



RESTORANT YORUM SINIFLANDIRILMASI VE SONUÇ ÇIKARIMI

içerik

01

VERİ SETİ ANALİZİ

02

VERİ TEMİZLEME

03

MAKİNE ÖĞRENMESİ İLE SINIFLANDIRMA

04

SONUÇ ÜRETİMİ

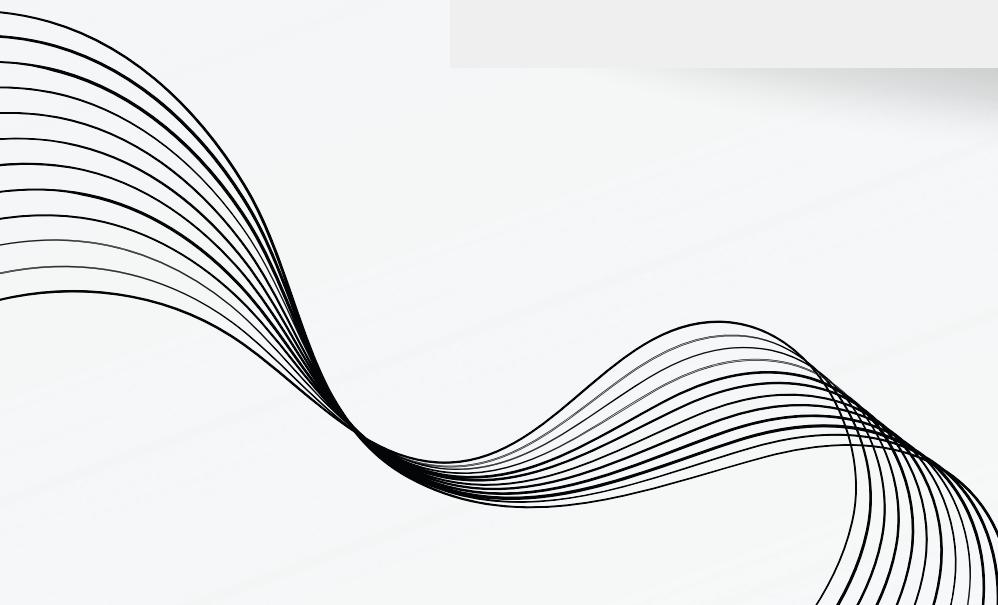
VERI SETİ ANALİZİ



YELP DATA VERİSİ
15 BİN RESTORANT
MİLYONLARCA YORUM



5 RESTORANT
12000 BIN YORUMA İNDİRGENMİŞTİR.



```
df_5businesses = dataset_business[dataset_business["business_id"].isin(business_mostreviews)] #içinde top 5 restoranı tut  
df_5businesses
```

	business_id	name	address	city	state	postal_code	latitude	longitude	stars	review_count	is_open	attributes	categories	hours
141	SZU9c8V2GuREDN5KgyHFJw	Santa Barbara Shellfish Company	230 Stearns Wharf	Santa Barbara	CA	93101	34.408715	-119.685019	4.0	2404	1	{'OutdoorSeating': 'True', 'RestaurantsAttire': ...}	Live/Raw Food, Restaurants, Seafood, Beer Bar,...	{'Monday': '0:0-0:0', 'Tuesday': '11:0-21:0', ...}
957	W4ZEKkva9HpAdZG88juwyQ	Mr. B's Bistro	201 Royal St	New Orleans	LA	70130	29.954387	-90.068363	4.0	2064	1	{'NoiseLevel': 'u'average', 'RestaurantsReser...}	Bars, Breakfast & Brunch, Restaurants, Barbequ...	{'Wednesday': '11:30-20:0', 'Thursday': '11:30...}
4497	GBTPC53ZrG1ZBY3DT8Mbcw	Luke	333 Saint Charles Ave	New Orleans	LA	70130	29.950742	-90.070416	4.0	4554	1	{'RestaurantsReservations': 'True', 'Restauran...}	German, Restaurants, Seafood, Cocktail Bars, F...	{'Monday': '7:0-21:0', 'Tuesday': '7:0-21:0', ...}

```
df_review.head() # İlk 5 veri kontrolü
```

	review_id	user_id	business_id	stars	useful	funny	cool	text	date
0	KU_O5udG6zpxOg-VcAEodg	mh_-eMZ6K5RLWhZyISBhwA	XQfwVwDr-v0ZS3_CbbE5Xw	3.0	0	0	0	If you decide to eat here, just be aware it is...	2018-07-07 22:09:11
1	BiTunyQ73aT9WBnpR9DZGw	OyoGAe7OKpv6SyGZT5g77Q	7ATYjTIgM3jUlt4UM3lypQ	5.0	1	0	1	I've taken a lot of spin classes over the year...	2012-01-03 15:28:18
2	saUsX_uimxRICVr67Z4Jig	8g_iMtSiwikVnbP2etR0A	YjUWPpl6HXG530lwP-fb2A	3.0	0	0	0	Family diner. Had the buffet. Eclectic assortm...	2014-02-05 20:30:30
3	AqPFMleE6RsU23_auESxiA	_7bHUi9Uuf5__HHc_Q8guQ	kxX2SOes4o-D3ZQBkiMRfA	5.0	1	0	1	Wow! Yummy, different, delicious. Our favo...	2015-01-04 00:01:03
4	Sx8TMOWLNUJBWer-0pcmoA	bcjbaE6dDog4jkNY91ncLQ	e4Vwtrqf-wpJfwesgvdgxQ	4.0	1	0	1	Cute interior and owner (?) gave us tour of up...	2017-01-14 20:54:15

VERİ MUHENDİSLİĞİ

```
def cleaning_text(text):
    # Noktalama işaretlerini ve sayıları ve fazladan boşlukları kaldırma
    text = re.sub(r"\w\s+", "", text) # Noktalama
    text = re.sub(r"\d+", "", text) # Sayılar
    text = re.sub(r"\s+", " ", text.strip()) # Fazladan boşluklar
    text = text.lower()
    # Stopwords kaldır ve lemmatize et
    text = " ".join([lemmatizer.lemmatize(word) for word in text.split() if word not in stop_words])
    return text
```

```
# 80 kelimededen fazla olan yorumları filtreleme, fazlası LLM modelinin kafasını karıştırıyor
df_5businessreviews = df_5businessreviews[df_5businessreviews['word_count'] <= 80]
```

```
# Yeni DataFrame, ve satır sayısı kontrolü
df_5businessreviews[['text', 'word_count']]
```

```
from imblearn.over_sampling import SMOTE
import numpy as np
import pandas as pd

# SMOTE işlemi için X (embedding'ler) ve y (hedef değişken)
smote = SMOTE(random_state=42, sampling_strategy={1.0: 1000, 0.0: 1000})

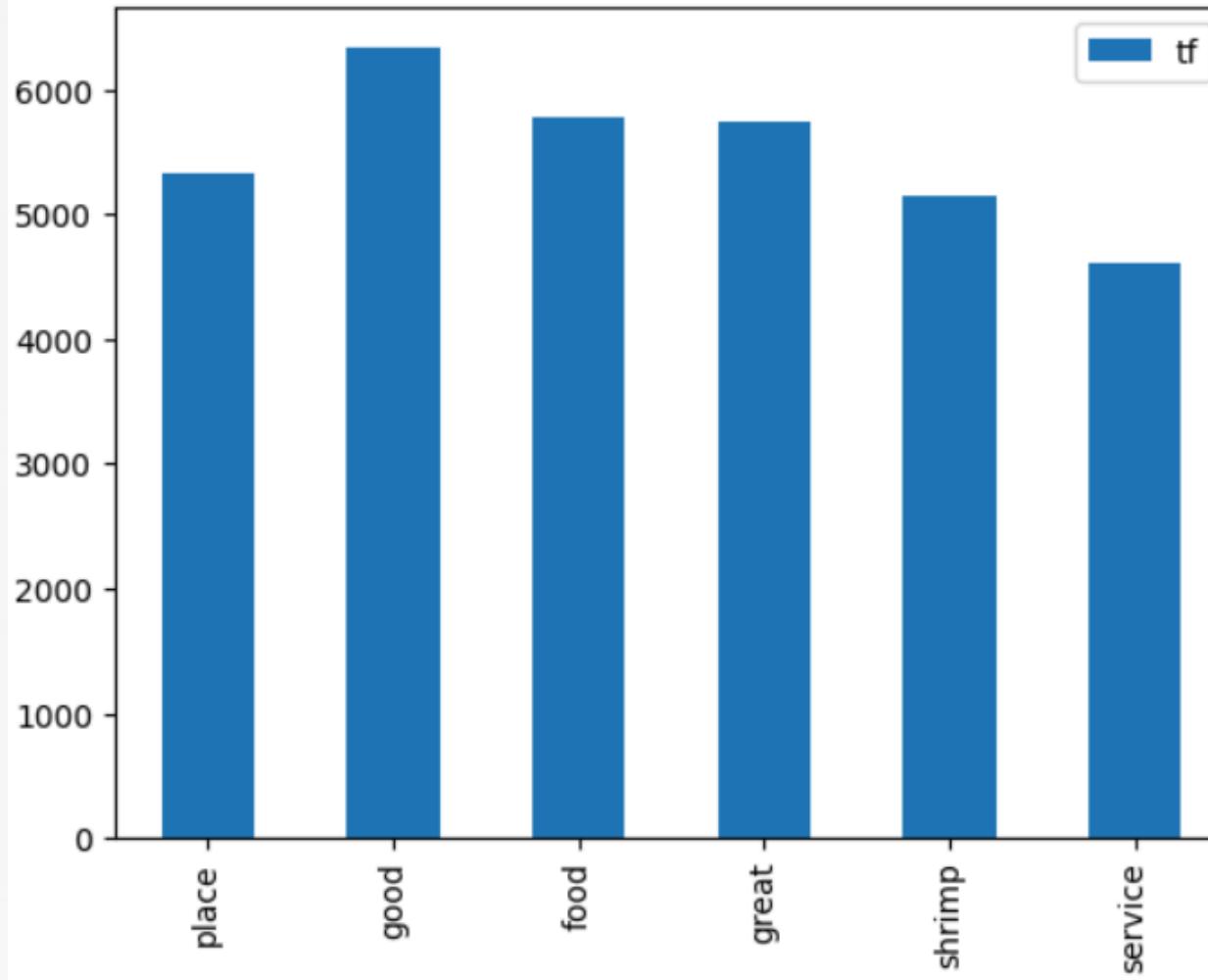
# Azınlık sınıflarını artırma sürecim
X_smote, y_smote = smote.fit_resample(X, y)

# Sınıf dağılımını print ettiğim kısım
print("SMOTE sonrası sınıf dağılımı:")
print(pd.Series(y_smote).value_counts())
SMOTE sonrası sınıf dağılımı:
stars
4.0    1606
5.0    1500
3.0    1000
2.0    1000
1.0    1000
Name: count, dtype: int64
```

		text	cleaned_text
37	We checked in around 2:30 pm. Check-in was quick and easy with complimentary valet parking (always appreciated). The Peppermill is huge and reminded us of the Caesars Palace in Las Vegas. They ...	checked around pm checkin quick easy complimentary valet parking always appreciated peppermill huge reminded u caesar palace la vega around eight restaurant site offering almost type food could wa...	
44	In a word... "OVERRATED!". The food took forever to come out. The burger was WAY overcooked (Especially when I ordered it Medium Rare) and the Seafood Jumbalaya was Decent. Not worth the price....	word overrated food took forever come burger way overcooked especially ordered medium rare seafood jumbalaya decent worth price better place eat french quarter horrible everything heard place woul...	

01

```
import matplotlib.pyplot as plt  
tf.columns = ["words", "tf"]  
tf[tf["tf"] > 4000].plot.bar(x="words", y="tf")  
plt.show()
```



```
from wordcloud import WordCloud  
text = " ".join(i for i in df_5businessreviews.cleaned_text)  
  
wordcloud = WordCloud(max_font_size=50,  
                      max_words=80,  
                      background_color="white").generate(text)  
  
plt.figure()  
plt.imshow(wordcloud, interpolation="bilinear")  
plt.axis("off")  
plt.show()
```



EMBEDDING

```
from sentence_transformers import SentenceTransformer
import numpy as np

# SBERT model yüklemeye sürecim
model = SentenceTransformer('paraphrase-MiniLM-L3-v2')

# Embedding'leri oluşturduğum alan
texts = df_5businessreviews['cleaned_text'].tolist() # Yorum metinlerini al
embeddings = model.encode(texts, batch_size=128, show_progress_bar=True) # Embedding'leri çıkar
```

Embedding'lerin boyutu: (4047, 384)

```
X = embeddings # SBERT embedding'leri
y = df_5businessreviews['stars'] # Yıldızlar hedef değişken
```

SİNİFLANDIRMA

- RANDOM FOREST

Siniflandırma Raporu:				
	precision	recall	f1-score	support
1.0	0.87	0.98	0.92	300
2.0	0.92	0.89	0.90	300
3.0	0.91	0.63	0.74	300
4.0	0.50	0.60	0.54	482
5.0	0.53	0.50	0.51	450
accuracy			0.69	1832
macro avg	0.75	0.72	0.72	1832
weighted avg	0.70	0.69	0.69	1832

- LOGISTIC REGRESSION

Siniflandırma Raporu:				
	precision	recall	f1-score	support
1.0	0.78	0.89	0.83	300
2.0	0.70	0.77	0.73	300
3.0	0.59	0.63	0.61	300
4.0	0.52	0.45	0.48	482
5.0	0.59	0.56	0.58	450
accuracy			0.63	1832
macro avg	0.64	0.66	0.65	1832
weighted avg	0.62	0.63	0.63	1832

SONUÇ ÜRETME

MODEL 1

- Restaurant ID: GBTPC53ZrG1ZBY3DT8Mbcw
Summary: the raw bar was fantastic. Those were the best Alaskan crab legs i've had . service was spot on. Professional but not overbearing or obvious . bread and butter at the table was excellent. Crun
-
- Restaurant ID: PY9GRFzr4nTZeINF346Q0w
Summary: the Peppermill is rated the casino with the loosest slots in casino . the rooms are stellar for the price, and there weren't any bed bugs in our bed . they have eight restaurants on site offering almost any type
-
- Restaurant ID: SZU9c8V2GuREDN5KgyHFJw
Summary: the rock crab special at the shellfish company on the wharf was a great value . the crab was huge, delicious and meaty, and the salad and soup were fresh . if you're looking for a
-
- Restaurant ID: W4ZEKkva9HpAdZG88juwyQ
Summary: the burger was WAY overcooked and the seafood jumbalaya was Decent . the bread pudding was a tad dry in the city that's supposed to be the mother lode of bread pudding .
-
- Restaurant ID: vN6v8m4D045Z4pp8yxxF_w
Summary: the shrimp and grits are hands down the best we've had in the city . the lump crab omelette with the sauce was just the cure for a dull grey rainy morning . everything on the menu is

SONUÇ ÜRETME

MODEL 1

- Restaurant ID: GBTPC53ZrG1ZBY3DT8Mbcw
Summary: the raw bar was fantastic. Those were the best Alaskan crab legs i've had . service was spot on. Professional but not overbearing or obvious . bread and butter at the table was excellent. Crun
-
- Restaurant ID: PY9GRFzr4nTZeINF346Q0w
Summary: the Peppermill is rated the casino with the loosest slots in casino . the rooms are stellar for the price, and there weren't any bed bugs in our bed . they have eight restaurants on site offering almost any type
-
- Restaurant ID: SZU9c8V2GuREDN5KgyHFJw
Summary: the rock crab special at the shellfish company on the wharf was a great value . the crab was huge, delicious and meaty, and the salad and soup were fresh . if you're looking for a
-
- Restaurant ID: W4ZEKkva9HpAdZG88juwyQ
Summary: the burger was WAY overcooked and the seafood jumbalaya was Decent . the bread pudding was a tad dry in the city that's supposed to be the mother lode of bread pudding .
-
- Restaurant ID: vN6v8m4D045Z4pp8yxxF_w
Summary: the shrimp and grits are hands down the best we've had in the city . the lump crab omelette with the sauce was just the cure for a dull grey rainy morning . everything on the menu is

SONUÇ ÜRETME

MODEL 1

BLEU Score: 0.0629

ROUGE Scores:

```
ROUGE-1: {'r': 0.6211904761904762, 'p': 0.36435578913839783, 'f': 0.4565336719233276}
ROUGE-2: {'r': 0.31714285714285706, 'p': 0.15452559185595893, 'f': 0.20747667767569716}
ROUGE-L: {'r': 0.5283333333333333, 'p': 0.30172638955247655, 'f': 0.38214499753465325}
```

SONUÇ ÜRETME

MODEL 2

Star Rating: 5

Review: Very nice place for a great lunch spot. Service was great and the service was good. We have been here a few times in the past several years. very good. My

Star Rating: 3

Review: My husband and I ate here the previous day, and we had the shrimp and grits.

MODEL 1-2

STREAMLIT

Yelp Review Generator

Eğitilen DistilGPT-2 modeli ile yıldız derecesine göre yorum oluşturun.

Yorum isteğinizi yazın (örnek: '5 yıldızlı yorum yap'):

4 yıldızlı yorumunu göreyim bakalım kankı

Yorum Oluştur

Oluşturulan Yorum:

Star Rating: 4 Review: Great place to get oysters and I love it! We were seated at the bar the next time I was in NOLA. They were awesome, and the food and service was phenomenal. The wait was only

Restoran ve Yorum Analizi

Beş farklı restoran arasından bir seçim yapın ve yorumları görün.

Bir restoran seçin:

Luke

Seçilen restoran: Luke

Yorum: *The raw bar was amazing with fantastic Alaskan crab legs and excellent bread and butter.*