

Metin Madenciliği ve Doğal Dil İşleme Kullanarak Marka

İtibar ve Şikayet Analizi

Anıl Alkan

Özet

Çalışmada Hepsiburada firmasına ait olan kullanıcı yorumları ve şikayetleri toplanmış, yorumların duygu analizi, şikayetlerin sınıflandırılması ve analizi yapılmıştır. Şikayet verisi'nin elde edilmesi için metin madenciliği teknikleri kullanılmış, duygu analizi ve şikayet sınıflandırması için Doğal Dil İşleme kütüphanelerinden olan NLTK ve Makine Öğrenimi algoritmalarından olan XGBoost, Naive Bayes, Lojistik Regresyon ve Keras algoritmaları kullanılmıştır.

Çalışma Python programlama dili ile Google Colab ve Deepnote platformlarında gerçekleştirilmiştir. Elde edilen çıktılar analiz edilerek Tableau raporlama aracı ile dashboard formatında görselleştirilmiştir.

Kullanılan Araçlar



Yöntem

Müşteri şikayetleri şikayetvar.com adresinden BeautifulSoup kütüphanesi kullanılarak web scraping yöntemi ile çekilmiştir. BeautifulSoup kütüphanesi kullanılarak çekilen veriler .csv formatında bir dosyaya kaydedildikten sonra hatalı, eksik gözlemler temizlenmiştir.

```
import requests
import numpy as np
import mysql.connector
import pandas as pd
import csv
from array import array
from bs4 import BeautifulSoup as bs
```

URL = 'https://www.sikayetvar.com/hepsiburada?page=0'

```
def sikayetvarr():
    for page in range(0,30):
        req = requests.get(URL + str(page), headers=headers)
        soup = bs(req.text, 'html.parser')

        titles = soup.find_all('h2', attrs={"complaint-title"})
        my_array = np.asarray(titles)

        for i in range(4,19):
            if page>1:
                print(my_array[i])
            else:
                print(my_array[i])
```

```
with open('newsikayet.csv', 'w', newline='') as file:
    writer = csv.writer(file)
    writer.writerow(["Text"])
    writer.writerow([titles[i].text])
    writer.writerow([titles[i].text])
```

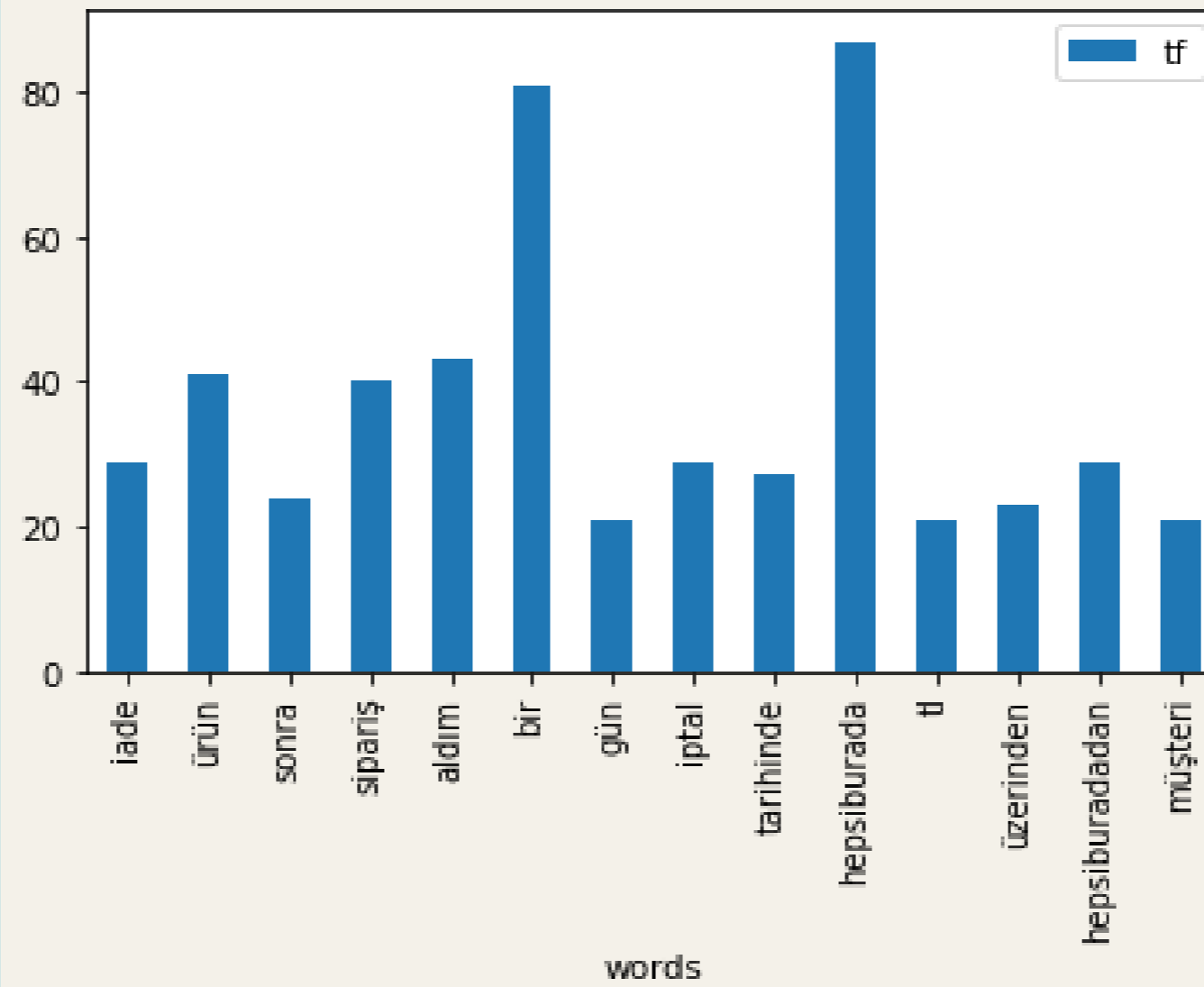
sikayetvarr()

NLP Uygulamaları

Duygu analizi için kullanılacak veriler, model kurulmadan önce veri ön işleme ve NLP(Doğal Dil İşleme) aşamalarından geçmektedir. Bu aşamada NLTK(Natural Language Toolkit) kütüphanesi kullanılarak büyük küçük harf dönüşümü, noktalama işaretleri ve sayıların kaldırılması, stopwords'lerin silinmesi, lemmatizasyon ve kelime köklerinin ayrılması işlemleri gerçekleştirilmiştir.



Veri ön işleme aşamasından sonra metin verisinde en çok tekrar eden kelimeler WordCloud ve Matplotlib kütüphaneleri ile görselleştirilmiştir.



Veri ön işleme aşamasından sonra sklearn count vectorizer ve tf-idf count vectorizer kütüphaneleri kullanılarak metnin vektörlere dönüşme işlemi gerçekleştirilmiştir.

	about	all	cent	cents	money	new	old	one	two	
doc	1	1	3	1	1	1	1	1	1	In theory (a)
Index	0	1	2	3	4	5	6	7	8	
doc	1	1	3	1	1	1	1	1	1	In practice (b)

Vektörize edilen veri eğitim ve test kümesi olmak üzere bölünüp, makine öğrenimi algoritması çalıştırılmaktadır.

```
text1 = "bu ürün çok iyi herkese tavsiye ederim"
text10 = "beklediğim gibi çıkmadı"
text13 = "berbat beğenmedim"

texts = [text1, text10, text13]

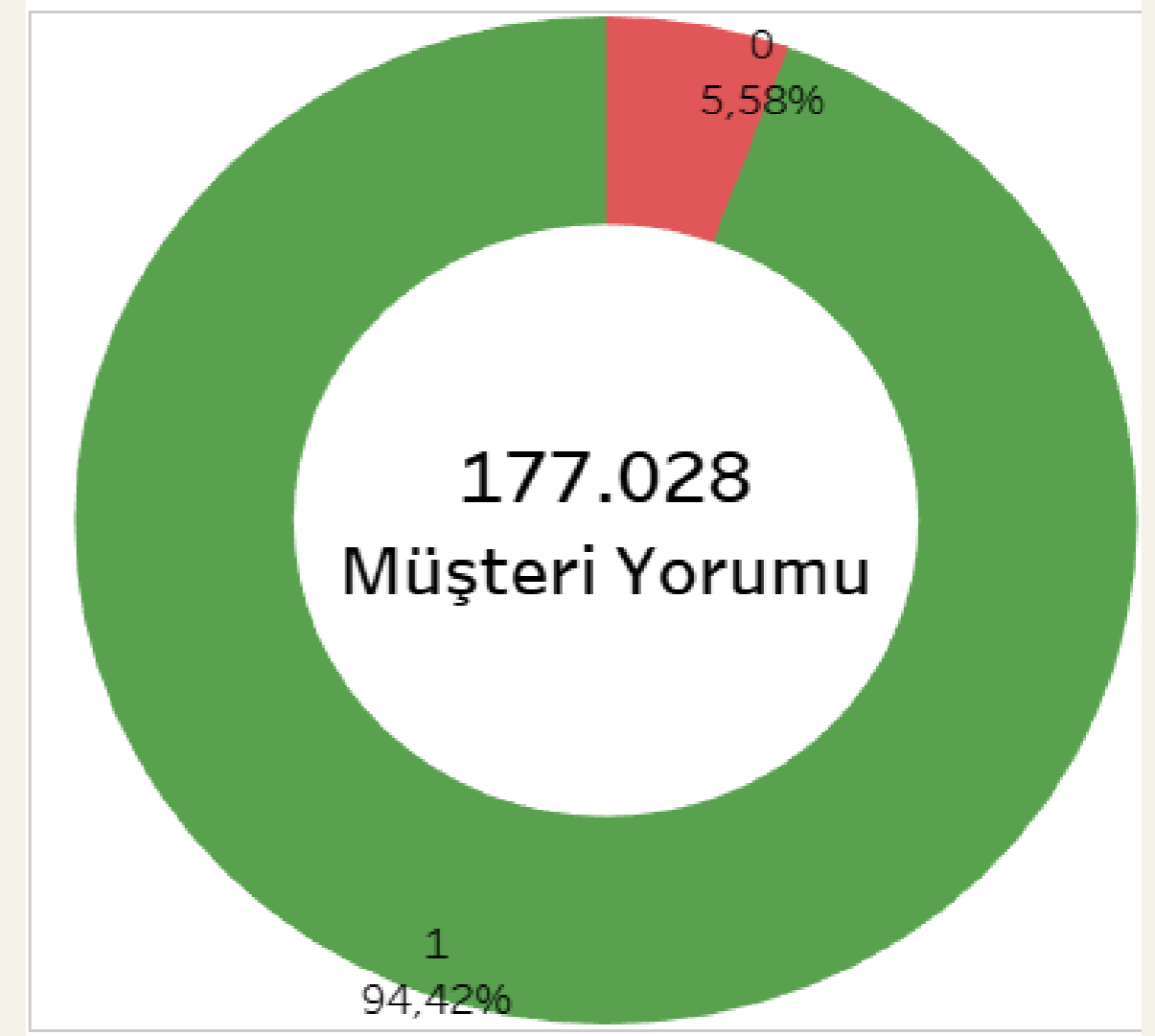
model.predict(tokens_pad)

array([[0.9933615 ],
       [0.51821446 ],
       [0.05696875 ]], dtype=float32)
```

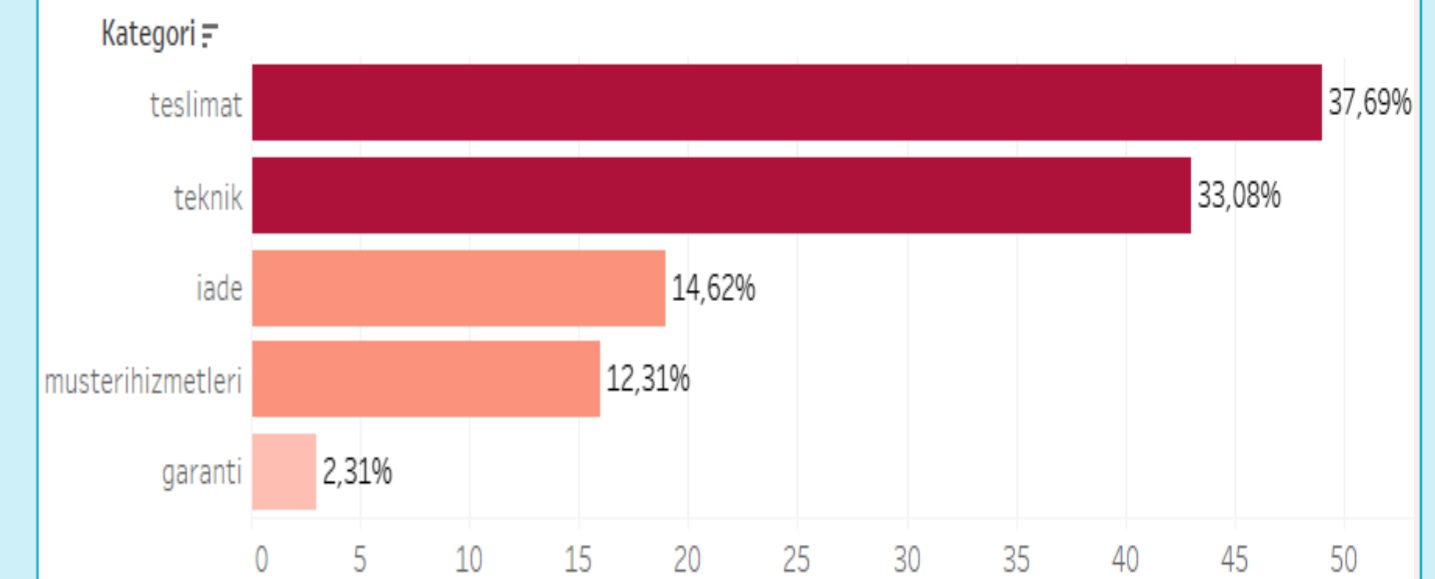
Sonuç

Çalışmada kullanılan Hepsiburada müşteri yorumları duygu analizi yöntemi ile sınıflandırılmış 177.028 adet müşteri yorumundan %94.42'si olumlu %5.58'i ise olumsuz olarak belirlenmiştir. Şikayetvar.com'dan alınan hepsiburada şikayet verileri 5 farklı kategoriye ayrılarak etiketlenmiş ve makine öğrenimi algoritmaları ile tahminlenmesi gerçekleştirilmiştir. Buna göre en fazla şikayet gelen kategori %37.69 oranı ile teslimat olmuştur. Şikayetleri oranlarını sırasıyla %33.08 ile teslimat, %14.62 ile iade, %12.31 ile müşteri hizmetleri, %2.31 ile garanti kategorileri izlemektedir.

Olumlu-Olumsuz Yorumlar



Şikayet Kategorileri



Kaynaklar

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python*. O'Reilly Media.

B. Şardağı Bozyiğit ve Ç. Tarhan, "MAKİNE ÖĞRENMESİ YÖNTEMLERİ KULLANILARAK SOSYAL MEDYADA MARKA İTİBAR ANALİZİ", *Yönetim Bilişim Sistemleri Dergisi*, c. 6, sayı. 2, ss. 57-76, Ara. 2020

<https://zeyrek.readthedocs.io/en/latest/>

<https://www.nltk.org>

İletişim Bilgileri

E-Posta: anilalkan96@gmail.com
Tel: +90 (541) 207 92 23

Mentor: Prof.Dr. Vahap TECİM

