

★ Problem Tanımı

Bu çalışmanın amacı, öğrencilerin final notu (**G3**) üzerinde etkili olan faktörleri analiz ederek, farklı regresyon modelleri ile **G3 notunu tahmin etmek** ve modellerin performansını karşılaştırmaktır. Kullanılan veri setinde öğrencilerin demografik bilgileri, aile yapısı, okul ve sosyal yaşam durumları ile ara sınav notları (**G1, G2**) gibi değişkenler yer almaktadır.

Temel Hedefler:

- Hangi modelin daha başarılı tahminler yaptığını belirlemek
- Model performanslarını karşılaştırarak aşırı öğrenme (**overfitting**) veya yetersiz öğrenme (**underfitting**) durumlarını incelemek
- Model çıktılarının pratik kullanım açısından değerlendirilmesi

Model Sonuçları

Model	Test MAE	Test MSE	Test RMSE	Test R ²	Genel Yorum
Linear Regression	0.917	1.856	1.363	0.872	Dengeli ve açıklanabilir, overfitting yok.
Ridge Regression	0.916	1.855	1.362	0.872	Linear Regression'a benzer, stabil sonuçlar.
Random Forest	0.971	2.711	1.647	0.812	Eğitimde çok başarılı, testte düşüş var (overfitting).
Optimized Random Forest	0.931	2.369	1.539	0.836	Hiperparametre optimizasyonu sonrası performansı arttı, overfitting azaldı.
XGBoost	1.036	2.738	1.655	0.811	Eğitimde mükemmel, testte düşüş → overfitting yüksek.
CatBoost Regressor	0.911	2.078	1.441	0.856	Test performansı iyi, eğitimde çok başarılı, hafif overfitting olabilir.

🔍 Çıkarımlar

✓ En İyi Performans Veren Modeller

- Linear Regression** ve **Ridge Regression**, dengeli performanslarıyla öne çıkıyor.
- CatBoost Regressor**, en düşük hata metriklerine sahip, güçlü sonuçlar verdi.

⚠ Aşırı Öğrenme Durumu

- XGBoost** ve optimize edilmemiş Random Forest modellerinde aşırı öğrenme tespit edildi.

- Random Forest optimizasyon sonrası daha dengeli hale geldi (**GridSearchCV** önemli katkı sağladı).

⚙ Genelleme Yeteneđi

- **Linear ve Ridge Regression** modelleri, eğitim ve test sonuçları arasında farkın az olması sayesinde **yüksek genelleme yeteneđine sahip**.
- CatBoost Regressor karmaşık yapıyı iyi öğreniyor, fakat hafif overfitting riski barındırıyor.

🔧 Pratik Kullanım Önerisi

- **Linear / Ridge Regression:** Küçük ve orta boyutlu veri setlerinde, hızlı ve güvenilir sonuçlar için uygun.
- **CatBoost:** Daha büyük, karmaşık veri setlerinde veya doğruluđun kritik olduđu durumlarda tercih edilebilir.
- **Random Forest:** Optimize edilerek orta seviye karmaşıklıkta veri setleri için dengeli bir seçim olabilir.

Genel Sonuç

Bu çalışmada, öğrencilerin yıl sonu notlarını tahmin etmeye yönelik çeşitli regresyon modelleri karşılaştırılmıştır. **Linear ve Ridge Regression**, test setinde en iyi genelleme başarısını göstererek güvenilir tahminler sağlamıştır. Daha karmaşık modeller ise eğitim verisinde mükemmel sonuçlar üretmelerine rağmen, test verisinde aynı başarıyı sürdürememiştir. Bu durum, veri setinin sınırlılığı ve model karmaşıklığı arasındaki dengenin iyi kurulması gerektiğini ortaya koymuştur.