



BİLECİK SEYH
EDEBALI
ÜNİVERSİTESİ

Veri Madenciliğine Giriş Final Sunumu

Hazırlayan: Alihan AKKUŞ

Ders: Veri Madenciliğine giriş

Danışman Hoca :Sayın Dr. Öğr. Üyesi Nur Kuban
TORUN



Kısaca

- Youtube'dan yayınlanan 2023 F1 yarışı özetinin videosunda insanların duygusal tepkilerinin analizi edilmiştir .R programlama dili kullanarak.
- Yapılana adımlar
 - 1.Ver çekimi
 - 2.Ver temizleme ve kaydetme
 - 3.veriyi işleme
 - 4.Veriyi görselleştirme



Kullanılan paketler

Burada
kullanılan
paketlerin nasıl
yüklendiği ve
nasıl aktif hale
getirildiği ek
olarak ne için
kullanılacakları
da yazmaktadır

```
install.packages("tuber")#youtubeden veri cekimi icin  
install.packages("tm") # metinleri duzenlemek icin  
install.packages("SnowballC") # metin koku icin  
install.packages("wordcloud") # kelime bulutu uretici  
install.packages("RColorBrewer") #renk paletleri  
install.packages("syuzhet") # duyarlılık analizi  
install.packages("ggplot2") # grifik cizmez icin
```

```
library("tuber")  
library("tm")  
library("SnowballC")  
library("wordcloud")  
library("RColorBrewer")  
library("syuzhet")  
library("ggplot2")
```



Veri çekme

Youtube'dan yorum çekme ve Kaydetme

- googledan geliştirici hesabı alıp,Tuber paketini kullanarak yorumları çekiyoruz.
- Tekrar eden yorumları temizleme işlemi yapıyoruz.
- Yorumları "Yorumlar.csv" şeklinde kaydediyoruz.
- Veri değişkenine kayıt ettiğimiz konumdan geri yükleme yapıyoruz

```
#vericekimi####  
#tuber paketi ile  
myclientid="*****"  
myclientsecret="*****"  
yt_oauth(myclientid,myclientsecret,token="")  
yorumlar=get_all_comments("2-jv1hc7c58")  
  
#tekrar eden yorumlari temizleme|  
yorumlar<-yorumlar[!( duplicated(yorumlar$textDisplay)),]  
#cekilen verilerin kayıt edilmesi  
dosya_yolu <- "yorumlar.csv"  
write.table(yorumlar$textDisplay, file = dosya_yolu, sep = "|"  
            quote = FALSE, row.names = FALSE)  
veri<-read.table(file.choose(),header = F,sep="|")
```



Veriyi işleme

Tm paketi ile veriyi kullanıma uygun hale getirme

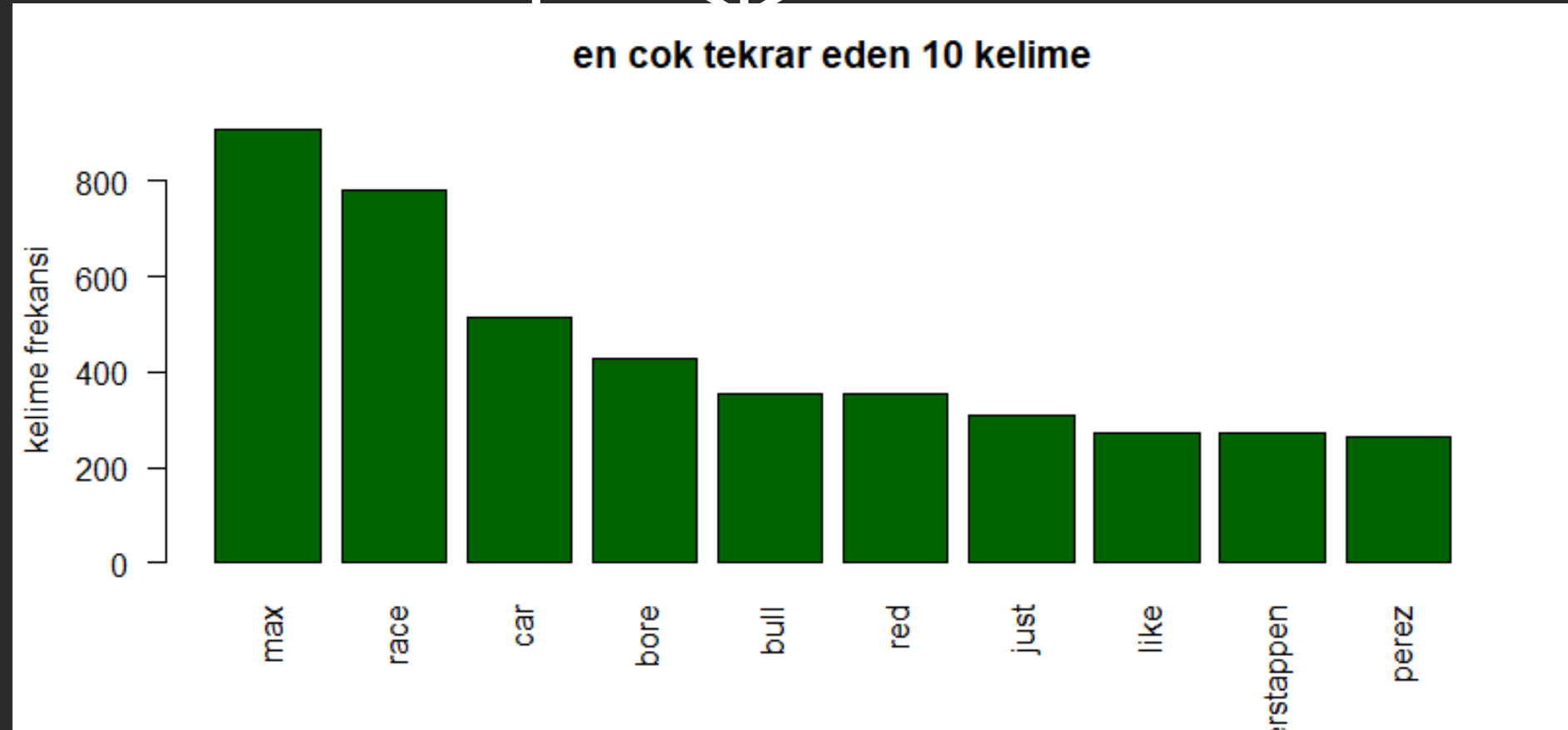
- Veri adlı değişkenimizi Corpus biçimine dönüştürüyoruz ve kveri (kullanılabilir veri) değişkenine atıyoruz.
- Sonrasında emojiler ,/','@','\\|' verimizden çıkartıyoruz.
- En son tümveriyi küçük harflere düştüme ,sayıları çıkartma vb. işlemlerini yapıyoruz

```
#veriyi isleme####  
#(tm paketini kullanarak)veri den gereksiz karakterlerin cikarilmasi islem  
kveri <- Corpus(VectorSource(veri))  
#emojilerin cikartilmasi  
temiz_yorum <- tm_map(kveri, content_transformer(function(x)  
    gsub("\\&\\#\\d+;", "", x)))  
toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ",  
kveri <- tm_map(kveri, toSpace, "/")  
kveri <- tm_map(kveri, toSpace, "@")  
kveri <- tm_map(kveri, toSpace, "\\|")  
#metini kucuk harflere donusturma  
kveri <- tm_map(kveri, content_transformer(tolower))  
#rakamlari cikartma  
kveri <- tm_map(kveri, removeNumbers)  
#ingilizce anlamsiz eklimerlin cikartilmasi  
kveri <- tm_map(kveri, removeWords, stopwords("english"))  
#noktalama isaretlerinin kaldırılması  
kveri <- tm_map(kveri, removePunctuation)  
#fazla boslukların kaldırılması  
kveri <- tm_map(kveri, stripWhitespace)  
#metinlerin köklerine inme  
kveri <- tm_map(kveri, stemDocument)
```

Veri Görselleştirme

- Temizlenmiş veri kullanarak en çok kullanılan kelimelerin bulunma ve sıralamsı yapılır
- En çok kullanılan kelimelerin ilk on tanesi seçilir
- Çubuk Grafiğinde gösterilir

```
57 #veri gorsellestirme####  
58 #terim belge matrisi oluşturma  
59 TextDoc_dtm <- TermDocumentMatrix(kveri)  
60 dtm_m <- as.matrix(TextDoc_dtm)  
61 #Frekansları azalan değere göre sıralama  
62 dtm_v <- sort(rowSums(dtm_m),decreasing=TRUE)  
63 dtm_d <- data.frame(word = names(dtm_v),freq=dtm_v)  
64  
65 #En çok tekrar eden 10 kelime  
66 head(dtm_d, 10)  
67 #En çok kullanılan kelimelerin sütun grafiği  
68 barplot(dtm_d[1:10,]$freq, las = 2, names.arg = dtm_d[1:10,]$word,  
69         col = "darkgreen", main = "en çok tekrar eden 10 kelime",  
70         ylab = "kelime frekansı")
```



- Word Cloud paketi ile minimum frekansı 80 olan ve maksimum 200 kelime olacak şekilde kelime bulutu oluşturma

A word cloud visualization of terms related to Formula 1 racing. The words are arranged in a circular pattern, with varying font sizes and colors. The most prominent words include "max", "race", "verstappen", "car", "bull", "checo", "alonso", "ferrari", "comment", "better", "watch", "need", "newmerced", "can", "bored", "red", "will", "best", "ever", "now", "back", "commentari", "drive", "sport", "year", "get", "drivers", "season", "lewis", "highlight", "still", "team", "great", "start", "drs", "fan", "give", "look", "know", "overtak", "croft", "hamilton", "leclerc", "much", "show", "lap", "way", "see", "well", "time", "even", "domin", "make", "think", "everi", "que", "like", "just", "one", "perez", "hard", "que". The colors range from dark blue to light green, with some words in red or orange. The overall shape is roughly circular, with words radiating from the center.



Kelime Bulut

- Syuzhet paketi ile kelimelerin duygu puanlaması yapılmaktadır.
- Üç farklı method kullanılmakta
 - a. Syuzhet methodu
 - en olumsuz (-1), en olumlu (+1) ile derecelendirir.
 - b. Bing methodu
 - -1 negatifi ve +1 pozitif duyguyu gösteren ikili ölçeklendirir
 - c. Affin methodu
 - -5 ile +5 arasında değişen tamsayı ölçeği kullanır
- En altta aynı kelimelere verdikleri dereceler bulunmaktadır

```
79 #syuzhet paketi ile duygu puanlama
80 #3 farklı method ile duygu analizi
81 #1.syuzhet methodu
82 syuzhet_vector <- get_sentiment(veri$V1, method="syuzhet")
83 #ilk bir kaci
84 head(syuzhet_vector)
85 # analizn ozeti
86 summary(syuzhet_vector)
87
88 #2.bing methodu
89 bing_vector <- get_sentiment(veri$V1, method="bing")
90 head(bing_vector)
91 summary(bing_vector)
92 #3.affin methodu
93 afinn_vector <- get_sentiment(veri$V1, method="afinn")
94 head(afinn_vector)
95 summary(afinn_vector)
96 #ucunun ayni anda hangi kelimeye ne deger verdigini görüntuleme
97 rbind(
98   sign(head(syuzhet_vector)),
99   sign(head(bing_vector)),
100   sign(head(afinn_vector))
101 )
102
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0	-1	0	0	0	0
[2,]	0	-1	-1	0	0	0
[3,]	0	-1	-1	0	0	0



Duygu analiiz

- Syuzhet paketi ile bütün verinin duygu analizi yapılmaktadır
- ve ilk 10 tanesinin görüntülenmektedir

```
103 #her satirin duygu analizi
104 d<-get_nrc_sentiment(veri$v1)|
105 #ilk 10 tanesinin görüntülenmesi
106 head (d,10)
```

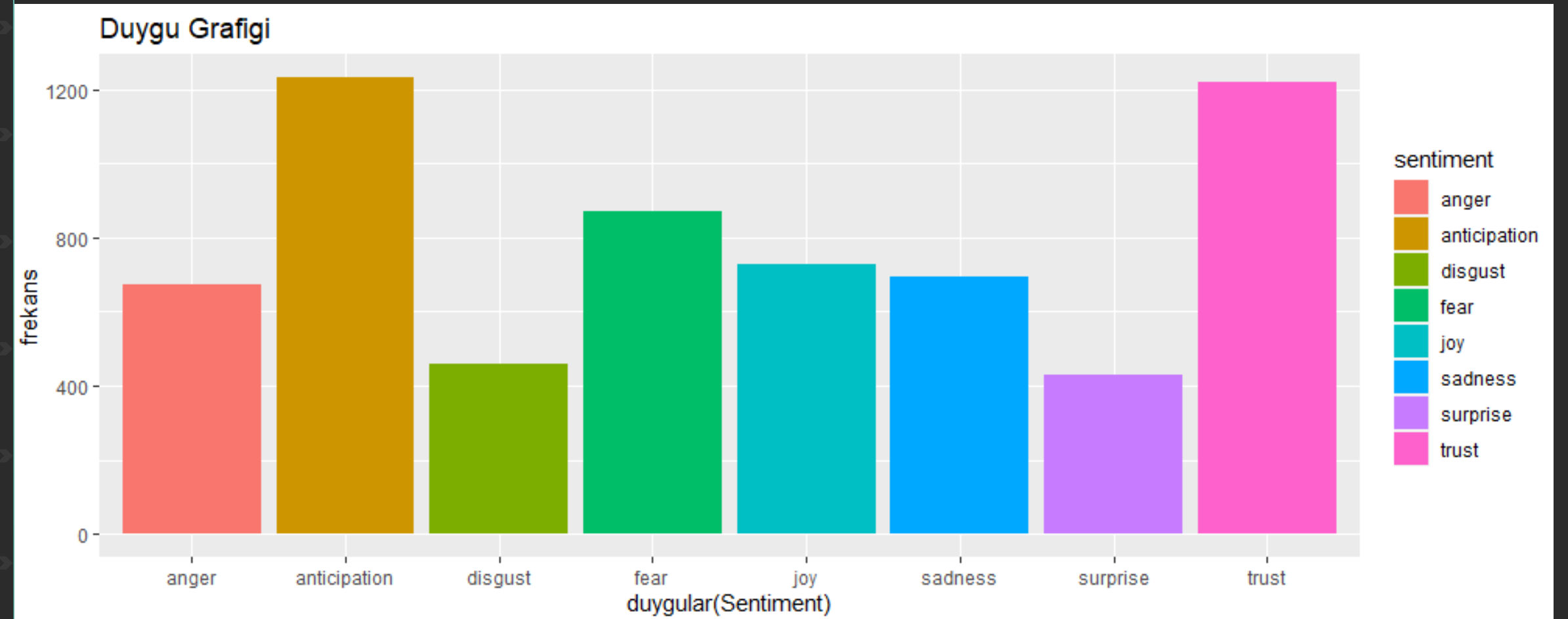
```
> head (d,10)
      anger anticipation disgust fear joy sadness surprise trust negative positive
1         0           0        0    0    0         0         0         0         0         0
2         0           0        0    0    0         0         0         0         0         0
3         0           0        0    0    0         0         0         0         0         0
4         0           0        0    0    0         0         0         0         0         0
5         0           0        0    0    0         0         0         0         0         0
6         0           0        0    0    0         0         0         0         0         0
7         2           0        2    0    0         1         0         0         3         3
8         0           0        0    0    0         0         0         0         0         0
9         0           0        0    0    0         0         0         0         0         0
10        0           0        0    0    0         0         0         0         0         0
```



Duygu Analizi Çubuk Grafiği

- sınıflandırılan kelimelerin duygu derecelerine göre toplanması
- Tüm yorumların Çubuk Grafiğinde dönüştürülmesi
- duygular
 - anger(Kızgın)
 - anticipation(beklenti)
 - disgust(iğrenme)
 - fear(korku)
 - joy(neşe)
 - sadness(hüzün)
 - surprise(şaşkın)
 - trust(güven)
 - negative(olumsuz)
 - positive(olumlu)

```
#cubuk grafigine donusumu  
#transpoze  
td<-data.frame(t(d))  
# rowSums ilevi,satir seklinde toplamiyi saglar  
td_new <- data.frame(rowSums(td[2:5022]))  
#donustume ve temizleme  
names(td_new)[1] <- "count"  
td_new <- cbind("sentiment" = rownames(td_new), td_new)  
rownames(td_new) <- NULL  
td_new2<-td_new[1:8,]  
#cubuk grafigi  
quickplot(sentiment, data=td_new2, weight=count, geom="bar", fill=sentiment,  
          xlab = "duygular(Sentiment)",ylab="frekans")+ggtitle("Duygu Grafigi")
```



Kaynakça

[HTTPS://WWW.RED-GATE.COM/SIMPLE-TALK/DATABASES/SQL-SERVER/BI-SQL-SERVER/TEXT-MINING-AND-SENTIMENT-ANALYSIS-WITH-R/](https://www.red-gate.com/simple-talk/databases/sql-server/bi-sql-server/text-mining-and-sentiment-analysis-with-r/)

[CHAT CPT](#)

[AİMRT YOUTUBEKANALI](#)

