



NECMETTİN ERBAKAN
ÜNİVERSİTESİ

Mühendislik Fakültesi

Bilgisayar Mühendisliği Bölümü

Doğal Dil İşleme Dersi Proje Formu

Proje Başlığı	
Ürün İncelemelerinin Sınıflandırması	

Öğrenci Bilgileri	
Öğrenci No	18010011067
Öğrenci Ad Soyad	Bayram YARIM

Arş. Gör. Sinem ÇINAROĞLU

Kasım 2022
Konya

1. Proje Konusunun Önemi

Ürün İncelemelerinin Sınıflandırması nedir?

Günümüzde insanlar için online alışveriş kaçınılmaz için bir ihtiyaç haline gelmiştir. Satın alınmak istenen ürünün internet üzerinden incelenmesi, fiyat kıyaslaması yapılması ve sipariş edilip kısa sürede eline ulaşması gibi pek çok avantajı bilinçli internet kullanıcılarını online alışverişe yönlentmektedir. E-ticaret sitelerinin son yıllarda elde ettikleri gelirler de, insanların online alışverişe gösterdikleri yoğun ilgiyi açık şekilde göstermektedir.

Ürün incelemesinin e-ticaret firmaları için önemi oldukça büyüktür. E-ticaret firmalarının yaptığı ürün incelemesi çalışmaları firmalara büyük katkılar sağlamaktadır. Ürün incelemesi sayesinde e-ticaret firmaları müşterilerine ürünleri hakkında daha kolay bir şekilde bilgi vermektedir.

Online alışveriş de yapılan ürün incelemeleri müşteriler içinde büyük önem arz etmektedir. Bu incelemelerin iyi – pozitif yıldızı yüksek olan ürünlerde müşterilerin alım tercihi iyi yönde olmakta, kötü – negatif yıldızı düşük olan ürünlerde müşterilerin alım tercihi pek olmamaktadır.

2. Projenin Amacı ve Hedefi

Projenin amacı ürün incelemelerinde çıkan yorumlara göre sınıflandırma yapmaktır. İyi yönde olumlu yorumlar pozitif , kötü yönde olumsuz yorumlar negatif olarak sınıflandırılacaktır. Çıkan bu sonuçlarla markalar hakkında çıkarımlar yapmakta projenin amaç ve hedefleri arasında yer almaktadır.

3. Projenin iş-zaman çizelgesi

Projede yer alacak başlıca iş paketleri (İP) ve hedefleri, her bir iş paketinin hangi sürede gerçekleştirileceği, başarı ölçütü ve projenin başarısına katkısı “İş-Zaman Çizelgesi” doldurularak verilir.

İş-zaman çizelgesi

İP. No.	İş Paketlerinin Adı ve Hedefleri
1	Veri Setlerinin Hazırlanması: Ürünlere ait yorumlar trendyol e-ticaret sitesinden çekip veri tabanına kaydetmek.
2	Veri Normalizasyonu: Verileri normalize etmek.
3	Veri Ayırıklaştırma: Kullanılacak ürünlere ait yorumları ayırıklaştırmak.
4	Makine Öğrenmesi: Makine öğrenmesi ile verilerden negatif ve pozitif sonuçlarını elde etmek

Proje Özeti

Projeyi iş zaman çizelgesine göre adım adım takip ederek uyguladım.

* Öncelikle trendyol e-ticaret sitesinden verileri bir uygulama ile sqllite veritabanına kaydettim. Sayfalar arasında gezinerek 4109 farklı üründen toplamda 43923 adet veri toparlanmıştır. Bu verilerde id, ürün id, ürün yorumu ve ürüne ait puanlandırma(1..5) verisi bulunmaktadır.(**Bknz : Resim-1 : Ürünlere ait veri seti**) Hazırlanan bu veri setini normalize etmek için programa gönderdim.

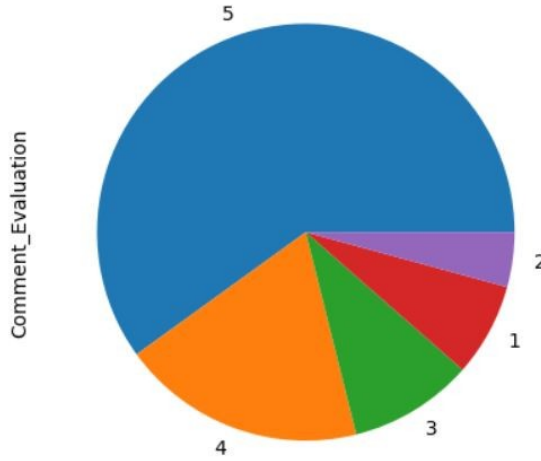
3004	119	Harika bir ürün bayıldım yumuşacık ve çok kaliteli M giyiniyor...	5
3005	119	Resimde olan geldi. Çok kaliteli ve şık. Bir beden büyük alınabi...	5
3006	119	Lastiği biraz sert geldi bu model için ama kumaşı gayet güzel, ...	4
3007	119	Pek rahat bulmadım. Lastiği biraz sıkı ve sert	3
3008	119	model guzel fakat yumusak bir kulot beklemeyin	2
3009	119	Yan kısımları çok sert	1

Resim 1.: Ürünlere ait veri seti

* Verileri programda kullanmak için python kütüphanelerinden yararlanarak verileri sqllite veritabanından okudum ve csv dosya formatına çevirdim. Elde edilen bu verileri normalize ettim. Yüklenen bu verilerde gereksiz sütunlar(id, product id) silindi. Verilere ait ürün yıldızlarını görsel olarak pasta grafiğinde gösterdim. (Bknz : Resim 2 : Ürün yıldızlarının görselleştirilmesi)

```
In [5]: # yorumların yıldızlarının görselleştirilmesi
df['Comment_Evaluation'].value_counts().plot.pie()
```

```
Out[5]: <AxesSubplot:ylabel='Comment_Evaluation'>
```



Resim 2.: Ürün yıldızlarının görselleştirilmesi

Bu işlemten sonra ürünlere ait yorumları ayırıklaştırma işlemine geçtim. Ayırıklaştırma işleminde yorum yıldızlarına göre verileri ikiye bölme işlemi gerçekleştirdim. Bölme işlemi yorum yıldızı 3 ve altında ise işlem sonucu Negatif(0) , 4 ve üstü ise Pozitif(1) olarak ayarladım. (Bknz : Resim 3 – Veri işleme pozitif ve negatif ayırıklaştırma)

```
In [45]: df.loc[(df["Comment_Evaluation"]>3), "Comment_Evaluation"]=1
df
```

```
Out[45]:
```

	Comment_Content	Comment_Evaluation	word_count	char_count	clean_text	clean_text_tok
0	cok guzel yaaaa	0	3	15	cok guzel yaaaa	[cok, guzel, yaaaa]
1	Çok beğendik kaliteli güzel bi kumaşı var	0	7	41	çok beğendik kaliteli güzel bi kumaşı var	[çok, beğendik, kaliteli, güzel, bi, kumaşı, var]
2	Çok güzel 🍷	0	3	12	çok güzel	[çok, güzel]
3	Hızlı teslimat.Yumuşacık,rahat ve bol gömlek	0	5	44	hızlı teslimat yumuşacık rahat bol gömlek	[hızlı, teslimat, yumuşacık, rahat, bol, gömlek]
4	hiç guzel degil	0	3	15	hiç guzel degil	[hiç, guzel, degil]
...
43918	Süper 🍷🍷🍷 kalıbı duruşu	0	4	26	süper kalıbı duruşu	[süper, kalıbı, duruşu]
43919	Altıyla üstü arasında sanki 2 beden var. Üstün...	0	27	182	altıyla üstü arasında sanki beden var üstünün ...	[altıyla, üstü, arasında, sanki, beden, var, ü...]
43920	Dokusu harika yumaşacık harika tavsiye ediyorum	0	6	47	dokusu harika yumaşacık harika tavsiye ediyorum	[dokusu, harika, yumaşacık, harika, tavsiye, e...]
43921	Ürün gayet güzeldi	0	3	18	ürün gayet güzeldi	[ürün, gayet, güzeldi]
43922	Kaliteli ürün	0	2	13	kaliteli ürün	[kaliteli, ürün]

Resim 3.: Veri işleme pozitif ve negatif ayırıklaştırma

Veri ayrıklaştırma işleminden sonra pandas, numpy ve nltk vs.. kütüphanelerinden hazır fonksiyonları kullanarak gereken işlemleri gerçekleştirdim. Bu işlemler yorumlardaki boşlukları, gereksiz karakterleri silmek, yorumları modellemek ve formatlamaktır.

En son olarak programa verilen girişi sonuçlandırmaktır. (Bknz: Resim 4 ve Resim 5 – İşlem sonuçları)

```
In [46]: d = {'Comment_Content': ["çok beğendim tekrar almayı düşünüyorum"]}
df_test=pd.DataFrame(d)

df_test['clean_text'] = df_test['Comment_Content'].apply(lambda x: finalpreprocess(x)) #preprocess the data
X_test=df_test['clean_text']
df_test.drop('Comment_Content', inplace=True, axis=1)
X_vector=tfidf_vectorizer.transform(X_test) #converting X_test to vector
y_predict = lr_tfidf.predict(X_vector) #use the trained model on X_vector
y_prob = lr_tfidf.predict_proba(X_vector)[:,1]
df_test['pred_prob']= y_prob
df_test['Pred']= y_predict
df_test['Pred'] = np.where(df_test['Pred'] == 0, 'Negative (-)', "Positive (+)")
print_full(df_test)
```

	clean_text	pred_prob	Pred
0	çok beğendim tekrar almayı düşünüyorum	0.97	Positive (+)

Resim 4.: İşlem Sonuçları - Pozitif

```
In [47]: d = {'Comment_Content': ["hiç beğenmedim hem de çok geç geldi"]}
df_test=pd.DataFrame(d)

df_test['clean_text'] = df_test['Comment_Content'].apply(lambda x: finalpreprocess(x)) #preprocess the data
X_test=df_test['clean_text']
df_test.drop('Comment_Content', inplace=True, axis=1)
X_vector=tfidf_vectorizer.transform(X_test) #converting X_test to vector
y_predict = lr_tfidf.predict(X_vector) #use the trained model on X_vector
y_prob = lr_tfidf.predict_proba(X_vector)[:,1]
df_test['pred_prob']= y_prob
df_test['Pred']= y_predict
df_test['Pred'] = np.where(df_test['Pred'] == 0, 'Negative (-)', "Positive (+)")
print_full(df_test)
```

	clean_text	pred_prob	Pred
0	hiç beğenmedim hem de çok geç geldi	0.01	Negative (-)

Resim 5.: İşlem sonuçları - Negatif

4. Projede kullanılacak yöntem, donanımlar ve yazılımlar ile ilgili bilgiler

Yazılım Araçları

Python – Jupyter-Notebook – Anaconda framework

Numpy – Pandas Kütüphaneleri

Java

Sqlite