



Akbank Makine Öğrenmesi Bootcamp: Yeni Nesil Proje Kampı

Bu projede ilk olarak katılımcılar makine öğreniminde iki temel yol arasında seçim yapacaklardır: sınıflandırma veya regresyon.

Öğrenciler tahmin amaçlı olarak verileri kategoriler halinde sınıflandırmayı veya girdi özelliklerine dayalı olarak sürekli değerleri tahmin etmeyi öğreneceklerdir.

Bu proje, öğrencilere yapay zeka ve makine öğrenmesi alanında sınıflandırma ve regresyon alanlarında veri analizi, model geliştirme ve değerlendirme teknikleri konusunda pratik deneyim kazandırmayı amaçlamaktadır.

1- Proje Konusu

İki seçeneğiniz var: Sınıflandırma veya Regresyon.

2 - Veri Setinizi Seçin

Bu aşamada, projenizde kullanacağınız veri setine karar vereceksiniz. Kendi veri setiniz varsa kullanabilir, verdiğimiz örneklerden birini seçebilir, veya bu dosyanın en altındaki kaynakları inceleyerek orada yer alan bir veri seti seçebilirsiniz.

Sınıflandırma için örnek veri seti: <https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009>

Regresyon için örnek veri seti <https://www.kaggle.com/datasets/vikrishnan/boston-house-prices/data>

3 - Bir Google Colab Dosyası Oluşturun

* Projeniz, .ipynb uzantılı bir dosyada gerçekleştirilmiş olmalı.

* Bu dosyada, kod satırlarının yanı sıra projenizin teknik detaylarını açıkladığınız yorum hücreleriniz olmalı.

4 - Bir GitHub Reposu & README.md Dosyası oluşturun

* Bir Github deposu oluşturmalsınız.

* Bu repoda .ipynb uzantılı proje dosyanız, veri seti dosyanız ve README.md dosyanız olmalı.

* Detaylı teknik anlatımlarınız .ipynb dosyasında yer almalıdır. README.md dosyasında özet bilgilerle projenizden bahsedebilirsiniz.

Örnek bir dosya düzeni için: <https://github.com/gokerguner/notebooks/blob/main/ml01/ml01.ipynb>

5 - Keşifsel Veri Analizi (EDA - Exploratory Data Analysis)

Pandas, matplotlib, seaborn vb. ilgili kütüphaneleri kullanarak seçtiğiniz veri setini açıklayınız. Görselleştirme kütüphaneleri ile görseller, grafikler oluşturup, ilgili pandas yöntemleriyle veri setine ait genel bilgileri aktarın. Analizlerinizi açıklamak için .ipynb içerisinde yorum hücrelerini kullanmalısınız.

Kaynak: <https://gokerguner.medium.com/machine-learning-1-7d4581caa291>

6 - Veri Ön İşleme

Veri kümenize bağlı olarak gerekli ön işleme adımlarını yapmalısınız. Veri setini temizlemek, normalleştirme, label-encoding veya one-hot encoding (kategorik değişkenleriniz varsa), veri kümenizi eğitim ve test kümelerine bölmek vb.

7 - Model Seçimi

- * Birkaç sınıflandırma/regresyon modeli seçin ve bunları ön işlemeden geçirdiğiniz verilerinizle eğitin.
- * Çapraz doğrulamayı(cross validation) kullanarak seçilen modellerin performanslarını inceleyin.
- * En iyi performansı gösteren model ile ilerleyin.

Proje dosyasında, yalnızca ilerlemeye karar verdiğiniz modele yer verebilirsiniz.

8 - Hiperparametre Optimizasyonu

- * Önceki adımda seçilen modelin hiper parametrelerini uygun bir yöntemle optimize edin (Grid Search, Randomized Search veya uygun gördüğünüz herhangi biri).

9 - Model Değerlendirme

Uygun model değerlendirme yöntemlerini kullanarak optimize edilmiş modeli değerlendirin.

- Örneğin Regresyon için kayıp hesaplaması: Ortalama Karesel Hata(Mean Squared Error), Ortalama Mutlak Hata(Mean Absolute Error) vb. ile,
- Sınıflandırma için karışıklık matrisi(Confusion matrix) oluşturarak:
 - Doğruluk(accuracy), kesinlik(precision), duyarlılık(recall), F1 puanı(F1 score) özelliklerine yer vererek.

<https://gokerguner.medium.com/machine-learning-2-korelasyon-matrиси-özelliк-seçimi-sınıfların-dengesizliđi-karar-ağaçları-af993bd8ea66>

Veri Seti Kaynakları

1. ***Kaggle*** (<https://www.kaggle.com/datasets>)
2. ***UCI Machine Learning Repository*** (<https://archive.ics.uci.edu/ml/index.php>)
3. ***Google Dataset Search*** (<https://datasetsearch.research.google.com/>)
4. ***AWS Public Datasets*** (<https://registry.opendata.aws/>)
5. ***Microsoft Research Open Data*** (<https://msropendata.com/>)
6. ***Data.gov*** (<https://www.data.gov/>)
7. ***Data.world*** (<https://data.world/>)
8. ***FiveThirtyEight*** (<https://data.fivethirtyeight.com/>)
9. ***The World Bank Data*** (<https://data.worldbank.org/>)
10. ***GitHub*** - Awesome Public Datasets Collection (<https://github.com/awesomedata/awesome-public-datasets>)