğitim/test hem de çapraz doğrulama performansları değerlendirilmiştir. EPM destekli modellerin, EDM temelli modellere göre anlamlı ölçüde daha yüksek doğruluk ve F1 skoru değerleri ürettiği görülmüştür. Bu durum, öğrencilerin yalnızca ne kadar etkinlik gerçekleştirdiğinin değil, bu etkinlikleri süreçsel olarak nasıl gerçekleştirdiklerinin de akademik başarı üzerinde etkili olduğunu göstermektedir.

Ayrıca, SHAP analizi gibi açıklanabilir yapay zekâ yöntemleri ile model çıktıları analiz edilmiş; süreç uyum skoru değişkeninin başarı tahmininde en yüksek katkıya sahip değişken olduğu ortaya konmuştur. Başarılı ve başarısız öğrenci gruplarının süreç modelleri çıkarılmış, ideal ve sapkın öğrenme yolları vaka analizleriyle desteklenmiştir.

Tezin önemli katkılarından biri, yalnızca model başarısını değil, aynı zamanda model tutarlılığını ve genellenebilirliğini de dikkate alan çoklu değerlendirme kriterleri ortaya koymasındır. 3.4.4 numaralı başlık altında geliştirilen “Tutarlılık Değerlendirme Kriterleri” çerçevesi, benzer çalışmalar için kıyaslama ölçütü sağlayabilecek bir standart önerisi niteliğindedir.

Sonuç olarak, bu çalışma hem yöntemsel hem uygulamalı yönleriyle süreç madenciliği, makine öğrenmesi ve eğitimsel analiz alanlarının kesişiminde özgün bir katkı sunmaktadır. Elde edilen bulgular, öğrenme analitiği uygulamalarında süreç temelli yaklaşımların önemini vurgulamakta; veri odaklı karar destek sistemlerinin geliştirilmesine ve öğrenci başarısının erken tahmin edilerek müdahale edilebilmesine yönelik önemli bir temel oluşturmaktadır.

KAYNAKÇA

- Adams, J. N., Peeperkorn, J., Brockhoff, T., Terrier, I., Göhner, H., Uysal, M. S., . . . van der Aalst, W. (2023). Discovering high-quality process models despite data scarcity. Retrieved from <https://arxiv.org/abs/2310.11332>
- Agrawal, R., Gunopulos, D., & Leymann, F. (1988). Mining Process Models from Workflow Logs. In *Sixth International Conference on Extending Database Technology, 1377*, pp. 469-483.
- Akhmedova, Z., & Nargisa, R. (2024). LMS Learning Management System Features. *Science and innovation in the education system*, 3(1), 85-94. doi:doi.org/10.5281/zenodo.10517194
- Alpaydın, E. (2020). *Introduction to Machine Learning 4th Edition*. MIT Press.
- AlQaheri, H., & Panda, M. (2022). An Education Process Mining Framework: Unveiling Meaningful Information for Understanding Students' Learning Behavior and Improving Teaching Quality. *Information*, 13 no. 1: 29. doi:<https://doi.org/10.3390/info13010029>
- Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 175-185.
- Alyadumi, Y. A., & Falcıoglu, P. (2023). Satisfaction of higher education students with blackboard learning system during Covid-19. *Journal of Management Marketing and Logistics*, 10(2), 72-84. doi:doi.org/10.17261/Pressacademia.2023.1744
- Atlam, E., Ewis, A., El-Raouf, M. M., Ghoneim, O., & Gad, I. (2022, July). A new approach in identifying the psychological impact of COVID-19 on university student's academic performance. *Alexandria Engineering Journal*, 61(7), 5223-5233. doi:10.1016/j.aej.2021.10.046
- Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In: *Larusson, J., White, B. (eds) Learning Analytics*, 61-75. doi:doi.org/10.1007/978-1-4614-3305-7_4
- Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. 253-272. doi:10.1017/CBO9781139519526.016
- Bates, T. (2020). *Ten Lessons for Online Learning from the COVID-19 Experience*. Canada: Commonwealth of Learning (COL).
- Batool, S., Rashid, J., Nisar, M. W., Kim, J., & Kwon, H. Y. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, 28, 905-971.
- Bhutoria, A. (2022). Personalized education and artificial intelligence in the United States, China, and India: A systematic review using a human-in-the-loop

- model. *Computers and Education: Artificial Intelligence*, 3.
doi:doi.org/10.1016/j.caeai.2022.100068
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university (4th ed.)*. Open University Press.
- Bilici, Z., & Özdemir, D. (2022). Data Mining Studies in Education: Literature Review For The Years 2014-2020. *Bayburt Eğitim Fakültesi Dergisi*, 17(33), 342-376. doi:doi.org/10.35675/befdergi.849973
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Bogarín, A., Cerezo, R., & Romero, C. (2015). Discovering students' navigation paths in Moodle. *Proceedings of the 8th International Conference on Educational Data Mining*, (pp. 556-557).
- Bogarín, A., Cerezo, R., & Romero, C. (2017). A survey on educational process mining. *WIRES Data Mining Knowl Discovery*, 8(1). doi:https://doi.org/10.1002/widm.1230
- Bonde, S. N., & Kirange, D. K. (2018). Survey on Evaluation of Student's Performance in Educational Data Mining. *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, (pp. 209-213). Coimbatore, India. doi:10.1109/ICICCT.2018.8473228
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145-1159. doi:doi.org/10.1016/S0031-3203(96)00142-2
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. doi:doi.org/10.1023/A:1010933404324
- Brozina, C., Knight, D., Kinoshita, T., & Johri, A. (2019). Engaged to Succeed: Understanding First-Year Engineering Students' Course Engagement and Performance Through Analytics. *IEEE Acces*, 7, 163686-163699. doi:10.1109/ACCESS.2019.2945873
- Çelik, U., & Akçetin, E. (2018). Process Mining Tools Comparison. *AJIT-E: Academic Journal of Information Technology*, 9(34), 97-104. doi:doi.org/10.5824/1309-1581.2018.4.007.x
- Cenka, B. A., Santoso, H. B., & Junus, K. (2022). Analysing student behaviour in a learning management system using a process mining approach. *Knowledge Management & E-Learning*, 14(1), 62-80. doi:doi.org/10.34105/j.kmel.2022.14.005
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 1-58. doi:doi.org/10.1145/1541880.1541882

- Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6), 318-331.
- Chen, J., Man, H., Jin, Q., & Huang, R. (2011, 10). Goal-Driven Navigation for Learning Activities Based on Process Optimization. *2011 International Conference on Internet of Things and 4th International Conference on Cyber, Physical and Social Computing* (pp. 389-395). Dalian, China: IEEE. doi:10.1109/iThings/CPSCoM.2011.97
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. doi:doi.org/10.1145/2939672.293978
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Diamantini, C., Genga, L., Mircoli, A., Potena, D., & Zannone, N. (2024). Understanding the stumbling blocks of Italian higher education system: A process mining approach. *Expert Systems with Applications*, 242, 122747. doi:doi.org/10.1016/j.eswa.2023.122747
- Dolak, R. (2019, November). Using Process Mining Techniques to Discover Student's Activities, Navigation Paths, and Behavior in LMS Moodle. *Innovative Technologies and Learning* (pp. 129-138). Springer, Cham. doi:doi.org/10.1007/978-3-030-35343-8_14
- Dougiamas, M., & Taylor, P. (2003). Moodle: Using Learning Communities to Create an Open Source Course Management System. *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*.
- Encord. (2023, July 18). *Encord Blog*. Retrieved from F1 Score in Machine Learning Explained.: <https://encord.com/blog/f1-score-in-machine-learning/>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 861-874.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996, Nov.). The KDD process for extracting useful knowledge from volumes of data. *Commun. ACM*, 39, 27-34. doi:10.1145/240455.240464
- Feng, G., & Chen, H. (2025). Educational process mining: A study using a public educational data set from a machine learning repository. *Education and Information Technologies*, 30, 8187-8214. doi:doi.org/10.1007/s10639-024-13130-y
- Feng, G., Fan, M., & Ao, C. (2022). Exploration and Visualization of Learning Behavior Patterns From the Perspective of Educational Process Mining. *IEEE Access*, 10, 65271-65283. doi:10.1109/ACCESS.2022.3184111.

- Fitriani, Y. (2020). Analisa Pemanfaatan Learning Management System (LMS) Sebagai Media Pembelajaran Online Selama Pandemi Covid-19. *Journal of Information System, Informatics and Computing*, 4(2), 1-8. doi:10.52362/jisicom.v4i2.312
- Freeman, S., Eddy, S., McDonough, M., Smith, M., Okoroafor, N., Jordt, H., & Wenderoth, M. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. doi:doi.org/10.1073/pnas.1319030111
- Grigorova, K., Malysheva, E., & Bobrovskiy, S. (2017). Application of Data Mining and Process Mining approaches for improving e-Learning Processes. *Information Technology and Nanotechnology 2017*, (pp. 115-121). doi:10.18287/1613-0073-2017-1903-115-121
- Günther, C. W., & Rozinat, A. (2012). Disco: discover your processes. *CEUR Workshop Proceedings*. 940, pp. 40-44. Tallinn, Estonia: Proceedings of the Demonstration Track of the 10th International Conference on Business Process Management (BPM 2012).
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques*. Springer.
- Hand, D. J. (2012). Assessing the performance of classification methods. *International Statistical Review*, 80(3), 400-414.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd. Edition. Springer.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 81-112. doi:doi.org/10.3102/003465430298487
- Hicheur Cairns, A., Gueni, B., Fhima, M., Cairns, A., David, S., & Khelifa, N. (2015). Process Mining in the Education Domain. *International Journal on Advances in Intelligent Systems*, 8(1&2), 219-232.
- Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., . . . Stachl, C. (2021). Machine learning for the educational sciences. *Review of Education*, 9(3). doi: https://doi.org/10.1002/rev3.3310
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms. *International Conference on Artificial Intelligence in Education (AIED)*, 157-171. doi:doi.org/10.1007/978-3-030-23204-7_14
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression (3rd ed.)*. Wiley.
- Iglesias-Pradas, S., Hernández-García, Á., Chaparro-Peláez, J., & Prieto, J. L. (2021). Emergency remote teaching and students' academic performance in higher education during the COVID-19 pandemic: A case study. *Computers in human behavior*, 119. doi:doi.org/10.1016/j.chb.2021.106713

- Intayoad, W., Kamyod, C., & Temdee, P. (2018). Process mining application for discovering student learning paths. *2018 International Conference on Digital Arts, Media and Technology (ICDAMT)* (pp. 220-224). IEEE. doi:10.1109/ICDAMT.2018.8376527
- Kang, H. (2013). The prevention and handling of the missing data. *Korean journal of anesthesiology*, 64(5), 402-406. doi:10.4097/kjae.2013.64.5.402
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.
- Khan, M., Naz, S., Khan, Y., Zafar, M., Khan, M., & Pau, G. (2023). Utilizing Machine Learning Models to Predict Student Performance From LMS Activity Logs. *IEEE Access*, 11, 86953-86962. doi:10.1109/ACCESS.2023.3305276
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, (pp. 1137-1143).
- Liu, F., Zhao, L., Zhao, J., Dai, Q., Fan, C., & Shen, J. (2022). Educational Process Mining for Discovering Students' Problem-Solving Ability in Computer Programming Education. *IEEE Transactions on Learning Technologies*, 15(6), 709-719. doi:10.1109/TLT.2022.3216276
- Lobos, K., Sáez-Delgado, F., Cobo-Rendón, R., Mella Norambuena, J., Maldonado Trapp, A., & Cisternas San Martin, N. (2021). Learning Beliefs, Time on Platform, and Academic Performance During the COVID-19 in University STEM Students. *Frontiers in Psychology*, 12:780852. doi:10.3389/fpsyg.2021.780852
- Loder, A. K. (2024). The Use of Educational Process Mining on Dropout and Graduation Data in the Curricula (Re-)Design of Universities. *Trends in Higher Education*, 3(1), 50-66. doi:doi.org/10.3390/higheredu3010004
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. *McKinsey Global Institute*.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Mittal, U., Sai, S., Chamola, V., & Sangwan, D. (2024). A Comprehensive Review on Generative AI for Education. *IEEE Access*, 12, 142733-142759. doi:10.1109/ACCESS.2024.3468368.
- Nguyen, N., & Winman, T. (2020). A CASE STUDY OF THE IMPLEMENTATION PROCESS OF CANVAS LEARNING MANAGEMENT SYSTEM IN HIGHER EDUCATION. *14th International Technology, Education and Development Conference* (pp. 1399-1405). Valencia, Spain: IATED. doi:doi.org/10.21125/inted.2020.0463

- Olisah, C. C., Smith, L., & Smith, M. (2022). Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. *Computer Methods and Programs in Biomedicine*, 220. doi:10.1016/j.cmpb.2022.106773
- Patil, M. M. (2021, July). Study of Learners Behaviour in virtual learning environment using Process Mining. *2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 1-6. doi:10.1109/CONECCT52877.2021.9622641
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, É. (2012). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Potavin, J., Jongswat, N., & Premchaiswadi, W. (2012). Applying Fuzzy-Genetic mining in conformance and dependency relations. *2012 Tenth International Conference on ICT and Knowledge Engineering*, (pp. 228-235). doi:10.1109/ICTKE.2012.6408560
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1, 81-106. doi:doi.org/10.1007/BF00116251
- Racca, A. R., Sulis, E., & Capecchi, S. (2022). Behavioral Web Tracking in e-Learning: An Educational Process Mining Application. *2022 26th International Conference Information Visualisation (IV)*, 269-274. doi:10.1109/IV56949.2022.00053
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, . . . Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913–929. doi:doi.org/10.1111/ECOG.02881
- Roci, A., & Davidrajuh, R. (2018, 10). A Polynomial-time Alpha-Algorithm for Process Mining. *International Journal of Simulation: Systems, Science and Technology*, 19, 12.1-12.7. doi:10.5013/IJSSST.a.19.05.12
- 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. doi:10.1109/TSMCC.2010.2053532
- Romero, C., & Ventura, S. (2020, May/June). Educational data mining and learning analytics: An updated survey. *Data Mining and Knowledge Discovery*, 10(3). doi:doi.org/10.1002/widm.1355
- Rozinat, A., & van der Aalst, W. (2008). Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33, 64-95. doi:10.1016/j.is.2007.07.001

- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(e3 :e0118432). doi:doi.org/10.1371/journal.pone.0118432
- Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, 2(160). doi:doi.org/10.1007/s42979-021-00592-x
- Sarno, R., Haryadita, F., Sunaryono, D., & Munif, A. (2015). Model discovery of parallel business processes using modified Heuristic Miner. 2015 *International Conference on Science in Information Technology (ICSITech)*, (pp. 30-35). doi:10.1109/ICSITech.2015.7407772
- Siemens, G., & Long, P. (2011, September/October). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46, 31-40.
- Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G., & Vasileva, T. (2024). AI-driven adaptive learning for sustainable educational transformation. *Sustainable Development*. doi:10.1002/sd.3221
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction 2nd Edition*. MIT Press.
- Syakur, A., Sugirin, & Widiarni. (2020, March). The Effectiveness of English Learning Media through Google Classroom in Higher Education. *Britain International of Linguistics, Arts and Education*, 2(1). doi:doi.org/10.33258/biolae.v2i1.218
- Tagoe, M., & Cole, Y. (2020). Using the Sakai Learning Management System to change the way Distance Education nursing students learn: are we getting it right? *Open Learning: The Journal of Open, Distance and e-Learning*, 35(3), 201-221. doi:doi.org/10.1080/02680513.2019.1704232
- Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168-192.
- Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., . . . Altman, R. B. (2001). Missing Value Estimation Methods for DNA Microarrays. *Bioinformatics*, 17(6), 520-525. doi:10.1093/bioinformatics/17.6.520
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Massachusetts.
- Umer, R., Susnjak, T., Mathrani, A., & Suriadi, S. (2017). On predicting academic performance with process mining in learning analytics. *Journal of Research in Innovative Teaching & Learning*, 10(2), 160-176. doi:10.1108/JRIT-09-2017-0022
- van der Aalst, W. (2011). *Process mining Discovery, Eonformance and Enhancement of Business Process*. Heidelberg: Springer. doi:10.1007/978-3-642-19345-3

- van der Aalst, W. (2012). Process Mining: Overview and Opportunities. *ACM Transactions on Management Information Systems (TMIS)*, 3(2), 1-17. doi:10.1145/2229156.2229157
- van der Aalst, W. (2016). *Process mining: data science in action*. Berlin, Germany: Springer. doi:10.1007/978-3-662-49851-4_1
- van der Aalst, W. (2018). *Process Mining: Data Science in Action (2nd Edition)*. Springer. doi:doi:10.1007/978-3-662-49851-4
- van der Aalst, W. M. (2012). Process Mining Manifesto. *Business Process Management Workshops - BPM 2011 International Workshops*. 99, pp. 169-194. Berlin: Springer. doi:doi.org/10.1007/978-3-642-28108-2_19
- van der Aalst, W., Bolt, A., & Van Zelst, S. J. (2017). RapidProM: Mine your processes and not just your data. *arXiv:1703.03740*. doi:doi.org/10.48550/arXiv.1703.03740
- van der Aalst, W., Dongen, B. F., Günther, C. W., Verbeek, H. M., & Weijters, A. J. (2009, 01). ProM: The process mining toolkit. In *Proceedings of the BPM 2009 Demonstration Track (BPMDemos 2009)*, (pp. 1-4). Ulm, Germany.
- van der Aalst, W., Weijters, A., & Maruster, L. (2004). Workflow Mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, 16-9, 1128-1142. doi:10.1109/TKDE.2004.47
- Vautier, A., Enns, K., & Cadaret, C. (2023). Impacts of the COVID-19 Pandemic on Student Performance and Perceptions of Learning. *NACTA Journal*, 67(1). doi:10.56103/nactaj.v67i1.94
- Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(11). doi:doi.org/10.1186/s40561-022-00192-z
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64-70. doi:doi.org/10.1207/s15430421tip4102_2