



VERİ MADENCİLİĞİ

DÖNEM SONU PROJE ÖDEVİ

SHELL YAKIT İSTASYONU METİN MADENCİLİĞİ

TUNAHAN KARASAKAL

YÖNETİM BİLİŞİM SİSTEMLERİ

DOÇ. DR. NUR KUBAN TORUN

İÇİNDEKİLER

1. GİRİŞ	/
1.1 Kullanılan Youtube Videosu	/
1.2 Python ile Veri Çekme	/
2. KULLANILAN PAKETLER VE R PROGRAMLAMA İŞLEMLERİ	/
2.1 Duygu Analizi	/
2.2 Polarity Analizi	/

1.Giriş

Günümüzde dijital platformlar, tüketici davranışlarını ve tercihlerini anlamada önemli bir veri kaynağı haline gelmiştir. YouTube gibi popüler video paylaşım platformlarında kullanıcıların paylaştığı yorumlar, belirli bir marka veya ürün hakkında önemli geri bildirimler sunar. Shell yakıt istasyonları gibi geniş bir müşteri kitlesine hitap eden bir marka için, bu yorumlar müşteri memnuniyetini ve beklentilerini anlamada kritik bir rol oynar. Bu çalışmada, YouTube videolarında Shell yakıt istasyonlarıyla ilgili yapılan yorumlar üzerinde metin madenciliği teknikleri uygulanarak, kullanıcıların algıları, ihtiyaçları ve memnuniyet seviyeleri analiz edilmektedir. Bu sayede, müşteri geri bildirimlerinden faydalanarak hizmet kalitesinin artırılmasına yönelik içgörüler elde edilmesi hedeflenmektedir.

1.1 KULLANILAN YOUTUBE VİDEOSU

<https://www.youtube.com/watch?v=8ulsbRaArvk>

1.2 PYTHON İLE VERİ ÇEKME

```
1 import requests
2 import csv
3
4 # API Anahtarınız ve Video ID'si
5 API_KEY = "AIzaSyAsh3rmhZilvhTC-1_CuHQdFSji_MC0dUw"
6 VIDEO_ID = "8uIsbRaArvk" # Çekmek istediğiniz video ID'sini buraya yazın.
7 BASE_URL = "https://www.googleapis.com/youtube/v3/commentThreads"
8
9 # Yorumları çeken fonksiyon
10 def get_video_comments(video_id, api_key, max_results=50):
11     comments = []
12     url = f"{BASE_URL}?part=snippet&videoId={video_id}&maxResults={max_results}&key={api_key}"
13
14     while url:
15         response = requests.get(url)
16         if response.status_code != 200:
17             print(f"API hatası: {response.status_code}, {response.text}")
18             break
19         data = response.json()
20
21         for item in data.get("items", []):
22             comment = item["snippet"]["topLevelComment"]["snippet"]["textDisplay"]
23             author = item["snippet"]["topLevelComment"]["snippet"]["authorDisplayName"]
24             published_at = item["snippet"]["topLevelComment"]["snippet"]["publishedAt"]
25             comments.append({"author": author, "comment": comment, "published_at": published_at})
26
27         # Next page token ile sayfalar arası geçiş
28         url = data.get("nextPageToken", None)
29         if url:
30             url = f"{BASE_URL}?part=snippet&videoId={video_id}&maxResults={max_results}&key={api_key}&pageToken={data['nextPageToken']}"
31
32     return comments
33
34 # Verileri CSV dosyasına yazma fonksiyonu
35 def write_to_csv(comments, filename='youtube_comments.csv'):
36     # Bu noktada header'ı düzgün bir şekilde tanımlayalım
37     header = ['author', 'comment', 'published_at'] # CSV başlıkları
38
39     with open(filename, mode='w', newline='', encoding='utf-8') as file:
40         writer = csv.DictWriter(file, fieldnames=header)
41
42         writer.writeheader()
43
44         for comment in comments:
45             writer.writerow(comment)
46
47     print(f"Yorumlar '{filename}' dosyasına yazıldı.")
48
49 # Yorumları çekme ve CSV dosyasına yazma
50 video_comments = get_video_comments(VIDEO_ID, API_KEY)
51 write_to_csv(video_comments)
```

Tablo.1

Çektiğimiz verilerimiz **tablo.1.1** ile gösterilmiştir.

author	comment	published_at
@jeffw8218	"It's crazy how all the comments are about how good of a test this was 🤔.Shell V-power is 93 octane, which means it will only make a difference compared to o	
@davidrogn3441	"I had a 1999 F350 with the 5.4 with nearly 250,000 miles on it. When I used the Shell v power premium fuel I got 2-3 miles per gallon better and the engine	
@bruceuhland8	"On my area, premium gas is usually \$1.00 to \$1.20 more a gallon. Too much to justify using it. I will put in some Techron in every few thousand miles instea	
@lennvilleserve4285	"Wondering if non ethanol 91-93 octane would do the same as Shell V-power 93? 🤔,2025-01-08T21:52:42Z	
@livinglegacy8429	"Do shell vs other top competitors like Mobile and Sunoco,2025-01-07T05:39:14Z	
@miloclanenont457	"I put cheap gas in my tank and my car ran like crap.,2025-01-06T00:44:49Z	
@Van1508	"Great video as always! The effort put in is greatly appreciated! God bless,2025-01-03T14:09:41Z	
@liveness2006	"Havnt got to look but have you tested top-tier diesel vs normal diesel? Wounding if its worth it for my diesel,2025-01-03T03:32:19Z	
@dlbsyst	"I know high octane gas is nonsense in cars that don't need it. What I want to know is which 87 octane gas is best. Which one offers the best performance, gas mi	
@Zapablast05	"Set video to 1.25x speed to match his talking pace of today's videos.,2024-12-25T07:25:01Z	
@Wayne6692	"2008 Impala 149K miles. Use shell v most times. Runs quiet, cannot hear it, and cleans system. With Lucas trans conditioner car runs like a champ on highway.",20	
@randallpetersen9164	"That music is a riff on Lazy Eye by Silversun Pickups. Actually more like a ripoff if I'm being honest.,2024-12-24T19:43:31Z	
@johnncallisterjr3913	"I like your generator cart set up nice..I wondering what is the best brand of gasoline on the us market i been using citgo for top tier gas..I have a 20	
@akesha4138	"Engines, esp v-8 larger engines last 50% longer when run on premium fuel. My older 96 Dodge Ram shows almost no engine wear after 300K.,2024-12-17T10:33:16Z	
@chucklucas8747	"Just bought a 2.0 Ecoboost Maverick i use shell middle grade hoping it will control carbon,2024-12-13T13:23:03Z	
@Ostromb232	"NICE.,2024-12-08T23:40:46Z	
@YouCantSeeMeENT	"Chevy spark 21 runs amazing on V-power great mileage! Just dont use the AC or Heat lol,2024-11-27T09:15:31Z	
@johnnd4348	"Oil Geek says using Top Tier gas actually reduces wear on valves and cylinder walls, thus improving MPG. He has the test results to prove it.",2024-11-27T02:29:58	
@johnnd4348	"Shell V Power has a better detergent additive package than cheap gas.,2024-11-27T02:21:55Z	
@VOLTAIRES_5	"Does that mean that I can get 87 octane even though my dash says premium fuel only?,2024-11-21T08:34:48Z	
@MunchiesNcrunchies	"ive always used Vpower maybe once every 4 months, cant say i regret it.",2024-11-20T23:14:11Z	
@Funkydood	"Greetings, Todd! Next time please pit Shell vs Texaco/Chevron!,2024-11-14T18:16:58Z	
@adamatter	"Go figure you I hadn't come across this video yet when I literally commented on your Techron one. I can say that when I've had my cars dyno tuned it seem	
@stephenanderson5630	"With the shell v power nitro I noticed about 2-3mpg gain even while towing a small trailer. Definitely better than cheap gas but you do pay a premium pri	

Tablo.1.1

2. KULLANILAN PAKETLER VE R PROGRAMLAMA İŞLEMLERİ

```
1 install.packages("tidyverse")
2 install.packages("lubridate")
3 install.packages("ggplot2")
4 install.packages("Amelia")
5 install.packages("wordcloud2")
6 install.packages("text")
7 install.packages("readxl")
8 library(tidyverse)
9 library(lubridate)
10 library(ggplot2)
11 library(Amelia)
12 library(wordcloud)
13 library(data.table)
14 library(dplyr)
15 library(wordcloud2)
16 library(stringi)
17 library(tm)
18 library(tidytext)
19 library(sentimentr)
20 library(syuzhet)
```

Tablo.2

Çektiğimiz verileri R üzerinden çağırıyoruz ve yapılması gereken temizleme işlemlerini yapıyoruz.(tablo.2.1)

```
#veri çekme ve temizleme bolumu
df <- fread("C:/Users/TUNAHAN/Desktop/notar/veri maden/youtubev3/youtube_comments.csv")
df <- df %>% select(-published_at)
df <- df %>%
  distinct(author, .keep_all = TRUE)

df$comment <- stri_encode(df$comment, to = "UTF-8")
```

Tablo.2.1

Emojilerimizi ayıklıyoruz ve frekanslarını hesaplıyoruz.(tablo.2.2)

```
# Emojileri ayıklama fonksiyonu
extract_emojis <- function(text) {
  regmatches(text, gregexpr("[\\p{So}\\p{Cn}]\\p{Emoji_Presentation}]+", text, perl = TRUE))
}

# Yorumlardan emojileri cikartma
all_emojis <- unlist(sapply(df$comment, extract_emojis))

# Emojilerin frekansini hesapla
emoji_freq <- as.data.frame(table(all_emojis))
colnames(emoji_freq) <- c("emoji", "frequency")
```

Tablo.2.2

Tekrar eden emojilerimiz ayıklayıp emoji bulutumuzu ayarlıyoruz.(tablo.2.3)

```
# Tekrar eden emojiler için her bir emojiden yalnızca birini al
unique_emoji_freq <- emoji_freq[!duplicated(emoji_freq$emoji),]

wordcloud2(emoji_freq, size = 1.3, color = 'random-light', backgroundColor = "red")
```

Tablo.2.3

Emoji bulutu(tablo.2.4)



Tablo.2.4

Bu kod, metin verisini temizlemek için bir dizi işlem uygular. İlk olarak, tüm yorumlar tek bir metin parçası olarak birleştirilir. Ardından, metinden gereksiz öğeler (büyük harfler, noktalama işaretleri, sayılar, yaygın kelimeler ve boşluklar) kaldırılır. Bu işlemler, kelime bulutu gibi analizler için veriyi daha uygun hale getirir.(tablo.2.5)

```
#kelime bulutu
all_reviews <- paste(df$comment, collapse = ' ')

# Metni temizleyelim
docs <- Corpus(VectorSource(all_reviews))
docs <- tm_map(docs, content_transformer(tolower)) %>%
  tm_map(removePunctuation) %>%
  tm_map(removeNumbers) %>%
  tm_map(removeWords, stopwords("en")) %>%
  tm_map(stripWhitespace)
```

Tablo.2.5

İstemediğimiz kelimeleri ve olmaması gereken kelimeleri ayırıyoruz.(tablo.2.6)

```
# Kendi özel stopwords listemizi ekleyebiliriz
custom_stopwords <- c("the", "and", "for", "but", "or", "so", "can", "just", "will",
docs <- tm_map(docs, removeWords, custom_stopwords)
```

Tablo.2.6

Kelimelerimizin frekanslarını hesaplıyoruz.(tablo.2.7)

```
# Kelimelerin frekanslarını hesaplama
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
word_freqs <- sort(rowSums(m), decreasing = TRUE)
word_freq_df <- data.frame(word = names(word_freqs), freq = word_freqs)
```

Tablo.2.7

Kelime bulutumuz için en çok kullanılan 100 kelimeyi bulmasını ve oluşturmasını istiyoruz.(tablo.2.8, tablo.2.9)

```
top_100_words <- head(word_freq_df, 100)
wordcloud(words = top_100_words$word,
  freq = top_100_words$freq,
  scale = c(3, 0.5), # Boyutlandırma
  colors = brewer.pal(8, "Dark2"))
```

Tablo.2.8



Tablo.2.9

2.1 DUYGU ANALİZİ

Duygu analizi işlemi, metin verisinde duygu analizi yaparak metnin duygusal içeriğini incelemeyi sağlar. İlk olarak, tüm temizlenmiş metin birleştirilir ve ardından `get_nrc_sentiment` fonksiyonu ile metindeki duygusal ifadeler (mutluluk, öfke, korku, vb.) analiz edilir. Analiz sonucunda her duygu kategorisi için bir değer elde edilir. Elde edilen duygusal veriler, bir veri çerçevesine dönüştürülür ve duygu kategorilerinin toplam frekansları hesaplanır. Bu sayede metnin genel duygusal tonunu anlamak mümkün olur. Bu tür bir duygu analizi, özellikle müşteri yorumları gibi verilerde kullanıcı algılarını anlamada faydalıdır.(tablo.3)

```
#duygu analizi
# Temizlenmiş tüm metni birleştiriyoruz
text_clean <- paste(sapply(docs, as.character), collapse = " ")
view(text_clean)
# Duygu analizi yapmak
sentiments <- get_nrc_sentiment(text_clean)

# Sonuçları veri çerçevesine dönüştürmek
sentiments_df <- as.data.frame(t(sentiments))
sentiments_df$sentiment <- rownames(sentiments_df)
rownames(sentiments_df) <- NULL

# Duyguları toplayalım
total_sentiments <- colSums(sentiments[, 1:8])

# Duygu kategorilerini bir veri çerçevesine yerleştirelim
total_sentiments_df <- data.frame(sentiment = names(total_sentiments), count = total_sentiments)
```

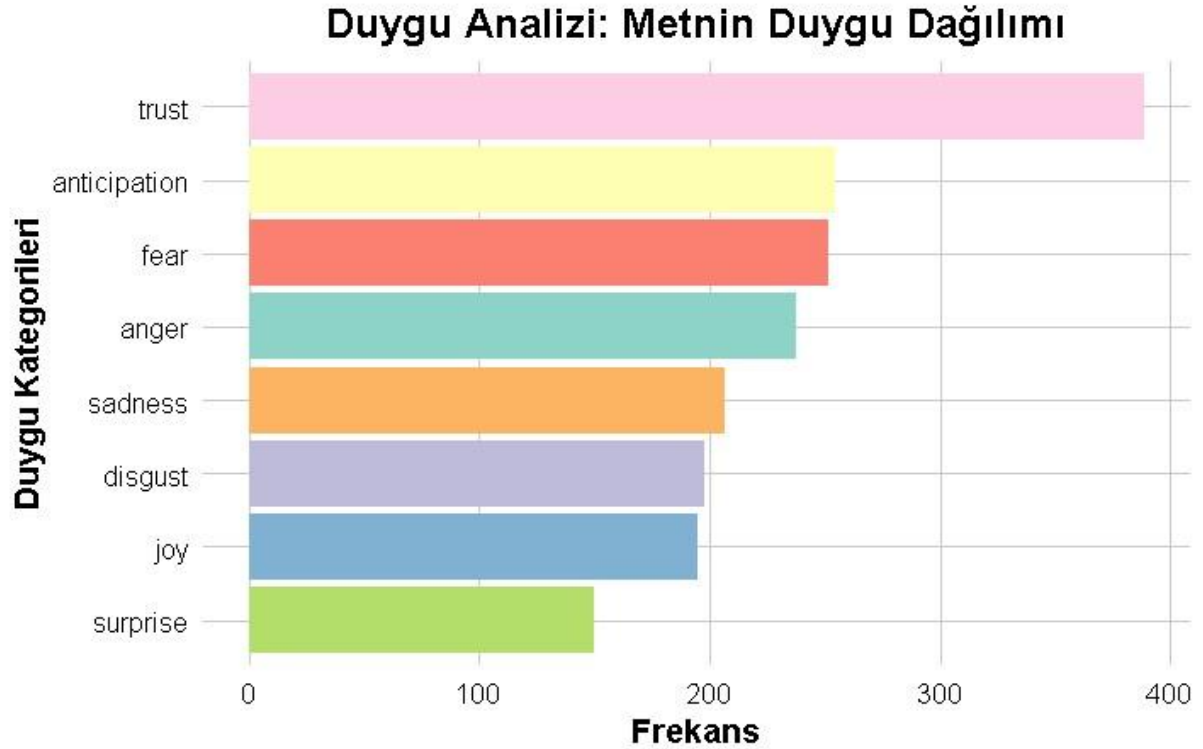

Tablo.3

Görselleştirmek için ggplot paketimizi kullanıyoruz(tablo.3.1)

```
ggplot(total_sentiments_df, aes(x=reorder(sentiment, count), y=count, fill = sentiment)) +
  geom_bar(stat="identity", show.legend = FALSE) +
  labs(
    title = "Duygu Analizi: Metnin Duygu Dağılımı",
    x = "Duygu Kategorileri",
    y = "Frekans"
  ) +
  coord_flip() +
  theme_minimal(base_size = 15) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "black", hjust = 0.5),
    axis.title = element_text(size = 14, face = "bold"),
    axis.text = element_text(size = 12, color = "black"),
    panel.grid.major = element_line(color = "grey80"),
    panel.grid.minor = element_blank(),
    plot.background = element_rect(fill = "white", color = "white")
  ) +
  scale_fill_brewer(palette = "Set3")
```

Tablo.3.1

Duygu analizi grafiklerimiz(tablo.3.2):



Tablo.3.2

2.2 POLARITY ANALİZİ

Polarity analizimiz için sentiment ile kelimelerimizi pozitif, negatif, nötr olarak ayırıyoruz. Ardından yeni bir sütun olarak ekleyip görselleştiriyoruz.(tablo.3.3)

```
#polarity
sentiment_result <- sentiment(df$comment)
sentiment_result$polarity <- ifelse(sentiment_result$sentiment > 0, "positive",
                                   ifelse(sentiment_result$sentiment < 0, "negative", "neutral"))

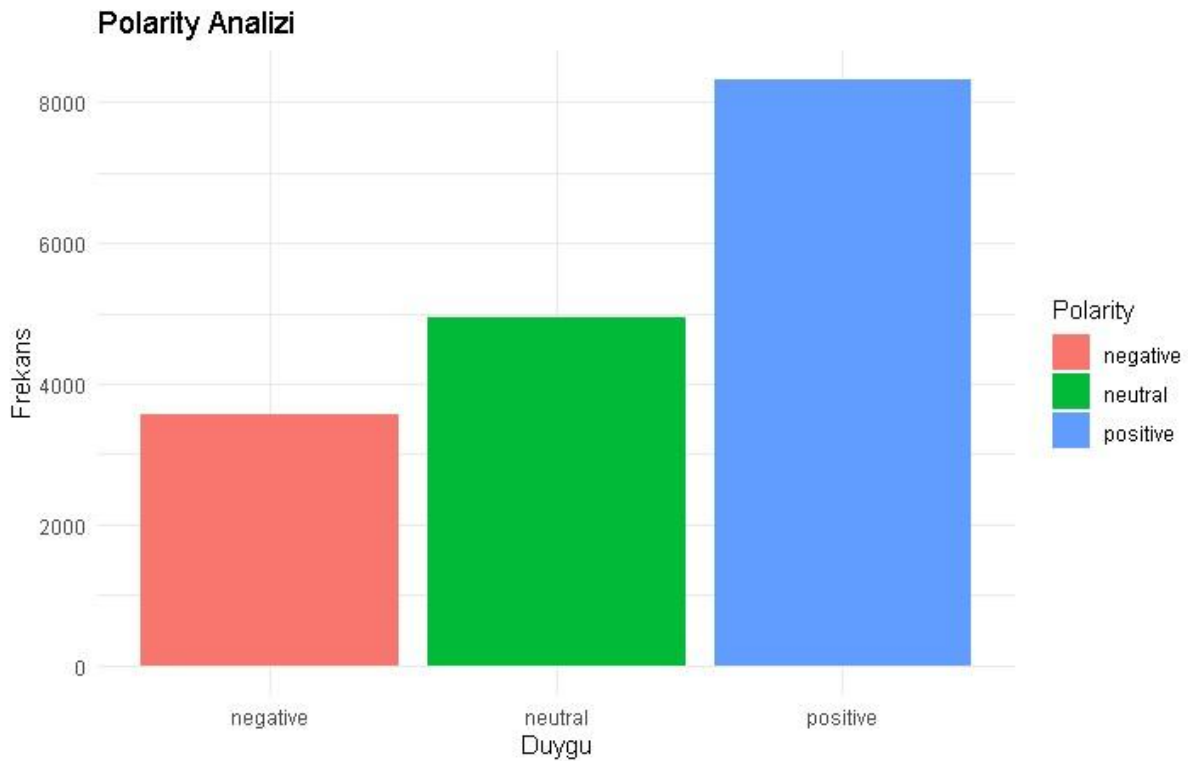
library(ggplot2)

# Polarity sütununu saymak
sentiment_counts <- as.data.frame(table(sentiment_result$polarity))
colnames(sentiment_counts) <- c("Polarity", "Count")

# Bar grafiği ile görselleştirme
ggplot(sentiment_counts, aes(x = Polarity, y = Count, fill = Polarity)) +
  geom_bar(stat = "identity") +
  labs(title = "Duygu Dağılımı", x = "Duygu", y = "Frekans") +
  theme_minimal()

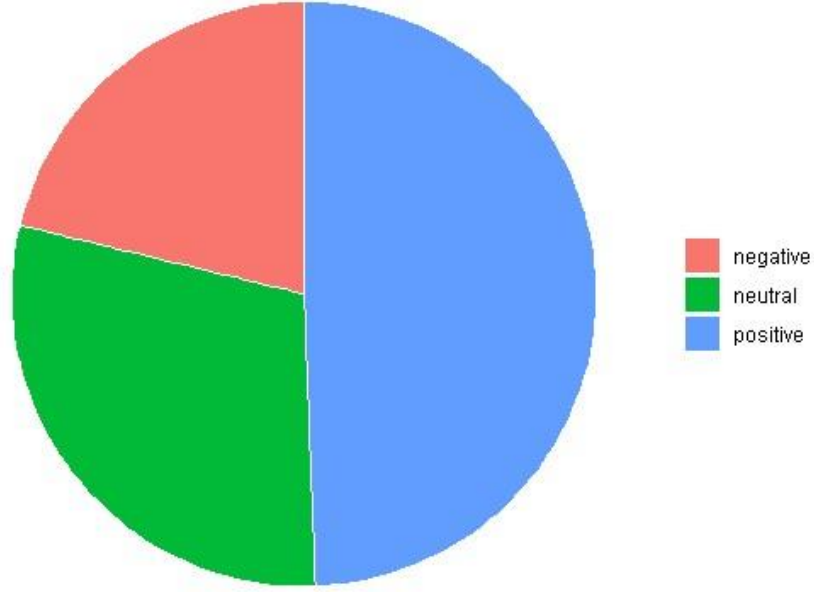
ggplot(sentiment_counts, aes(x = "", y = Count, fill = Polarity)) +
  geom_bar(width = 1, stat = "identity", color = "white") +
  coord_polar("y", start = 0) +
  labs(title = "Duygu Analizi Pasta Grafiği") +
  theme_void() +
  theme(legend.title = element_blank())
```

Tablo.3.3



Tablo.3.4

Polarity Analizi Pasta Grafiği



Tablo.3.5

TEŞEKKÜRLER

KAYNAKÇA

Torun, N. K. ve Şengül A., (2022). Kripto Para Birimlerinin Twitter Verileri ile Metin Madenciliği Kapsamında İncelenmesi. Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi, 6(1), 54-65

<https://medium.com/@seherkumsar/veri-biliminde-api-kullanımı-7b21aa28f3d9>

<https://www.newslabturkey.org/2019/10/16/r-ekosisteminde-metin-madenciliginasil-yapilir/>

<https://www-tidyttextmining-com.translate.goog/topicmodeling? x tr sl=en& x tr tl=tr& x tr hl=tr& x tr pto=tc>