

# GradeSense

## LMS Aktivite Verileri ile Öğrenci Başarısının Tahmini

---

Ders Adı:

Veri Madenciliği

Proje Ekibi:

- Fatma Abazlı – 22040101142
- Hazal Al Mohammad Algharbi 22040101140
- İmaduddin Hamo – 22040101010

Danışman:

Dr. Yıldız Karadayı

GitHub Proje Linki: <https://github.com/GradeSense210/GradeSense>

PROJE AMACI

# Öğrenci Başarısını Erken

## Aşamada Tahmin Etmek

LMS aktivitelerine bakarak öğrencinin dersi geçip geçmeyeceğini (PASS / FAIL) erken aşamada tahmin etmek, öğrencilere zamanında destek sağlamak ve akademik başarıyı artırmak için kritik öneme sahiptir.

### Basit Modeller

Hızlı ve yorumlanabilir referans noktaları sunar.

### Gelişmiş Modeller

Karmaşık öğrenci davranışlarını ve etkileşimlerini yakalar.

### Derin Öğrenme

Zaman serisi verilerini, etkileşimleri ve gizli örüntüleri daha iyi öğrenir.

# Kullanılan Veri Seti: Kaynak ve İçerik

Öğrenci başarısı tahmini için, Moodle tabanlı Öğrenme Yönetim Sistemi (LMS) üzerinden toplanan kapsamlı bir aktivite veri seti kullanılmıştır.

## Veri Kaynağı ve Genel Yapı

Projede Moodle tabanlı bir LMS sisteminden toplanan öğrenci aktivite verileri kullanılmıştır. Her satır, bir öğrencinin LMS üzerindeki toplam aktivite özetini temsil eder.

- Açık kaynak veri seti: [GitHub](#)
- 25.260 satır (öğrenci-ders kayıtları)
- 69 özellik, 1 hedef değişken

## Özellik Tipleri

Veri seti, öğrencilerin ders sürecindeki etkileşim davranışlarını temsil eden çeşitli aktivite ve zaman bazlı özelliklerden oluşur:

- **Course Activities:** Ders ve kaynak görüntülemeler
- **Assignment Activities:** Ödev görüntüleme, gönderme zamanları
- **Quiz Activities:** Quiz görüntüleme, deneme sayısı, tamamlama süreleri
- **Forum Activities:** Forum görüntüleme, tartışma katılımı
- **Time-based Features:** Platformda geçirilen süre, aktivite zaman ölçümleri

## Hedef Değişken ve Ön İşleme

Modelin hedefi, erken uyarı sistemi oluşturmak için öğrencinin dersi geçip geçmeyeceğini tahmin etmektir.

- **TARGET:** Öğrencinin dönem sonu notu
- **PASS\_FAIL: 1 (Geçti)** eğer  $\text{TARGET} \geq 5$ , **0 (Kaldı)** eğer  $\text{TARGET} < 5$
- **Ön İşleme:** -1 değerleri NaN olarak, aktivite eksik değerleri 0 olarak ele alındı. Gereksiz sütunlar kaldırıldı ve dengeli sınıf dağılımı için Stratified Split uygulandı.

# Kullanılan Modeller: Kapsamlı Bir Bakış

Öğrenci başarısı tahmini için hem geleneksel makine öğrenimi algoritmaları hem de derin öğrenme tabanlı yaklaşımalar kullanılmıştır.

## Geleneksel Modeller (Baseline)

- Logistic Regression
- Decision Tree
- KNN (K-Nearest Neighbors)
- Naive Bayes
- SVM (Support Vector Machine)
- LDA (Linear Discriminant Analysis)

## Gelişmiş & Derin Öğrenme Modelleri

- MLP (Multi-Layer Perceptron)
- 1D-CNN (1 Boyutlu Evrişimsel Sinir Ağı)
- GRU (Gated Recurrent Unit)
- TabNet (Tabular Veri için Özel Ağ)
- Autoencoder + Sınıflandırıcı
- Wide & Deep (Hibrit Yaklaşım)

# Her Model Ne Problemi Çözüyor?

Her bir modelin amacı ve öğrenci davranışlarını analiz etmedeki rolü aşağıda açıklanmıştır.

Logistic Regression

max\_iter=500 (yakınsama için iterasyon sayısı)

Decision Tree

max\_depth=8 (ağaçın maksimum derinliği, aşırı uydurmayı engeller)

KNN

n\_neighbors=5 (sınıflandırma için komşu sayısı)

Naive Bayes

default (varsayılan ayarlar)

SVM

kernel=rbf (Radial Basis Function çekirdeği, doğrusal olmayan ayırım için)

LDA

default (varsayılan ayarlar)

MLP

hidden=(128,64), dropout=0.25 (iki gizli katman ve aşırı uydurmayı önleyici dropout oranı)

1D-CNN

conv(16,32), kernel=3 (16 ve 32filtreli, 3 boyutlu evrişim katmanları)

GRU

hidden=128, layers=2 (128 birimli iki GRU katmanı)

TabNet

n\_d=32, n\_steps=5 (karar adımları ve boyut ayarları)

Autoencoder

latent\_dim=32 (gizli temsilin boyutu)

Wide & Deep

deep=(256,128), dropout=0.2 (derin bileşen için katmanlar ve dropout)

# Detaylı Model Performans Karşılaştırması (Test Seti)

Tabloda, öğrenci başarısı tahmini için kullanılan tüm modellerin test seti üzerindeki detaylı performans karşılaştırmasının sunmaktadır. Farklı özellik mühendisliği yaklaşımıları (Orijinal, Chi-Square, PCA) ve derin öğrenme çerçeveleri (PyTorch) altındaki performansları da dahil edilmiştir.

Model	Accuracy	F1 Score	ROC-AUC
Tabloda, öğrenci başarısı tahmini için kullanılan tüm modellerin test seti üzerindeki detaylı performans karşılaştırmasının sunmaktadır. Farklı özellik mühendisliği yaklaşımıları (Orijinal, Chi-Square, PCA) ve derin öğrenme çerçeveleri (PyTorch) altındaki performansları da dahil edilmiştir.	Logistic Regression (Original)	0.897	0.914
	Logistic Regression (Chi-Square)	0.897	0.914
	Logistic Regression (PCA)	0.893	0.910
	Decision Tree (Original)	0.910	0.925
	Decision Tree (Chi-Square)	0.911	0.926
	Decision Tree (PCA)	0.823	0.853
	KNN (Baseline)	0.739	0.548
	Naive Bayes (Baseline)	0.756	0.577
	MLP (PyTorch)	0.911	0.926
	1D-CNN (PyTorch)	0.909	0.923
	SVM (PCA)	0.943	0.954
	LDA (PCA)	0.936	0.949
	SVM (Chi-Square)	0.961	0.968
	LDA (Chi-Square)	0.953	0.962
	AE + Classifier	0.958	0.965
	GRU (PyTorch)	0.962	0.969
	TabNet (PyTorch)	0.962	0.969
	Wide & Deep (Best)	0.964	0.971

# En İyi Model: Wide & Deep

Kapsamlı karşılaştırmalar sonucunda **Wide & Deep** modeli, öğrenci başarısı tahmininde en üstün performansı sergilemiştir.



## Yüksek Accuracy

Genel doğru tahmin oranı en yüksek modeldir.



## Üstün F1-Score

Dengeli hassasiyet ve geri çağrıma (recall) ile sınıflandırma başarısı.



## Mükemmel ROC-AUC

Pozitif ve negatif sınıfları ayırmadaki üstün yeteneği.



## Hibrit Öğrenme Yeteneği

Hem basit lineer hem de karmaşık derin ilişkileri öğrenir.

LMS öğrenci aktiviteleri, akademik başarıyı tahmin etmek için güçlü bir göstergedir ve **Wide & Deep** gibi gelişmiş modellerle bu potansiyel en üst düzeyde değerlendirilebilir.