



Amazon Yorumları için Duygu Analizi



Amazon üzerinden satışlarını gerçekleştiren ev tekstili ve günlük giyim odaklı üretimler yapan **Kozmos**, ürünlerine gelen yorumları analiz ederek ve aldığı şikayetlere göre özelliklerini geliştirerek satışlarını artırmayı hedeflemektedir. Bu hedef doğrultusunda yorumlara duygu analizi yapılarak etiketlenecek ve etiketlenen veri ile sınıflandırma modeli oluşturulacaktır.

Veri seti belirli bir ürün grubuna ait yapılan yorumları, yorum başlığını, yıldız sayısını ve yapılan yorumu kaç kişinin faydalı bulduğunu belirten değişkenlerden oluşmaktadır.

4 Değişken

5611 Gözlem

489 KB

Star	Ürüne verilen yıldız sayısı
HelpFul	Yorumu faydalı bulan kişi sayısı
Title	Yorum içeriğine verilen başlık, kısa yorum
Review	Ürüne yapılan yorum



Proje Görevleri

.....

Görev 1: Metin Ön İşleme

Adım 1: amazon.xlsx verisini okutunuz.

Adım 2: Review değişkeni üzerinde ;

- a.** Tüm harfleri küçük harfe çeviriniz.
- b.** Noktalama işaretlerini çıkarınız.
- c.** Yorumlarda bulunan sayısal ifadeleri çıkarınız.
- d.** Bilgi içermeyen kelimeleri (stopwords) veriden çıkarınız.
- e.** 1000'den az geçen kelimeleri veriden çıkarınız.
- f.** Lemmatization işlemini uygulayınız.

Görev 2: Metin Görselleştirme

Adım 1: Barplot görselleştirme işlemi için;

- "Review" değişkeninin içerdiği kelimelerin frekanslarını hesaplayınız, tf olarak kaydediniz.
- tf dataframe'inin sütunlarını yeniden adlandırınız: "words", "tf" şeklinde
- "tf" değişkeninin değeri 500'den çok olanlara göre filtreleme işlemi yaparak barplot ile görselleştirme işlemini tamamlayınız.

Adım 2: WordCloud görselleştirme işlemi için;

- "Review" değişkeninin içerdiği tüm kelimeleri "text" isminde string olarak kaydediniz.
- WordCloud kullanarak şablon şeklinizi belirleyip kaydediniz.
- Kaydettiğiniz wordcloud'u ilk adımda oluşturduğunuz string ile generate ediniz.
- Görselleştirme adımlarını tamamlayınız. (figure, imshow, axis, show)

Görev 3: Duygu Analizi

Adım 1: Python içerisindeki NLTK paketinde tanımlanmış olan **SentimentIntensityAnalyzer** nesnesini oluşturunuz.

Adım 2: SentimentIntensityAnalyzer nesnesi ile polarite puanlarını inceleyiniz;

- "Review" değişkeninin ilk 10 gözlemi için polarity_scores() hesaplayınız.
- İncelenen ilk 10 gözlem için compound skorlarına göre filtreleyerek tekrar gözlemleyiniz.
- 10 gözlem için compound skorları 0'dan büyükse "pos" değilse "neg" şeklinde güncelleyiniz.
- "Review" değişkenindeki tüm gözlemler için pos-neg atamasını yaparak yeni bir değişken olarak dataframe'e ekleyiniz.

NOT: SentimentIntensityAnalyzer ile yorumları etiketleyerek, yorum sınıflandırma makine öğrenmesi modeli için bağımlı değişken oluşturulmuş oldu.

Görev 4: Makine Öğrenmesine Hazırlık

Adım 1: Bağımlı ve bağımsız değişkenlerimizi belirleyerek datayı train test olarak ayırınız.

Adım 2: Makine öğrenmesi modeline verileri verebilmemiz için temsil şekillerini sayısala çevirmemiz gerekmektedir;

- a.** TfidfVectorizer kullanarak bir nesne oluşturunuz.
- b.** Daha önce ayırmış olduğumuz train datamızı kullanarak oluşturduğumuz nesneye fit ediniz.
- c.** Oluşturmuş olduğumuz vektörü train ve test datalarına transform işlemini uygulayıp kaydediniz.

Görev 5: Modelleme (Lojistik Regresyon)

Adım 1: Lojistik regresyon modelini kurarak train dataları ile fit ediniz.

Adım 2: Kurmuş olduğunuz model ile tahmin işlemleri gerçekleştiriniz;

- a.** Predict fonksiyonu ile test datasını tahmin ederek kaydediniz.
- b.** classification_report ile tahmin sonuçlarınızı raporlayıp gözlemleyiniz.
- c.** cross validation fonksiyonunu kullanarak ortalama accuracy değerini hesaplayınız.

Adım 3: Veride bulunan yorumlardan rastgele seçerek modele sorulması;

- a.** sample fonksiyonu ile "Review" değişkeni içerisinde örneklem seçerek yeni bir değere atayınız.
- b.** Elde ettiğiniz örneklemini modelin tahmin edebilmesi için CountVectorizer ile vektörleştiriniz.
- c.** Vektörleştirdiğiniz örneklemini fit ve transform işlemlerini yaparak kaydediniz.
- d.** Kurmuş olduğunuz modele örneklemini vererek tahmin sonucunu kaydediniz.
- e.** Örneklemi ve tahmin sonucunu ekrana yazdırınız.

Görev 6: Modelleme (Random Forest)

Adım 1: Random Forest modeli ile tahmin sonuçlarının gözlenmesi;

- RandomForestClassifier modelini kurup fit ediniz.
- Cross validation fonksiyonunu kullanarak ortalama accuracy değerini hesaplayınız.
- Lojistik regresyon modeli ile sonuçları karşılaştırınız.

miuul

miuul.com