

# Proje: Ürün Başlığından Açıklama ve Kategori Tahmini

**Süre:** min. 3 gün - max. 7 gün

## Amaç

Ürün başlığı verisi kullanarak iki temel görev yapılacaktır:

- **Başlıktan Açıklama Üretimi** (Üretken Yapay Zeka - GenAI)
- **Başlıktan Kategori Tahmini** (Gözetimli Öğrenme - ML)

Proje çıktısı Google Colab veya Kaggle üzerinde çalıştırılacak ve GitHub'a yüklenecektir.

## 1. Veri Seti Hazırlığı

- Açık kaynak erişimli bir e-ticaret veri seti seçiniz.
- **Önerilen veri setleri:**
  - [E-Commerce Behavior Data from Multi-Category Store](#)
  - [Retailrocket E-commerce Dataset](#)
  - [Online Retail II \(UCI\)](#)
  - [Amazon Reviews 2023 \(HuggingFace\)](#)
- Yalnızca *product\_title*, *product\_description*, *category* gibi alanları kullanınız.
- Veri setinden ilk 1000–3000 satırla çalışabilirsiniz.

## 2. Keşifsel Veri Analizi (EDA)

- *head()*, *describe()*, *value\_counts()* gibi temel incelemeler yapınız.

- Basit grafiklerle kategori dağılımını gösteriniz (*Matplotlib* veya *Seaborn* kullanabilirsiniz).
- En az 2–3 grafik oluşturup, veriyi yorumlayınız.

### 3. Başlıktan Açıklama Üretimi (GenAI)

- HuggingFace üzerinde *t5-small* veya *flan-t5-small* gibi küçük modelleri kullanarak başlıklardan açıklama üretiniz.
- Yaklaşık 10 örnek için açıklama çıktısı alınız ve karşılaştırınız.
- `pipeline("text2text-generation")` fonksiyonu yeterli olacaktır.

### 4. Kategori Tahmini (Gözetimli Öğrenme)

- Başlıkları *TfidfVectorizer* ile vektörleştiriniz.
- *LogisticRegression* veya *RandomForest* sınıflandırıcı modellerinden biri ile eğitim yapınız.
- Modelinizi *train\_test\_split* kullanarak test edin.
- Sonuçları *accuracy* ve *F1 score* ile değerlendiriniz.
- Basit bir *classification\_report* çıktılayınız.

### 5. GitHub Teslimi

- Proje dosyanızı *.ipynb* formatında kaydedip **GitHub**'a yükleyiniz.
- **README.md** dosyanızda projenizin kısa özetini, yöntemlerinizi ve önemli çıktıların ekran görüntülerini paylaşınız.
- Proje teslimi min 3 gün - max 7 gün olarak tanımlanmıştır.

### Opsiyonel

- Kümeleme: *TF-IDF* + *KMeans* ile benzer ürün gruplarını keşfediniz.
- Daha gelişmiş embedding için *sentence-transformers* kullanabilirsiniz.
- Uzun açıklamalar varsa *T5 model*i ile özet çıkarabilirsiniz.

## Önerilen Araçlar

| Amaç           | Araçlar / Kütüphaneler           |
|----------------|----------------------------------|
| Ortam          | Google Colab veya Kaggle         |
| NLP & GenAI    | HuggingFace Transformers         |
| Veri İşleme    | pandas, scikit-learn             |
| Görselleştirme | Matplotlib, Seaborn              |
| Vektörleme     | TfidfVectorizer                  |
| Modelleme      | LogisticRegression, RandomForest |

## İletişim

Herhangi bir sorunuz olduğunda iletişime geçebilirsiniz.

**Özlem Ekici**

Mail: [ozlem@turkwise.com.tr](mailto:ozlem@turkwise.com.tr)