**Data Set:** https://amazon-reviews-2023.github.io/ web sitesinden aşağıdaki kategorilere ait review ve meta data setleri indirilecektir:

* All\_Beauty
* Digital\_Music
* Health\_and\_Personal\_Care

Her proje için ayrıca belirtilen adımları takip ederek sonuçlar elde edilecek ve raporlanacaktır.

İndirilen veriler aşağıdaki gibi okunabilmektedir:

A screenshot of a computer code

Description automatically generated

A screenshot of a review

Description automatically generated

A screenshot of a computer

Description automatically generated

**Proje 1: Duygu Analizi (Sentiment Analysis)**

**Amaç:**

Review veri setinden metin tabanlı bir sentiment analiz modeli oluşturmak. Model, kullanıcı yorumlarını pozitif veya negatif olarak sınıflandırır.

**Adımlar:**

1. **Veri Etiketleme:**
   * Rating >= 3 olan yorumlar "pozitif," < 3 olan yorumlar "negatif" olarak etiketlenecektir.
2. **Veri Birleştirme ve Hazırlama:**
   * Üç kategoriye ait review veri setleri birleştirilecek.
   * Veriler karıştırılarak train-test split yapılacaktır (80/20 veya 70/30).
3. **Ön İşleme:**
   * Metin temizleme: Noktalama işareti, durak kelimeler (stop words), küçük/büyük harf dönüşümü vb.
   * Sayısallaştırma: TF-IDF, Word2Vec veya BERT gibi teknikler kullanılarak metinler sayısallaştırılacak.
4. **Modelleme:**
   * Makine öğrenmesi modelleri (Logistic Regression, Random Forest vb.) ve LSTM gibi bir derin öğrenme modeli ile sentiment analizi yapılacaktır.
   * Modellerin doğruluğu, F1-skoru ve karışıklık matrisi gibi performans metrikleri ile karşılaştırılacak.
5. **Görsel Analiz:**
   * Her kategori için ayrı bir Word Cloud oluşturulacaktır.
6. **Store Skorlama**
   * Average\_rating ,sentiment analiz sonucu gelen olumlu ve olumsuz yorumlardan, rating\_number değerinden bir skorlama yaparak başarılı store’ları listeleme. Store başarı sıralaması projesi yapılacak.

**Teslimat:** Kod, çıktılar ve yorumları içeren bir Jupyter Notebook dosyası.

**Proje 2: Görsel Sınıflandırma (Image Classification)**

**Amaç:**

Meta veri setindeki large image görsellerini kullanarak kategori sınıflandırması yapmak. "Hangi görsel hangi kategoriye aittir?" sorusuna yanıt aranacaktır.

**Adımlar:**

1. **Veri Hazırlık:**
   * Meta veri setindeki "image" sütunundan large görselleri indirilecektir.
   * İndirme şekli:

Aşağıdaki gibi images sütunundaki 0. Indeksteki değerin görseli okunmaktadır.

A screen shot of a computer program

Description automatically generated

* + Donanım sınırı aşılırsa, her kategoriden belirli sayıda örnek alınarak çalışılabilir.

1. **Ön İşleme:**
   * Görüseller normalize edilecek ve uygun boyutlara (224x224 gibi) getirilecektir.
2. **Modelleme:**
   * Derin öğrenme tabanlı bir model (ConvNet) veya transfer öğrenme modelleri (VGG, ResNet, MobileNet vb.) kullanılarak sınıflandırma yapılacaktır.
3. **Sonuçlar:**
   * Modelin doğruluk oranı, karışıklık matrisi ve kategori bazlı performans metrikleri raporlanacaktır.

**Teslimat:** Kod, çıktılar ve analizlerin bulunduğu bir Jupyter Notebook dosyası.

**Proje 3: Öneri Sistemi (Recommendation System)**

**Amaç:**

Meta veri setindeki ürün açıklamalarını kullanarak benzer ürünleri tespit eden bir öneri sistemi geliştirmek.

**Adımlar:**

1. **Veri Hazırlık:**
   * Meta veri setindeki ürün açıklamaları (tittle ve detail) kullanılacaktır.
2. **Metin Temizleme ve Encode Etme:**
   * Açıklamalar metin temizleme işlemlerinden geçirilecektir.
   * Sentence Transformers veya TF-IDF kullanılarak açıklamalar vektörize edilecektir.
3. **Benzerlik Hesaplama:**
   * Cosine similarity veya diğer benzerlik algoritmaları kullanılarak ürünler arasındaki benzerlik oranları tespit edilecektir.
4. **Sonuçlar:**
   * Belirli bir ürün için en benzer ürünler listelenecek.
   * Bu liste, kullanıcıya öneri olarak sunulabilecektir.

**Teslimat:** Kod, analizler ve çıktıların bulunduğu bir Jupyter Notebook dosyası.