

# T.C. SAKARYA ÜNİVERSİTESİ

# İŞLETME FAKÜLTESİ YÖNETİM BİLİŞİM SİSTEMLERİ BÖLÜMÜ

# VERİ GÖRSELLEŞTİRME

Oğuzhan GEDİKLİ B171306301

SAKARYA,2021

# İÇİNDEKİLER

1.	ΚĬ	RLÍ VERÍ	3
1	1.1.	"Selenium" İle Veri ekme	3
1	1.2.	Veri Çıktısı	4
2.	TE	EMİZ VERİ	5
2	2.1.	Verinin Temizlenmesi	5
2	2.2.	Sık Geçen Kelimeler	6
2	2.3.	Kelime Bulutunun Oluşturulması	7
3.	MA	AKİNE ÖĞRENMESİ	8
3	3.1.	Modelin Oluşturulması	
3	3.2.	Verilerin Sayısallaştırılması	9
	3.2	2.1. TF-IDF Modeli	9
	3.2	2.2. BoW Modeli	10
3	3.3.	Modelin Eğitilmesi	11
3	3.4.	CrossValidation	eti tanımlanmamış.
3	3.5.	Çekilen Verinin Üzerinde Modelin Uygulanması	13
4.	DU	J <b>YGU ANALİZİ</b>	14
5.	SO	NUÇLAARIN GÖRSELLEŞTİRİLMESİ	15
4	5.1.	Logistic Regression Görselleştirmesi	15
4	5.2.	Navie Baves Görsellestirme	16

## 1. KİRLİ VERİ

### 1.1. "Selenium" İle Veri ekme

İlk aşamada Twitter üzerinden veri çekmek için Python'da "Selenium" kütüphanesini kullanıyoruz. Selenium kütüphanesi bize Twitter üzerinden istediğimiz kelimeler ve kelime gurupları içerisinden son atılan Twetlerin içeriklerine, beğeni sayısına, yorum sayısı, kullanıcı adı, ismi ve retweet sayısına ulaşmamızı sağlamakta.

Ana dosyada tamamına ulaşacağımız Selenium kodlarının birkaç örneği aşağıdaki gibidir.

• Kullanıcıdan girmesi talep edilen verinin Twitter'a giriş yapıldıktan sonra arama ekranına yazdırılması. Arama sonuçları içerisinden son veriye ulaşılması;

• Arama ekranını aşağıya kaydırma işleminin ardından Twetler içerisinden verinin seçilmesi;

```
sayfakaynagi = browser.page_source
        soup = BeautifulSoup(sayfakaynagi, "html.parser")
        tweetler = soup.find all("div", attrs={"data-testid": "tweet"})
        print(tweetler)
        for tweet in tweetler:
            try:
                Isim = tweet.find("div", attrs={
                    'css-901oao css-bfa6kz r-jwli3a r-1qd0xha r-a023e6 r-b88u0q r-ad9z0x r-
bcqeeo r-3s2u2q r-qvutc0'}).text
                KullaniciAdi = tweet.find("div", attrs={
                    'css-901oao css-bfa6kz r-111h2gw r-18u37iz r-1qd0xha r-a023e6 r-
16dba41 r-ad9z0x r-bcqeeo r-qvutc0'}).text
                Tweet = tweet.find("div", attrs={
                    'css-901oao r-jwli3a r-1qd0xha r-a023e6 r-16dba41 r-ad9z0x r-bcqeeo r-
bnwqim r-qvutc0'}).text
                YorumSayisi = tweet.find("div", attrs={"data-testid": "reply"}).text
                RetweetSayisi = tweet.find("div", attrs={"data-testid": "retweet"}).text
                BegeniSayisi = tweet.find("div", attrs={"data-testid": "like"}).text
                writer.writerow([Isim, KullaniciAdi, YorumSayisi, RetweetSayisi, BegeniSayisi
 Tweet ])
```

### 1.2. Veri Çıktısı

Selenium ile çektiğimiz veriyi öncelikle CSV dosyası olarak kaydediyor ardından bu CSV dosyasını aynı kod içerisinde EXCEL dosyasına çeviriyoruz ve daha düzenli bir hale getiriyoruz.

• Çekilen verinin Pandas kütüphanesiyle kaydedilmesine ardından EXCEL dosyasına dönüştüren Python kodları;

```
from pandas import read_csv
ss = read_csv("tweetler.csv")
ss.to_excel("tweetler_excel.xlsx")
```

• Excel'e kaydedilmiş ilk çıktının son hali;

1	A	В	С	D
1	Tweet	Isim	KullaniciAdi	BegeniSayisi
2	Olm gelsenize Türkiye çok güzel oldu artık tüp kuyrukları yok, asgari ücr	Evohe	@evoheee	
3	"Asgari ücret insan onuruna yaraşır bir düzeye yükseltilmeli"	yenimuhalefet.com	@yenimuhalefet	1
4	Asgari ücret en az 30 bin tl olsun vekillerin maaşoda 2500 olsun	Aykut Salih Akkan	@AyktSalihAkkn09	
5	Arnavutluk'a 522 adet konut hediye etmişiz. Siz asgari ücret kaç TL olaca	Umut Gültekin	@umtgltkn1903	1
6	Taksiciler bile istediği zammı alıyor. Bizim asgari ücret için de işçi sendil	Fabrique in Turkey	@MevzuNeKardes	
7	#asgariucret tahminim asgari ücret 2706 tl olur ama yine geçen yıl yaptıl	Panama	@panamareis	
8	Asgari ücret isterseniz 5 bin olsun. Bu enflasyonla bi boka yaramaz.	ŞURA	@ulansura	
9	Erdoğan'dan asgari ücret açıklamasıhttps://erzurumajans.com/erdogan-	Erzurum Ajans	@ERZURUMAJANS	
10	O zamanlar asgari ücret 350 liraydı	Atilla Balkan	@AtillaBalkan	
11	Gt yap Beraber takipçi kasalımYoruma çökünBeraber büyütelim hesapla	(GT VAR)	@sessizlik00_00	2
12	Partimizin asgari ücret önerisinin en az 3.100 Tl olması gerektiğini anlat	CHP Ümraniye İlçe Başkan	@ChpUmraniyeIlce	2
13	Asgari ücret olmuşBir kırtık bişeMaaşı alınca kalmıyor neşeSenin üstüne	Sezer	@Sezer28845220	2
14	DİSK Çukurova Bölge Temsilciliği olarak Adana'dan seslendik: "Sefalet ü	H.YAŞAR GÜNDOĞDU	@HYAARGNDODU2	
15	Gt yap Beraber takipçi kasalımYoruma çökünBeraber büyütelim hesapla	GT VAR	@Kedi633	4
16	Tuik : Ne asgari ücret ne de asgari ücret artış oranı biz belirlemiyoruz.AÜ	Wilma Çakmaktaş	@wilmakedi	5
17	Erken secçim olursa bu fiyatlar düşecekmi ?? Asgari üçret 5.000 TL mi ol	FENERBAHCE 1907	@1907_MuratY	
18	Ne bir arabam var nede evimOnca yıl çalışmşamBüküldü belimÜç kuruş	Sezer	@Sezer28845220	
19	Bitane montum var tam 9yıllıkGören diyer ola bu kılıgdurmadan mama i	Sezer	@Sezer28845220	1
20	3.liste#taburespor #iddaa #fenerinmacivar #asgariücret #Turkcell #Yanır	Shef	@TabureShef	
21	2.liste#taburespor #iddaa #fenerinmacivar #asgariücret #Turkcell #Yanır	Shef	@TabureShef	
22	Asgari ücret 10 bin olsa az ise CHP li belediyelerde çalışan işçilerin maaş	Resul Kusur	@resul3380	
23	Erdoğan'dan asgari ücret çıkışıhttps://odatv4.com/erdogandan-asgari-u	Odatv	@odatv	1
24	Bu günde yoksun, Yarında olma, Yarında gelsen, Yanında olmam	Özgür Balkan	@ozgurbalkan34	
25	Asgari ücret şu oldu bu oldu hiç fark etmez,Sen 1 senede 2 yada 3 e katl	Naim Kolay	@NaimKolay	
26	1.liste#taburespor #iddaa #fenerinmacivar #YanındayızErolBulut #asgari	Shef	@TabureShef	
27	Asgari ücret??	işşiz adam derneği	@Pala35410814	
28	asgari ücret yerlerde alım gücü yok ekonomi bitik insanlar fakir çalışmal	bora	@boraabalci	
29	DİSK 3 bin 800 liralık asgari ücret talebini yineledi	On4 Haber	@on4haber	
30	Eğer sürpriz olmazsa asgari ücret 2750 TL dolaylarında olur	beso	@besirduzce	
31	#Asgariücret zam yapıyorsunuz birkaç gün olmadan ulaşıma yiyecek içeç	Seyit Ahmet	@SeyitMirik	
32	Asgari ücret yükseltisin sözüne bile muhalefet edenler saraydamı yaşıy	Kurban Tayfun	@tayfun_kurban	1

Tablo 1: Twitter üzerinden çekilen verinin ilk(kirli) hali.

### 2. TEMIZ VERİ

#### 2.1. Verinin Temizlenmesi

Bir önceki adımda çekilen Twitter verileri fazlasıyla kirli. Bundan sonraki adımların uygulanabilmesi için bu çekilen verinin önişleme sürecine girmesi gerekmektedir. Bu süreçte veri üzerinde öğrenmeyi zorlaştıracak her şeyi temizliyor ve daha anlaşılır hale getiriyoruz.

• Twetlerin içindeki linklerin, hashtaglerin, etiketlerin, noktalama işaretlerinin, sayıların ve GT'lerin temizlenmesi;

```
data['Tweet'] = data['Tweet'].str.replace (r'http\S+', '')
data['Tweet'] = data['Tweet'].str.replace (r'@\S+', '')
data['Tweet'] = data['Tweet'].str.replace (r'#\S+', '')
data['Tweet'] = data['Tweet'].str.replace('[^\w\s]','')
data['Tweet'] = data['Tweet'].str.replace('\d','')
data['Tweet'] = data['Tweet'].str.replace('gt', '')
```

• Twitter verisini Selenium ile çektiğimiz için içerisinde RT'ler de mevcut. Bu yüzden aşağıdaki kod tekrar eden verileri temizler;

```
data.drop_duplicates(subset=["Tweet","KullaniciAdi","BegeniSayisi","Isim"], keep=Fa
lse,inplace=False)
```

• Verinin bütün harflerini küçültmek, anlam ifade etmeyen kelimeleri çıkarmak ve en az kullanılan kelimeleri silmek için aşağıdaki kodu kullanıyoruz;

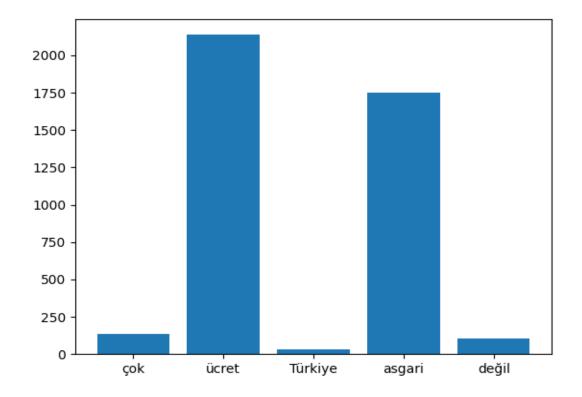
```
    data['Tweet'] = data['Tweet'].apply(lambda x: " ".join(x.lower() for x in x.split()))
    data['Tweet'] = data['Tweet'].apply (lambda x: " ".join (x for x in x.split () i f x not in sw))
    azgecenkelimeler = pd.Series(" ".join(data['Tweet']).split()).value_counts
    silinecekkelimeler = pd.Series(" ".join(data['Tweet']).split()).value_counts()[-10:]
    data['Tweet'] = data['Tweet'].apply(lambda x: " ".join(x for x in x.split() if x not in silinecekkelimeler))
```

#### 2.2. Sık Geçen Kelimeler

Verimiz temizlendikten sonra sık geçen kelimeleri ortaya çıkarmak verinin kullanımını daha da kolaylaştıracak ve ne ile uğraştığımızı görmemizi sağlayacaktır.

• Sık geçen kelimelerin bulunması için ve bu verilerin grafik haline getirilmesi için kullanılan kod aşağıdaki gibidir.

• Sık geçen kelimelerden oluşan grafik;



Grafik 1: Sık geçen kelimeler grafiği.

### 2.3. Kelime Bulutunun Oluşturulması

Kelime bulutu verimizin son halini oluşturacağımız sırada güzel bir görsellik sağlar ve kullanılan verinin içeriği hakkında fikir verir.

• Kelime bulutunu oluşturabilmek için Python üzerinden kullanılan kodlar aşağıdaki gibidir.

```
    Tweet = " ".join(i for i in data.Tweet)
    print("Kelime bulutunun oluşturulması : \n")
    wordcloud = WordCloud(background_color = "black").generate(Tweet)
    plt.imshow(wordcloud, interpolation = "bilinear")
    plt.axis("off")
    plt.tight_layout(pad = 0)
    plt.show()
```

Kelime bulutunun görseli;



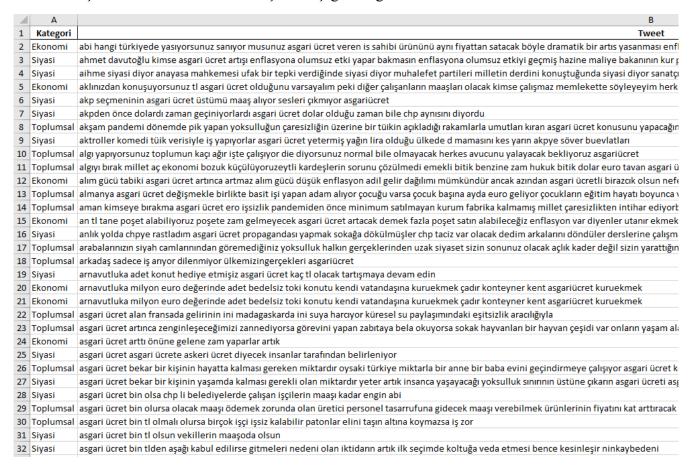
Resim 1: Temizlenmiş verinin kelime bulutu.

# 3. MAKİNE ÖĞRENMESİ

#### 3.1. Modelin Oluşturulması

Temizlediğimiz veri üzerinden yeni bir makine öğrenmesi seti oluşturuyoruz. Toplamda 3 kategoriden 300 adet veriyi seçiyoruz ve bu seçilmiş verilere her bir kategoriye denk gelen 100 twet ile modelimizi oluşturuyoruz.

Oluşturulan test verisinin örnek çıktısı aşağıdaki gibidir.



Tablo 2: Oluşturulan makine öğrenmesi setinin Excel çıktısı.

## 3.2. Verilerin Sayısallaştırılması

Makine öğrenmesi gerçekleştirirken makinenin bizim verimiz içerisindeki cümleleri tanıyabilmesi için bu veri kaynağını sayısal hale çevirmemiz gerekmektedir. Bunun için iki yöntem kullanılmıştır. Bunlardan ilki TF-IDF, ikincisi ise BOW\_VECTOR modelidir.

#### 3.2.1. TF-IDF Modeli

Tf-idf modeli bir kelimenin herhangi bir doküman içerisinde ne kadar önemli olduğunu belirtmek için tasarlanmış istatiksel bir ölçüttür. Bir verinin doküman içerisindeki frekansına ulaşmamızı sağlar.

• TF-IDF modelini uygulamak için Python üzerinden "sklearn" kütüphanesini kurmamız gerekmektedir. Kütüphaneyi kurduktan sonra verimizi sayısallaştırmak için kullanacağımız kod;

```
    vectorizer = TfidfVectorizer ()
    train_vectors = vectorizer.fit_transform (X_train)
    test_vectors = vectorizer.transform (X_test)
    #print(train_vectors.shape, test_vectors.shape)
    #print(train_vectors)
```

• TF-IDF modeli çıktısının Excel'de görünümü;

4	Α	В	С	D	E	F	G	Н	1	J	K	L	M	N	0	Р
1		0														
2	0	(0, 173)0.	.127718597	05316055	(0, 518)0.2	9417375396	5040213 (0	0, 341)0.230	542491541	88752 (0,	325)0.2075	8492914466	5725 (0, 15	82)0.27357	9070573033	366 (0, 113
3	1	(0, 303)0.	.250597887	95397264	(0, 2168)0.	2694625772	2342791 (0	0, 197)0.269	462577234	2791 (0, 1	064)0.2372	1317790175	842 (0, 16	36)0.26946	2577234279	91 (0, 344)
4	2	(0, 1634)	0.17571433	66124713	(0, 232)0.1	757143366	124713 (0,	1142)0.163	412827451	8599 (0, 2	053)0.1634	1282745185	99 (0, 176	66)0.175714	3366124713	3 (0, 516)0.
5	3	(0, 137)0.	.211933859	19226876	(0, 2348)0.	.2341396588	3782912 (0	0, 1176)0.18	180246654	408713 (0,	, 1109)0.25	1765393671	19826 (0, 2	2407)0.1818	0246654408	3713 (0, 16
6	4	(0, 1452)	0.16301059	665308285	(0, 170)0.	1419567355	5353732 (0	0, 687)0.201	170412343	8652 (0, 1	075)0.1693	4345591151	1015 (0, 69	94)0.201170	4123438652	2 (0, 1704)
7	5	(0, 2170)	0.20743273	70272419	(0, 2545)0.	.2074327370	0272419 (0	0, 738)0.207	432737027	2419 (0, 2	340)0.1746	1502488413	3946 (0, 22	213)0.20743	2737027241	19 (0, 874)
8	6	(0, 345)0.	.188419890	86244668	(0, 1084)0.	.2026038998	35138977	(0, 1180)0.1	.705501528	8897246 (	0, 468)0.20	2603899851	138977 (0,	60)0.20260	3899851389	977 (0, 360
9	7	(0, 240)0.	.459298522	2001522 (	0, 849)0.42	7143690174	13433 (0, 1	126)0.45929	852220015	22 (0, 120	7)0.258999	2148702577	73 (0, 1282	2)0.3721746	4175471354	4 (0, 1612)
10	8	(0, 113)0.	.212613094	39645446	(0, 430)0.1	.8716746608	3364873 (0	0, 1417)0.19	772835602	617458 (0,	, 2130)0.21	2613094396	645446 (0,	242)0.1977	2835602617	7458 (0, 48
11	9							0, 2567)0.23						•		
12	10							(0, 2521)0.1						,		
13	11				•			75)0.208402			•				•	
14	12							(0, 8)0.1914			•			•		
15	13							0, 1140)0.50			•					
16	14							(0, 1993)0.3								
17	15	,						0, 1616)0.27								
18	16							1857)0.274			· .					
19	17			•				0, 653)0.388			· .		- • •			
20	18							0, 327)0.257			•			•		
21	19							0, 1651)0.55								_
22	20							1910)0.200								
24	22							. 2552)0.305 (0, 1730)0.2						•		
25	23							. 1532)0.284								
26	24							(0, 1968)0.1						•		
27	25	1 1						0, 1983)0.21						• •		
28	26							991)0.3631			•			•		
29	27							0, 163)0.406			•			•		
30	28							(0, 1219)0.2						•		
31	29							622)0.2311								
		, , ,									,					

Tablo 3: TF-IDF yöntemiyle sayısallaştırılmış veri.

#### 3.2.2. BoW Modeli

BoW modelinin içerisinde bulunan **CountVectorizer** sınıfı verimiz içerisinde bulunan kelime koleksiyonunu terim matrisine dönüştürmektedir. Cümle içerisindeki kelimeleri sayar.

• BoW modelinin kullanımına ilişkin kodlar şu şekildedir;

```
    BoW_Vector = CountVectorizer (min_df=0., max_df=1.)
    BoW_Matrix = BoW_Vector.fit_transform (X_train)
    print (BoW_Matrix)
```

• BoW modeli çıktısının Excel'de görünümü;

4	Α	В	С	D	E	F	G	Н	I	J	K
1		0									
2	0			•							1164)1 (0, 11
3	1										590)1 (0, 15
4	2					-	-				(0, 1789)1 (0
5	3	(0, 358)1	(0, 167)1	(0, 2526)1	(0, 1409)1	(0, 1201)1	(0, 520)1 (	0, 2424)1 (	0, 1834)1	(0, 1039)1	(0, 1191)1 (0
6	4										185)1 (0, 252
7	5	(0, 167)1	(0, 2526)1	(0, 2129)1	(0, 1036)1	(0, 1148)1	(0, 2513)2	2 (0, 897)1	(0, 2575)1	(0, 926)1	(0, 1916)2 (0
8	6	(0, 235)1	(0, 1456)1	(0, 167)1	(0, 2526)1	(0, 1608)1	(0, 1452)1	(0, 894)1 (	0, 1816)1	(0, 1619)1	(0, 1989)1 (0
9	7	(0, 167)1	(0, 2526)1	(0, 733)1	(0, 1612)1	(0, 1282)1	(0, 1207)1	(0, 126)1 (	0, 849)1 (0	), 240)1	
10	8	(0, 358)1	(0, 167)1	(0, 2526)1	(0, 1382)1	(0, 509)1 (	0, 2100)1 (	0, 269)1 (0	, 2462)1 (0	), 672)1 (0,	1696)1 (0, 8
11	9	(0, 358)1	(0, 167)1	(0, 2526)1	(0, 2126)1	(0, 2553)1	(0, 567)1 (	0, 447)1 (0	, 1027)1 (0	), 1287)1 (0	), 1767)1 (0,
12	10	(0, 37)1 (	0, 1456)1	(0, 167)1 (	0, 2526)1 (	0, 1176)1 (	0, 551)1 (0	, 242)1 (0,	2322)1 (0,	2317)1 (0,	187)1 (0, 69
13	11	(0, 358)1	(0, 167)3	(0, 2526)2	(0, 2126)1	(0, 2317)1	(0, 2539)1	(0, 773)1 (	0, 1011)1	(0, 1554)1	(0, 48)1 (0, 2
14	12	(0, 205)1	(0, 1172)1	(0, 167)1	(0, 2526)1	(0, 894)1 (	0,551)2 (0	, 2321)1 (0	, 1001)1 (0	, 820)1 (0,	142)1 (0, 15
15	13	(0, 173)1	(0, 1628)1	(0, 1140)1	(0, 1137)1	(0, 2272)1	L				
16	14	(0, 167)1	(0, 2526)1	(0, 474)1	(0, 942)1 (	0, 278)1 (0	, 2523)1 (0	, 1240)1 (0	, 1993)1 (0	), 1114)1 (0	), 224)1
17	15	(0, 358)1	(0, 167)1	(0, 2526)1	(0, 1814)1	(0, 983)1 (	0,458)1 (0	, 251)1 (0,	2385)1 (0,	61)1 (0, 94	13)1 (0, 146)
18	16	(0, 167)1	(0, 2526)1	(0, 1795)1	(0, 487)1	(0, 577)1 (	0, 311)1 (0	, 816)1 (0,	1585)1 (0,	479)1 (0,7	77)1 (0, 1842
19	17	(0, 167)1	(0, 1608)1	(0, 1520)1	(0, 748)1	(0, 1803)1	(0, 1656)1	(0, 818)1 (	0, 653)1 (0	), 2578)1 (0	), 2020)1
20	18	(0, 198)1	(0, 167)1	(0, 2526)1	(0, 2020)1	(0, 2048)1	(0, 283)1 (	0, 2140)1 (	0, 402)1 (0	), 914)1 (0,	1664)1 (0, 1
21	19	(0, 167)1	(0, 2526)1	(0, 353)1	(0, 2020)1	(0, 1651)2	(0, 2111)1	(0, 1462)1			
22	20	(0, 1808)	1 (0, 325)1	(0, 2491)1	(0, 167)1	(0, 2526)1	(0, 1525)1	(0, 98)1 (0	, 1322)1 (0	), 452)1 (0,	749)1 (0, 11
23	21	(0, 173)1	(0, 2097)1	(0, 1296)1	(0, 2105)1	(0, 853)1	(0, 471)1 (	0, 34)1 (0,	2552)1 (0,	981)1 (0, 2	2430)1
24	22	(0, 173)2	(0, 167)1	(0, 1651)1	(0, 1879)1	(0, 456)1 (	0, 1516)1 (	0, 1954)1 (	0, 2051)1	(0, 2127)1	(0, 1543)1 (0
25	23	(0, 2526):	1 (0, 1651)	1 (0, 1534)	1 (0, 1118)	1 (0, 1460)	1 (0, 180)	1 (0, 1308)1	l (0, 1117):	1 (0, 1532)	1 (0, 1642)1
26	24	(0, 358)2	(0, 167)3	(0, 2526)2	(0, 203)1 (	0, 2047)2 (	0, 2534)1 (	0, 1112)1 (	0, 2207)1	(0, 1076)1	(0, 1995)1 (0
27	25	(0, 173)1	(0, 167)4	(0, 2526)2	(0, 353)1 (	0, 1579)1 (	0, 2126)2 (	0, 474)1 (0	, 1240)1 (0	), 2532)1 (0	0, 2023)1 (0,
28	26	(0, 167)1	(0, 2526)1	(0, 1651)1	(0, 602)1	(0, 1792)1	(0, 165)1 (	0, 1294)1 (	0, 991)1 (0	), 961)1 (0,	1145)1
29	27	(0, 173)1	(0, 557)1	(0, 2552)1	(0, 133)1 (	0, 1761)1 (	0, 163)1 (0	, 1435)1 (0	, 2330)1		
30	28	(0, 358)1	(0, 1172)1	(0, 170)1	(0, 1207)1	(0, 2126)1	(0, 2322)1	(0, 1376)1	(0, 1187)1	(0, 2054)2	(0, 642)1 (0
31	29	(0, 358)3	(0, 2427)1	(0, 167)2	(0, 2526)2	(0, 170)1 (	0, 2048)1 (	0, 2054)1 (	0, 294)1 (0	), 1331)1 (0	), 1049)1 (0,
32	30	(0, 173)1	(0, 167)1	(0, 2526)1	(0, 1653)1	(0, 1032)2	(0, 1011)1	(0, 2097)1	(0, 853)1	(0, 1430)1	(0, 564)1 (0,

Tablo 4: BoW yöntemiyle sayısallaştırılmış veri.

#### 3.3. Modelin Eğitilmesi

Modelin eğitilmesi için 5 farklı yöntem kullanılmıştır. Bu yöntemler sırasıyla Naive Bayes, Logistic Regression, Decision Tree, RandomForest, Gradient Boosting'dir.

• Model eğitilirken kullanılan yöntemlerin kodları ve çıktıları aşağıdaki gibidir.

```
#NaiveBaves
clf = MultinomialNB ()
clf.fit (train_vectors, y_train)
prediction = clf.predict (test vectors)
# Logistic Regression
LogicReg = LogisticRegression ()
LogicReg.fit (train vectors, y train)
prediction = LogicReg.predict (test_vectors)
#DecisionTree
dTmodel = DecisionTreeClassifier ()
dTmodel.fit (train_vectors, y_train)
prediction = dTmodel.predict (test vectors)
#RandomForest
rForest = RandomForestClassifier ()
rForest.fit (train vectors, y train)
prediction = rForest.predict (test vectors)
#GradientBoosting
 grBoosting = GradientBoostingClassifier ()
 grBoosting.fit (train vectors, y train)
prediction = grBoosting.predict (test vectors)
```

• 5 farklı yöntemle eğitilmiş modelin çıktısı;

```
Naive Bayes ::
                         RandomForest ::
 [[9 1 1]
                          [[6 4 1]
 [1 7 1]
                          [1 6 2]
                           [2 4 4]]
 [3 1 6]]
0.7333333333333333
                         0.5333333333333333
Logistic Regression ::
                         GradientBoosting ::
 [[9 1 1]
                          [[8 2 1]
 [1 6 2]
                           [3 5 1]
 [3 1 6]]
                           [5 1 4]]
                          0.566666666666667
0.7
DecisionTree ::
 [[7 3 1]
 [4 3 2]
 [4 3 3]]
0.4333333333333333
```

Tablo 5: 5 farklı yöntemle eğitilen modellerin çıktısı.

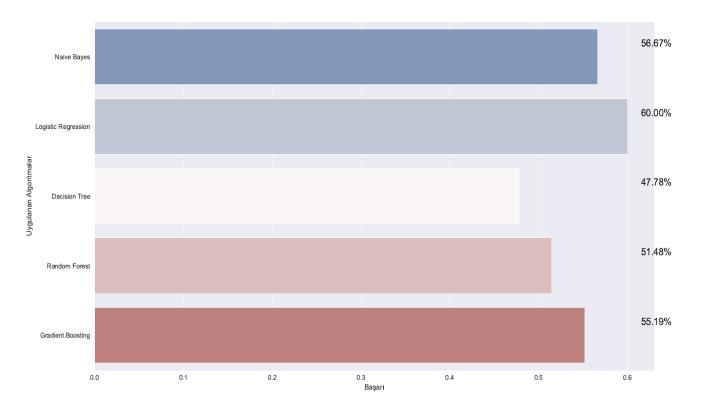
#### 3.4. Cross-validation

Bir önceki çalışmamızda gerçekleştirdiğimiz model eğitilmesinin bir adım öteye götürüp Crossvalidation yapıyoruz. Burada her yöntem 5 kez farklı verilerle çalışarak bir ortalama çıkarmakta.

• Cross-validation uygulamasının kodları aşağıdaki gibidir.

```
scores = cross_val_score (clf, train_vectors, y_train, cv=5)
scores2 = cross_val_score (LogicReg, train_vectors, y_train, cv=5)
scores3 = cross_val_score (dTmodel, train_vectors, y_train, cv=5)
scores4 = cross_val_score (rForest, train_vectors, y_train, cv=5)
scores5 = cross_val_score (grBoosting, train_vectors, y_train, cv=5)
methods = ["Naive Bayes", "Logistic Regression", "Decision Tree", "Random Forest ", "Gradient Boosting"]
accuracy = [scores.mean (), scores2.mean (), scores3.mean (), scores4.mean (), scores5.mean ()]
```

• Cross-validation yapıldıktan sonra modelin farklı yöntemlerle eğitim yüzdeleri aşağıdaki grafikte gösterilmiştir.



Grafik 2: Cross-validation çıktısının grafiği.

# 3.5. Çekilen Verinin Üzerinde Modelin Uygulanması

Çektiğimiz veriyi temizledikten sonra makine öğrenme seti oluşturuyoruz ve oluşturduğumuz makine öğrenme setini sayıya çevirip eğitiyoruz. Eğittiğimiz bu modeli de yine sayısal değere dönüştürdüğümüz temiz verimiz üzerine uyguluyoruz.

• Naive Bayes ve Logistic Regresyon ile tahmin ediyoruz. Bu tahmin kodları aşağıdaki gibidir;

```
    #Naive Bayes İle Tahmin
    predicted = clf.predict(test_vectors_)
    tahmin = pd.DataFrame(predicted)
    tahmin.rename(columns = {0:'tahmin'}, inplace = True)
    df["NaiveBayesTahmini"] = tahmin
    #Logistic Regresyon ile Tahmin
    predicted = LogicReg.predict(test_vectors_)
    tahmin = pd.DataFrame(predicted)
    tahmin.rename(columns = {0:'tahmin'}, inplace = True)
    df["LogisticTahmini"] = tahmin
```

• İki şekilde tahmin ettiğimiz verimizin tahminlerinin karşılaştırmasının Excel dosyasından ufak bir kesiti;

	А	В	С	D	E
1	Tweet	NaiveBayesTahmini	NBkategoriTahminleri	LogisticTahmini	LGkategoriTahminleri
2	olm gelsenize türkiye güzel oldu artık tüp kuyrukları yok asgari ücret o	1	Ekonomi	1	Ekonomi
3	asgari ücret insan onuruna yaraşır bir düzeye yükseltilmeli	2	Toplumsal	2	Toplumsal
4	taksiciler bile istediği zammı alıyor bizim asgari ücret işçi sendikası ye	1	Ekonomi	1	Ekonomi
5	asgariucret tahminim asgari ücret tl olur yine geçen yıl yaptıklarını jes	1	Ekonomi	1	Ekonomi
6	asgari ücret isterseniz bin olsun enflasyonla bi boka yaramaz	1	Ekonomi	1	Ekonomi
7	erdoğandan asgari ücret açıklaması	1	Ekonomi	1	Ekonomi
8	gt yap beraber takipçi kasalım yoruma çökün beraber büyütelim hesa	1	Ekonomi	1	Ekonomi
9	asgari ücret olmuş bir kırtık bişe maaşı alınca kalmıyor neşe senin üstü	1	Ekonomi	1	Ekonomi
10	disk çukurova bölge temsilciliği olarak adanadan seslendik sefalet üci	2	Toplumsal	2	Toplumsal
11	gt yap beraber takipçi kasalım yoruma çökün beraber büyütelim hesa	1	Ekonomi	1	Ekonomi
12	tuik asgari ücret asgari ücret artış oranı belirlemiyoruzaütk tarafından	0	Siyasi	0	Siyasi
13	bitane montum var tam yıllık gören diyer ola kılıg durmadan mama isc	1	Ekonomi	1	Ekonomi
14	liste taburespor iddaa fenerinmacivar asgariücret turkcell yanındayıze	2	Toplumsal	2	Toplumsal
15	liste taburespor iddaa fenerinmacivar asgariücret turkcell yanındayıze	2	Toplumsal	2	Toplumsal
16	erdoğandan asgari ücret çıkışı	1	Ekonomi	1	Ekonomi
17	günde yoksun yarında olma yarında gelsen yanında olmam asgariuecr	2	Toplumsal	2	Toplumsal
18	asgari ücret oldu oldu fark etmez sen senede yada e katlanmış temel	1	Ekonomi	1	Ekonomi
19	liste taburespor iddaa fenerinmacivar yanındayızerolbulut asgariücre	2	Toplumsal	2	Toplumsal
20	asgari ücret	1	Ekonomi	1	Ekonomi
21	disk bin liralık asgari ücret talebini yineledi	0	Siyasi	1	Ekonomi
22	sürpriz olmazsa asgari ücret tl dolaylarında olur	1	Ekonomi	1	Ekonomi
23	asgariücret zam yapıyorsunuz gün olmadan ulaşıma yiyecek içeçek giy	0	Siyasi	0	Siyasi
24	asgari ücret yükseltisin sözüne bile muhalefet edenler saraydamı yaş	2	Toplumsal	1	Ekonomi
25	asgari değil insanca bir yaşam istiyoruz duvar gazetemiz emekçilerle b	2	Toplumsal	2	Toplumsal
26	quote ww eczyt rysas thyao yanındayızerolbulut asgariuecret sınavage	2	Toplumsal	2	Toplumsal
27	asgari ücret kadar olacak asgariücret detaylar videomuzda	1	Ekonomi	1	Ekonomi
28	disk bin liralık asgari ücret talebini yineledi	0	Siyasi	1	Ekonomi
29	kalbi zengin insan ormanamama bir site keşfettim ülkemizin bir köşe	1	Ekonomi	2	Toplumsal
30	quote www eczyt rysas thyao yanındayızerolbulut asgariuecret sınavage	1	Ekonomi	2	Toplumsal
31	bir gazeteciyi dava etmişsin bin tl maaş olan ongun maaşını em kötü b	1	Ekonomi	0	Siyasi
32	disk bin liralık asgari ücret talebini yineledi	0	Siyasi	1	Ekonomi

Tablo 6: Naive Bayes ve Logistic Regresyon tahminlerinin karşılaştırması.

## 4. DUYGU ANALİZİ

Son olarak modelleri uyguladığımız veri setimizin "Turkish Bert" yöntemi ile duygu analizini gerçekleştiriyoruz.

• Son aşamaya getirdiğimiz verimizin Turkish Bert duygu analizi kodları aşağıdaki gibidir;

```
model = AutoModelForSequenceClassification.from_pretrained("savasy/bert-base-
   turkish-sentiment-cased")
 tokenizer = AutoTokenizer.from_pretrained("savasy/bert-base-turkish-sentiment-
 sa= pipeline("sentiment-analysis", tokenizer = tokenizer, model = model)
   sentiment list = []
   for i in dataTEST["Tweet"]:
       sentiment_list.append(sa(i))
   spredict list = []
   for i in range(0, len(sentiment list)):
       spredict_list.append(sentiment_list[i][0])
   spredict_list = pd.DataFrame(spredict_list)
   spredict list.head()
   dataTEST["label"] = spredict_list["label"]
  dataTEST["score"] = spredict list["score"]
  dataTEST.groupby("label").size()
   train_df["sentiment"] = dataTEST["label"]
   duygu_analizi = train_df.head(100)
```

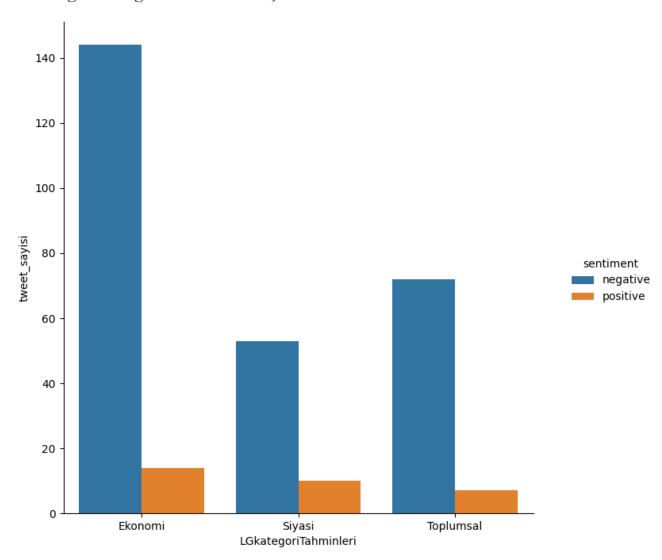
• Modelleri uyguladığımız test verisinin kategori tahminlerini ve duygularını içeren Excel çıktısının bir kısmı aşağıdaki gibidir;

	A	В	С	D	E	F
1	Tweet	NaiveBayesTahmini	NBkategoriTahminleri	LogisticTahmini	LGkategoriTahminleri	Duygu
2	olm gelsenize türkiye güzel oldu artık tüp kuyrukları yok asgari ücret olmuş direniyorsu	2	Toplumsal	2	Toplumsal	negative
3	asgari ücret insan onuruna yaraşır bir düzeye yükseltilmeli	2	Toplumsal	2	Toplumsal	negative
4	taksiciler bile istediği zammı alıyor bizim asgari ücret işçi sendikası yerine taksiciler gits	2	Toplumsal	2	Toplumsal	negative
5	asgariucret tahminim asgari ücret tl olur yine geçen yıl yaptıklarını jest yaparlarsa acayip	1	Ekonomi	1	Ekonomi	negative
6	asgari ücret isterseniz bin olsun enflasyonla bi boka yaramaz	1	Ekonomi	1	Ekonomi	negative
7	erdoğandan asgari ücret açıklaması	1	Ekonomi	1	Ekonomi	negative
8	gt yap beraber takipçi kasalım yoruma çökün beraber büyütelim hesapları takip takiples	1	Ekonomi	2	Toplumsal	negative
9	asgari ücret olmuş bir kırtık bişe maaşı alınca kalmıyor neşe senin üstüne toprak atalar a	1	Ekonomi	1	Ekonomi	negative
10	disk çukurova bölge temsilciliği olarak adanadan seslendik sefalet ücreti değil insan on	2	Toplumsal	2	Toplumsal	negative
11	gt yap beraber takipçi kasalım yoruma çökün beraber büyütelim hesapları takip takiples	1	Ekonomi	2	Toplumsal	negative
12	tuik asgari ücret asgari ücret artış oranı belirlemiyoruzaütk tarafından belirleniyorbirgür	0	Siyasi	0	Siyasi	negative
13	bitane montum var tam yıllık gören diyer ola kılıg durmadan mama isder evdeki çoluk se	1	Ekonomi	1	Ekonomi	negative
14	liste taburespor iddaa fenerinmacivar asgariücret turkcell yanındayızerolbulut	2	Toplumsal	2	Toplumsal	negative
15	liste taburespor iddaa fenerinmacivar asgariücret turkcell yanındayızerolbulut	2	Toplumsal	2	Toplumsal	positive
16	erdoğandan asgari ücret çıkışı	1	Ekonomi	1	Ekonomi	negative
17	günde yoksun yarında olma yarında gelsen yanında olmam asgariuecret cengizholding n	2	Toplumsal	2	Toplumsal	positive
18	asgari ücret oldu oldu fark etmez sen senede yada e katlanmış temel gıda fiyatlarına bir	0	Siyasi	1	Ekonomi	negative
19	liste taburespor iddaa fenerinmacivar yanındayızerolbulut asgariücret turkcell	2	Toplumsal	2	Toplumsal	negative
20	asgari ücret	1	Ekonomi	1	Ekonomi	negative
21	disk bin liralık asgari ücret talebini yineledi	1	Ekonomi	1	Ekonomi	negative
22	sürpriz olmazsa asgari ücret tl dolaylarında olur	0	Siyasi	1	Ekonomi	negative
23	asgariücret zam yapıyorsunuz gün olmadan ulaşıma yiyecek içeçek giyim vb alanlarda za	0	Siyasi	0	Siyasi	negative
24	asgari ücret yükseltisin sözüne bile muhalefet edenler saraydamı yaşıyor lüks yaşamları	2	Toplumsal	1	Ekonomi	negative
25	asgari değil insanca bir yaşam istiyoruz duvar gazetemiz emekçilerle buluştu avcılar asg	2	Toplumsal	2	Toplumsal	negative
26	quote www eczyt rysas thyao yanındayızerolbulut asgariuecret sınavagelmiyorum doco x	1	Ekonomi	2	Toplumsal	negative
27	asgari ücret kadar olacak asgariücret detaylar videomuzda	1	Ekonomi	1	Ekonomi	negative
	disk bin liralık asgari ücret talebini yineledi	1	Ekonomi	1	Ekonomi	negative
29	kalbi zengin insan ormanamama bir site keşfettim ülkemizin bir köşesine seçtiğimiz şel	2	Toplumsal	2	Toplumsal	negative
30	quote www eczyt rysas thyao yanındayızerolbulut asgariuecret sınavagelmiyorum doco x	1	Ekonomi	1	Ekonomi	negative
31	bir gazeteciyi dava etmişsin bin tl maaş olan ongun maaşını em kötü bir bin liraya düşür	1	Ekonomi	1	Ekonomi	negative
32	disk bin liralık asgari ücret talebini yineledi	1	Ekonomi	1	Ekonomi	negative

Tablo 7: Duygu analizi sonrası tahmin edilmiş verimizin son hali.

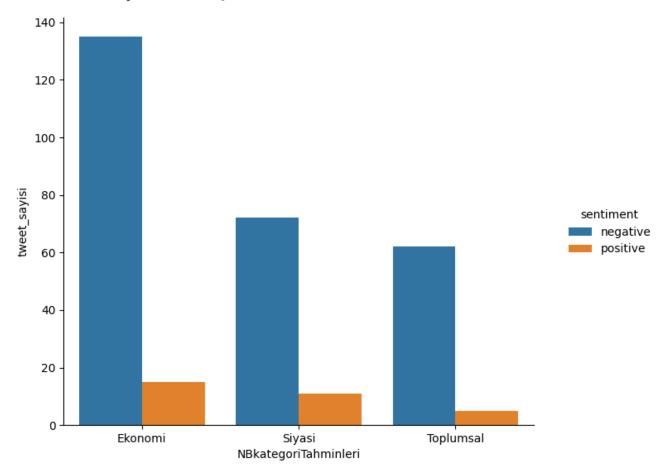
# 5. SONUÇLAARIN GÖRSELLEŞTİRİLMESİ

# 5.1. Logistic Regression Görselleştirmesi



Grafik 3: Logistic Regression Görselleştirmesi.

# 5.2. Naivee Bayes Görselleştirme



Grafik 4: Naivee Bayes Görselleştirmesi.