

<https://www.kaggle.com/competitions/mercedes-benz-greener-manufacturing/data>

Case Tanımı

Problem:

4000+ feature içeren Mercedes dataset'i üzerinde üretim süresini (y) tahmin etmek.

Birden fazla algoritma çalıştırılacak:

Support Vector Machine (SVM) (kernel trick, yüksek boyutlu veriye uygun)

Random Forest (ensemble, non-linear decision boundary)

Gradient Boosting (daha iyi generalization)

- Modellerin hem performans skorları (MAE, R^2) hem de matematiksel dayanakları kıyaslanacak.
- Sonuçlar görselleştirmelerle desteklenecek.
- Çözüm sonunda Docker imajı hazırlanacak ve Azure'a deploy edilecek.

Data Scientist 1

- SVM kodunu tamamlayacak (Optuna ile C, γ optimizasyonu).
- Matematiksel açıklamaları rapora ekleyecek.

Data Scientist 2

- Random Forest ve Gradient Boosting kodlarını tamamlayacak.
- PCA & Autoencoder ile dimension reduction yapacak.
- MLflow ile model karşılaştırma loglarını tutacak.

Data Analytics 1

- EDA → feature distribution, correlation heatmap.
- Görselleştirme (PCA scatter, error histogram).

Data Analytics 2

- Feature selection (low variance, correlation >0.95 drop).
- Yorum: hangi feature grupları üretim süresini etkiliyor?

Data Analytics 3

- Azure deployment için Dockerfile + FastAPI `serve.py` hazırlığı.
- Dashboard hazırlığı (R^2 trendi, üretim süresi sapması).

Beklenen Kod Yapısı

```
src/
├── data/
│   └── data_loader.py      # Kaggle dataset yükleme + split
├── features/
│   └── preprocessing.py    # PCA, Autoencoder, scaling
├── models/
│   ├── base_model.py      # ortak interface
│   ├── svm_model.py       # SVM class
│   ├── rf_model.py        # RandomForest class
│   ├── KNN_model.py       # KNN class
│   └── gbm_model.py       # GradientBoosting class
├── training/
│   └── trainer.py         # training loop + Optuna tuning
├── evaluation/
│   └── metrics.py         # MAE,  $R^2$ , residuals
├── visualization/
│   └── plots.py           # PCA scatter, error dist, SHAP
├── serve.py               # FastAPI servis (Azure için)
└── pipeline.py            # ML Pipeline kurgusu
```