T.C.

Bilecik Şeyh Edebali Üniversitesi
İktisadi ve İdari Bilimler Fakültesi
Yönetim Bilişim Sistemleri



Veri Madenciliği YouTube Veri ve Duygu Analizi

Mustafa ARAÇLI

İçindekiler

Ozet	3
Veri Madenciliği Nedir?	4
Seçilen Konu ve Kullanılan Videolar	5
Geliştirici Hesabı	7
YouTube Veri Çekme İşlemi	9
Verileri Temizleme	10
Kelime bulutu	14
Avatar	15
Duygu Analizi	15
Sonuç	19
Kaynakça	20

Özet

Veri madenciliği, veri analitik veya veri keşifte kullanılan bir yöntemdir. Veri madenciliği, veri tabanları, veri depolama sistemleri, veri dağıtım sistemleri veya internet gibi büyük veri kaynaklarından anlamlı bilgi veya trendler çıkarmanın yollarını arama işlemidir. Veri madenciliği süreci genellikle veri toplama, veri temizleme, veri dönüştürme, veri depolama, veri analizi ve veri görselleştirme adımlarından oluşur. Bu çalışmada bulut bilişim başlığı altında oyun sektörü için YouTube yorumlarından analiz yapılmıştır. Analizde toplam 14198 yorum filtrelenerek bunlardan 21881 tane kelime işleme alınmıştır. Kelime bulutu ve duygu analizi yöntemleri kullanılmıştır. Analiz için API bağlantısı Google Cloud Console dan alınmış olup editör olarak ta R Studio kullanılmıştır.

Veri Madenciliği Nedir?

Veri madenciliği, verileri analiz etme, keşfetme ve anlamlı bilgiler veya trendler çıkarmanın yollarını arama işlemidir. Bu bilgiler, pazarlama, finans, sağlık hizmetleri, üretim veya hükümet gibi farklı alanlarda kullanılabilir. Veri madenciliği, genellikle veri toplama, veri temizleme, veri dönüştürme, veri depolama, veri analizi ve veri görselleştirme adımlarından oluşur. Yapay zeka, makine öğrenimi ve istatistik gibi farklı yöntemleri kullanır.

Veri Madenciliği Süreci Nasıl İşler?

Veri madenciliği süreci genellikle aşağıdaki adımlardan oluşur:

Veri toplama: Veri madenciliği için kullanılacak veri kaynaklarını belirleme ve bu verileri toplama. Veri madenciliği, genellikle büyük veri kaynaklarından yararlanır, ancak veriler, veri tabanları, veri depolama sistemleri, veri dağıtım sistemleri veya internet gibi farklı kaynaklardan da toplanabilir.

Veri temizleme: Toplanan veriler içinde hata, eksik bilgi veya gereksiz bilgileri temizleme. Veri temizleme, veri madenciliği sürecinin önemli bir adımıdır çünkü veri madenciliği analizlerinde kullanılan verilerin kalitesi çok önemlidir.

Veri dönüştürme: Verileri, analiz için daha uygun hale dönüştürme. Veri dönüştürme, veri madenciliği sürecinde verilerin saklanması, analizi ve görselleştirilmesi için gerekli formatlara getirilmesi anlamına gelir.

Veri depolama: Verileri, analiz için kullanılabilir hale getirme ve depolama. Veri depolama, veri madenciliği sürecinin bir adımıdır, çünkü verilerin analiz edilmesi ve görselleştirilmesi için erişilebilir olması gerekir.

Veri analizi: Verileri analiz etme ve anlamlı bilgileri veya trendleri çıkarmanın yollarını arama. Veri analizi, veri madenciliği sürecinin en önemli adımlarından biridir çünkü verilerin anlamlı bilgiler veya trendler çıkarılması amacını taşır.

Veri görselleştirme: Anlamlı bilgileri veya trendleri, grafikler, diagramlar veya haritalar gibi farklı araçlarla görselleştirme. Veri görselleştirme, veri madenciliği sürecinin son adımıdır ve verilerin anlamlı bilgiler veya trendler çıkarılması sonucunda elde edilen bilgilerin kolayca anlaşılmasını sağlar.

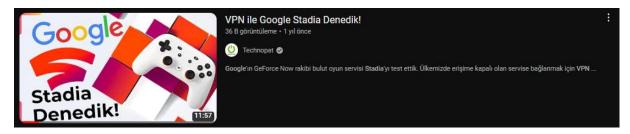
Seçilen Konu ve Kullanılan Videolar

Oyunlarda bulut bilişim

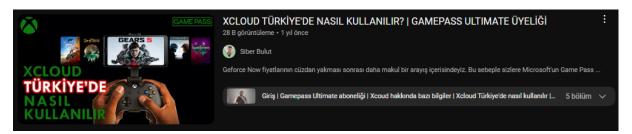
Bulut oyun, oyunların ve oyun hizmetlerinin internet üzerinden sunulmasıdır. Bu, oyunların ve oyun hizmetlerinin, oyuncuların cihazlarına yerleştirilmeden önce internet üzerinde çalıştırılmasını veya erişilmesini sağlar. Bulut oyunları, oyunların ve oyun hizmetlerinin internet üzerinden sunulması sayesinde oyunculara erişilebilirlik, stabilite, yüksek performans, güncelleme ve bakım ve maliyet azaltımı gibi avantajlar sunar. Bulut oyunları genellikle oyunların çalıştırılması için gerekli tüm donanım ve yazılımın internet üzerinde barındırıldığı oyun sunucularına bağlantı sağlar. Bu sayede oyuncular oyunların grafik, ses ve diğer özellikleri cihazlarının özelliklerine göre değil oyun sunucularının özelliklerine göre deneyimleyebilirler.



276 yorum



159 yorum



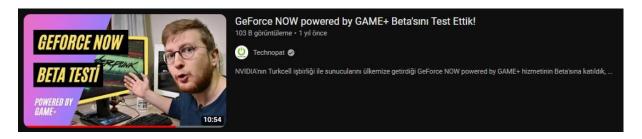
227 yorum



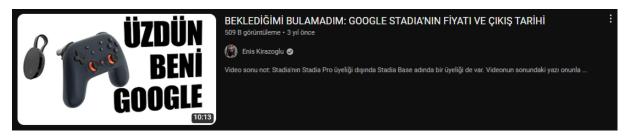
1047 yorum



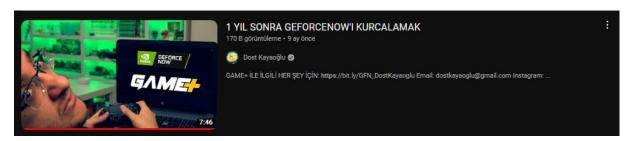
924 yorum



7068 yorum



2285 yorum



675 yorum

Toplam yorum miktarı = 14194 yorum,

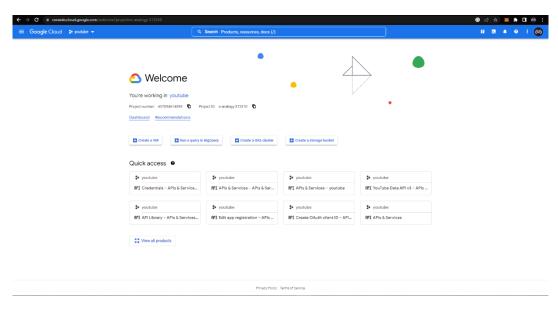
Toplam kullanılan kelime miktarı = 92461 kelime,

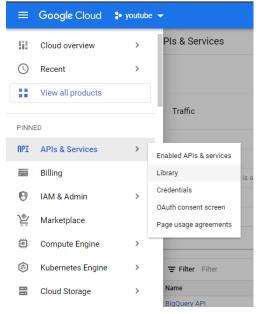
Toplam işlem gören kelime sayısı = 21881 kelime

Üzerinden değerlendirmeler yapılmıştır.

Geliştirici Hesabı

R Studio ile youtube'dan yapabilmek için öncelikle Google Cloud Console sayfasına gidip Google hesabımız ile giriş yapıyoruz. Eğer hesabımız yoksa basit bir şekilde yönergelere uyarak giriş yapıyoruz.

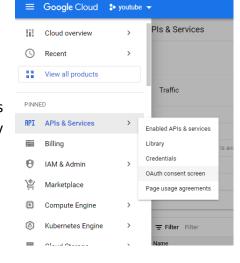




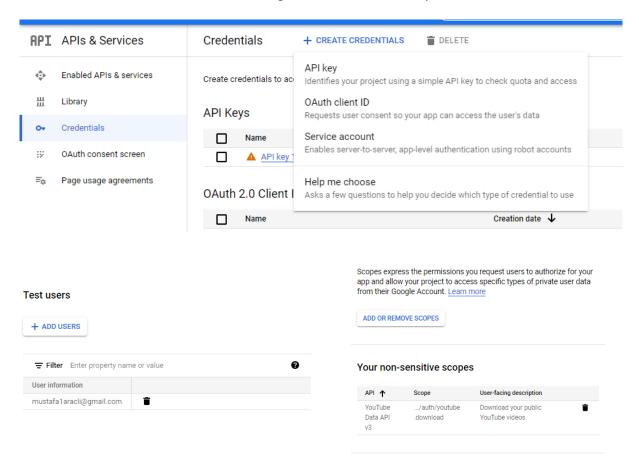
Ardından hamburger ikonuna tıklayarak APIs & Services > library' ye girerek YouYube Data API v3 api' sini aktif ediyoruz.



Yine tekrar hamburger ikonuna tıklayarak APIs & Services > OAuth consent screen' a girerek açılan sayfa da onay işlemlerini tamamlıyoruz.



Yine tekrar hamburger ikonuna tıklayarak APIs & Services > Credentials' a girerek açılan sayfa Create Credentials kısmına giriyoruz. Çıkan sekmede OAuth client ID' ye girip belirtilen yol ile anahtar ve şifremizi alıyoruz. Burada önemli nokta kütüphane kısmına YouTube API' si eklenmeli ve users kısmına kullanıcı hangi mail hesabını kullanıyorsa o mail hesabını eklemeli.



Bu işlemler yapıldıktan sonra anahtar ve şifremiz oluşturulup bize verilecektir.



YouTube Veri Çekme İşlemi

Youtube' dan veri çekmek için R Studio editörünü kullanacağız. R genel bir tabirle istatistiksel hesaplama, grafik ve veri analizi için kullanılan açık kaynaklı özgür bir yazılım dilidir.

Editörümüzü açtıktan sonra "tuber" adlı paketi indirmeli ve aktif etmeliyiz.

```
install.packages("tuber")
library(tuber)
```

Ardından id ve şifremizi bir değişkenin ile eşleştiriyoruz. Burada ki amacımız her seferinde id ve şifremizi yazmak yerine kısa adlar ile çağırmaktır.

yt_oauth(id,pass)

Bu komutu yazdığımızda API ile editörümüz arasında bağlantı kurulmak için bir tarayıcı sayfası açılır ve yönergeler takip edilerek işlemler tamamlanır.

Bu aşama ile birlikte API ile bağlantımız kurulmuş durumdadır. Sıra video yorumları çekme işlemindedir.

```
amazon <- get_all_comments(video_id = "0wP5d050f10")
write.table(amazon, "amazon.csv")
stadia <- get_all_comments(video_id = "gg9CbWmhxNw")
write.table(stadia, "stadia.csv")
xcloud <- get_all_comments(video_id = "ok3EYsvLODQ")</pre>
write.table(xcloud, "xcloud.csv")
geforce <- get_all_comments(video_id = "4GRJNbYQm00")</pre>
write.table(geforce, "geforce.csv")
geforce_s <- get_all_comments(video_id = "MsMwkz75oJU")</pre>
write.table(geforce_s, "geforce_s.csv")
techopadplus <- get_all_comments(video_id = "AUsg9QSwUps")
write.table(techopadplus, "geforceplus.csv")
geforceD <- get_all_comments(video_id = "Ken8arVoQgo")</pre>
write.table(geforceD, "geforceD.csv")
stadia2 <- get_all_comments(video_id = "-oPcBmITLhU")</pre>
write.table(stadia2, "stadia2.csv")
```

Yukarıda ki kodda önceki yaptığımız gibi her videodan gelecek yorumlar farklı bir değişkene tanımlandı. Get_all_comments komutunun içine yorumunu çekeceğimiz videonun "/" işaretinden sonraki kısmı kullanılır. Satır çalıştırıldığında API ile bağlantı kurarak yorum sayısı ile süre arasında doğru orantılı olarak karşımıza gelir.

Write.table komutu çektiğimiz yorumlar belirtilen yola .csv formatında kaydedilir.

Verileri Temizleme

Bu aşamaya kadar veri çekme işlemini tamamladık, sıra verileri temizleme işlemindedir. Öncelikle çektiğimiz verilerin içinde işimize yarmayan verileri Excel üzerinden siliyoruz. Eğer yazılar farklı tipte gözüküyorsa yeni bir Excel dosyası açıp veri>metinden veri al kısmına tıklayıp verileri "UTF-8" formatında yeni Excel'e aktarıyoruz ve .csv olarak kaydediyoruz.

Bu aşamadan sonra verilerimiz temizlenmeye hazır. Editörümüzü açıp kullanacağımız paketleri yüklüyoruz ve aktif ediyoruz.

```
install.packages("tm")
install.packages("RCurl")
install.packages("magrittr")
install.packages("dplyr")
install.packages("tidyverse")
install.packages("ggplot2")
install.packages("funModeling")
install.packages("lubridate")
install.packages("stringr")
install.packages("tidytext")
install.packages("wordcloud2")
install.packages("wordcloud")
install.packages("RColorBrewer")
install.packages("tibble")
install.packages("tidyr")
install.packages("readr")
install.packages("readxl")
install.packages("ggpubr")
install.packages("formattable")
install.packages("ggstance")
install.packages("psych")
install.packages("GGaly")
install.packages("rstatix")
install.packages("sentimentr")
install.packages("webshot")
install.packages("htmlwidgets")
install.packages("syuzhet")
install.packages("nabor")
install.packages("data.table")
install.packages("gutenbergr")
install.packages("stopwords")
install.packages("NLP")
```

```
ibrary(NLP)
library(tm)
library(RCurl)
library(magrittr)
library<mark>(dplyr)</mark>
library(tidyverse)
library(ggplot2)
library(funModeling)
library(lubridate)
library(stringr)
library(tidytext)
library(wordcloud2)
library(wordcloud)
library(RColorBrewer)
library(tibble)
library<mark>(tidyr)</mark>
library(readr)
library(readxl)
library(ggpubr)
library(formattable)
library(ggstance)
library(psych)
library(GGaly)
library<mark>(rstatix)</mark>
library(sentimentr)
library(webshot)
library(htmlwidgets)
library(syuzhet)
library(nabor)
library(data.table)
library(gutenbergr)
library(stopwords)
```

Sıra dosyamızı çağırma işleminde kolaylık olsun diye değişken kullanıp read.csv komutu ile dosyamızı içe aktarıyoruz.

```
bulut <- read.csv(file.choose())
View(bulut)</pre>
```

Veri setimizdeki linkleri siliyoruz.

```
bulut$text<-str_replace_all(bulut$text,"http[^[:space:]]*","")</pre>
```

Veri setimizdeki "@" ve "#" işaretlerini kaldırıyoruz.

```
bulut$text<-str_replace_all(bulut$text,"#\\S+","")
bulut$text<-str_replace_all(bulut$text,"@\\S+","")</pre>
```

Veri setimizdeki noktalama işaretlerini kaldırıyoruz.

```
bulut$text<-str_replace_all(bulut$text, "[[:punct:][:blank:]]+", " ")</pre>
```

Veri setimizdeki sayıları temizliyoruz.

```
bulut$text<-removeNumbers(bulut$text)</pre>
```

Veri setimizdeki ASCII formatına uymayan karakterleri temizliyoruz.

```
bulut$text<-str_replace_all(bulut$text,"[<].*[<]"," ")
bulut$text<-gsub("\ uFFFD","",bulut$text,fixed = TRUE)</pre>
```

Veri setimizdeki alfabetik olmayan harfleri temizliyoruz.

```
bulut$text<-str_replace_all(bulut$text,"[^[:alnum:]]"," ")
```

Veri setimizdeki çıkarmak istediğimiz gereksiz kullanmayacağımız kelimeleri çıkartıyoruz.

Veri setimizdeki bağlaçlar ve gereksiz tekrarlamalar kaldırıyoruz.

```
bulut$text<-removeWords(bulut$text,liste)</pre>
```

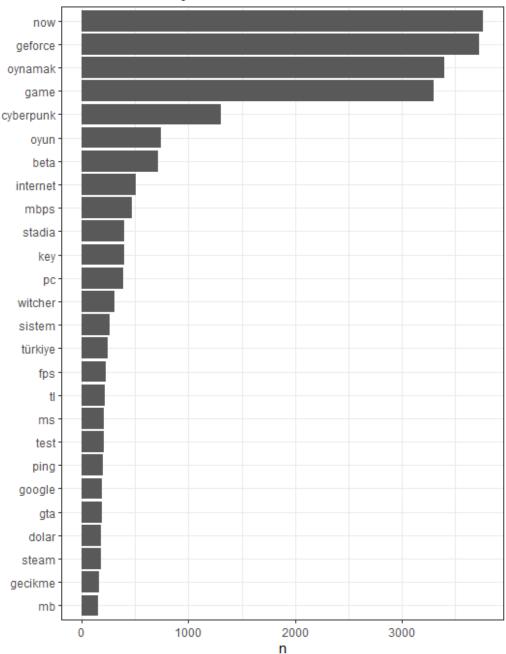
Veri setimizdeki kelimeleri alt alta sıralıyoruz.

```
sirali <- bulut %>% select(text) %>% mutate(linenumber = row_number()) %>% unnest_tokens(word,text)
```

Yorumlarda 150 den fazla tekrar eden kelimeleri grafik şeklinde çiziyoruz.

```
sirali %>%
  count(word, sort = TRUE) %>%
  filter(n > 150) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col() +
  theme_bw()+
  xlab(NULL) +
  coord_flip() +
  ggtitle("Yorumlarda en çok kullanılan kelimeler")
```

Yorumlarda en çok kullanılan kelimeler



En Çok Kullanılan Kelimelerin Anlamı

GeForce Now: NVIDIA'nın bulut tabanlı oyun akış servisidir. GeForce NOW dizüstü bilgisayarınıza, masaüstü bilgisayarınıza, Mac'inize, SHIELD TV'nize ve Android cihazınıza bulut üzerinden doğrudan gerçek zamanlı oyun deneyimi sunar. Oyun hesaplarınıza bağlanıp ister kütüphanenizdeki oyunları, ister ücretsiz oyunları dilediğiniz gibi oynayabilir, bulut kayıtlarıyla da (desteklenen oyunlar için), nerede olursanız olun, destekleyen herhangi bir cihazdan kaldığınız yerden devam edebilirsiniz.

Game: Bir oyuncu veya gamer, özellikle video oyunu, masa üstü rol yapma oyunu, müşterek bahis, beceri tabanlı iskambil oyunu ve interaktif oyunlar oynayan bir kişidir. Bazı oyuncular rekabetçidir, yani bazı oyunlarda para için rekabet edebilirler.

CyberPunk 2077: Cyberpunk 2077, CD Projekt tarafından geliştirilen ve yayımlanan bir video oyunudur. Oyun, ileri teknoloji ve sibernetiklerin yaygınlaşmış bir karamsar gelecekte geçmektedir. Oyunda açık dünya ortamı ve bölünen bir hikaye bulunur ve oyuncular "V" adlı bir merc olarak rol alırlar. Oyunun amacı, oyuncunun karakterini istediği gibi geliştirerek, Night City'de görevleri tamamlayarak zenginliğe ve güce ulaşmasıdır.

Beta: Beta, bir yazılım veya oyunun geliştirme aşamasındaki son sürümüdür. Bu sürüm, genellikle daha önce sadece belirli bir kullanıcı grubuna veya test ekibi tarafından kullanılır ve hata düzeltmeleri ve iyileştirmeler içerir.

Mbps: Mbps, "megabits per second" olarak adlandırılır. Bu birim, bir saniyede veri transfer hızını ölçmek için kullanılır.

Stadia: Stadia, Google tarafından işletilen bir bulut oyun hizmetidir. Hizmet, Kasım 2019 tarihinde başlatılmış olup yeterli yüksek hızlı internet kullanıyor olmaları koşuluyla, dünyanın dört bir yanındaki çok sayıda veri merkezi aracılığıyla oyunculara yüksek dinamik aralık desteği ile saniyede 60 kare sayısında 4K çözünürlüğe kadar video oyunları aktarabildiği belirtildi.

Fps: Aslında iki farklı şeyin kısaltması olarak kullanılır. Bunlardan biri bir oyun türü olan First Person Shooter (Birinci şahıs nişancı), diğeri ise bu yazımızın ana konusu olan Frames Per Second (Saniyedeki kare sayısı). FPS, görüntü cihazının performansını ölçmeye yardımcı olan bir birimdir. Görüntüleme ekranında her saniye gerçekleşen tam kare tarama sayısı FPS değeri olarak gösterilir.

MS: Milisaniye (ms) cinsinden ölçülen bu sinyal, verinin bilgisayarınızdan internet sunucusuna gidip gelme süresini bilmenizi sağlar. Özet olarak bu ölçüm, bilgisayar ve sunucu arasındaki 'gecikme' olarak adlandırılır.

Ping: Kullanılan bilgisayardan çıkan veri ve isteklerin diğer sunucuya iletilmesi arasındaki süre olarak olarak ölçülüyor. Örnek vermek gerekirse çevrimiçi bir oyun esnasında verilen bir komutun ya da videoyu açma-kapama gibi isteklerin karşı tarafa ne kadar sürede iletildiği ping ile açıklanıyor.

Steam: Oyunların yayınlandığı online market sistemidir.

İki farklı şekilde kelime bulutu yapıyoruz.

1.

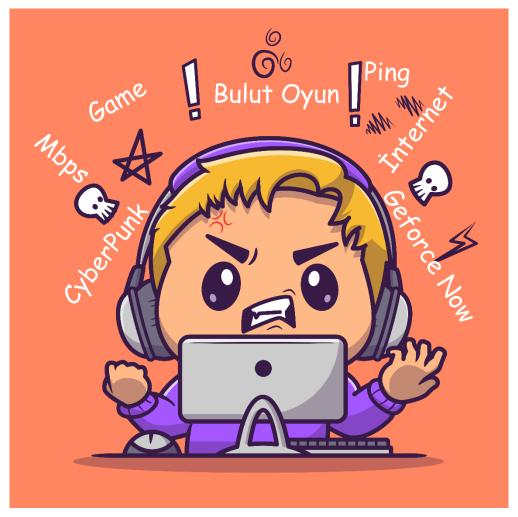
```
sirali %>%
  count(word) %>%
  with(wordcloud(word,n, max.words=40,colors = brewer.pal(11,"Spectral")))
```



2.



Ele aldığımız konu genellikle oyuncu kısma hitap ettiği için en çok kullanılan kelimeler doğrultusunda avatar tasarımı yapıldı.



Duygu Analizi

Duygu analizi, genellikle sosyal medya, müşteri geri bildirimleri, e-postalar veya çağrı transkriptleri gibi veri kaynaklarından yararlanır. Duygu analizi, genellikle pazarlama, finans, sağlık hizmetleri, üretim veya hükümet gibi farklı alanlarda kullanılır.



Gerekli paketler yüklenir ve ayarlar yapılır.

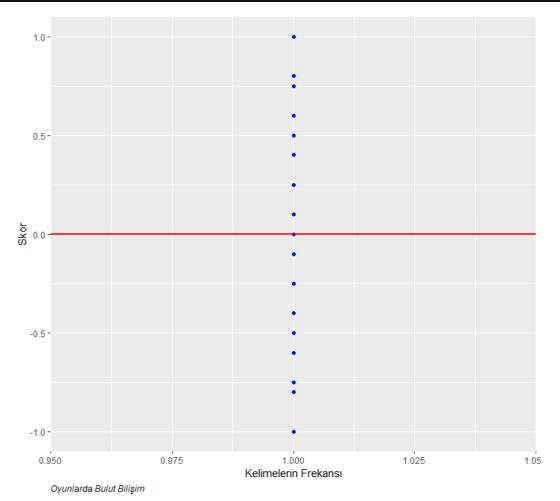
```
install.packages("sentimentr")
library(sentimentr)

polarite<-sentiment(pluss$word)
stat.desc(polarite$sentiment, basic=T) %>% pander()

polarite<-sentiment(pluss$word)
tablo<-cbind(pluss$word, polarite[,c(3,4)])
view(pluss)</pre>
```

Kelimelerin frekansını görmek için aşağıdaki kodu kullanıyoruz.

```
ggplot(tablo, aes(word_count, sentiment))+
  geom_point(color="blue")+
  geom_hline(yintercept = mean(tablo$sentiment), color="red", size=1)+
  labs(y = "Skor", x = "Kelimelerin Frekansı") +
  labs(caption = "Oyunlarda Bulut Bilişim")+
  theme(plot.caption = element_text(hjust = 0, face = "italic"))
```



Duygu gruplarına göre sınıflar

```
      duygu
      toplam
      oran

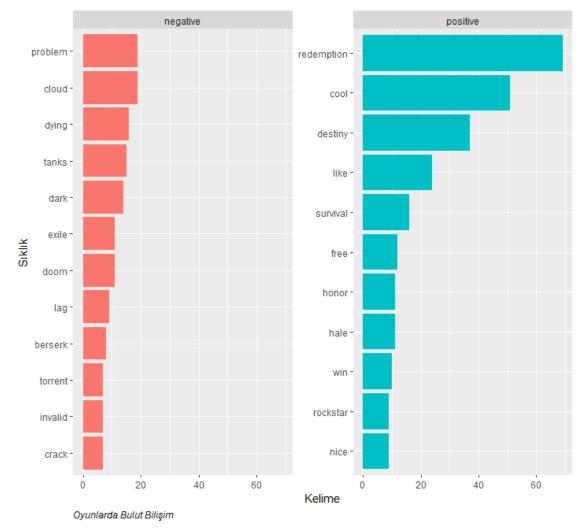
      <chr>>
      <int>><dbl>>

      1 positive
      366
      56.1

      2 negative
      286
      43.9
```

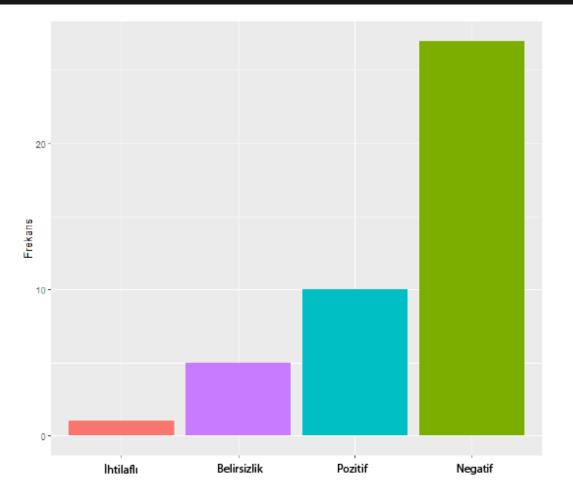
En çok tekrar eden olumlu ve olumsuz cümleler

```
duygu[,c(1,3,2)]%>% group_by(sentiment) %>% arrange(desc(n)) %>%
  top_n(10) %>%
  ggplot(aes(x=reorder(word,n), y=n, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(y = "Kelime",x = "Siklik") +
  coord_flip() +
  labs(caption = "Oyunlarda Bulut Bilişim")+
  theme(plot.caption = element_text(hjust = 0, face = "italic"))
```



4 farklı duygu analizine göre değerlendirmek için aşağıdaki kodları kullanırız.

```
install.packages("textdata")
library(textdata)
duygu<-pluss %>% inner_join(get_sentiments("afinn"),by="word")
stat.desc(duygu$value, basic=T) %>% pander()
library(syuzhet)
set.seed(1985)
duygu4<-pluss %>%inner_join(get_sentiments("loughran"),by="word")
duygu4 %>% group_by(sentiment) %>% summarise(toplam=n()) %>%
  mutate(oran=round(toplam/sum(toplam)*100,2)) %>%
  arrange(desc(oran)) %>% rename("duygu"="sentiment")
duygu4 %>%
  group_by(sentiment) %>% summarise(n=n())%>%
  ggplot(aes(reorder(sentiment, n), n, fill=sentiment)) +
  geom_bar(stat="identity", show.legend = FALSE)+
labs(y = "Frekans", x = "Duygu")+
  labs(caption = "Oyunlarda Bulut Bilişim")+
  theme(plot.caption = element_text(hjust = 0, face = "italic"))
```



Sonuç

Oyunlarda bulut bilişim konulu veri madenciliği sürecinde toplam 8 video üzerinden analizimi gerçekleştirdim. Videolardan toplam 14194 yorum ve bunların ayıklanmış hali 21881 kelime ile işlem yaptım. Seçtiğim konunun ana odağı oyun sektörü çevresinde gelişmektedir. Veri madenciliği sürecinde, veri toplama, veri temizleme, veri dönüştürme, veri depolama, veri analizi ve veri görselleştirme adımları gerçekleştirildi. Analiz sonucunda, elde edilen bilgilerin pazarlama, finans, sağlık hizmetleri, üretim veya hükümet gibi farklı alanlarda kullanılabileceği tespit edildi. Yaptığım analizler sonucunda çoğu kullanıcı "GeForce Now" servisini kullanıyor ve hakkında konuşuyor. Bunun yanı sıra kullanıcılar yorumlarda oyunlar hakkında sorular sorup yanıtlıyor. Yaşadıkları sıkıntıları internet hızı, takılmalar, gecikme gibi konular hakkında paylaşım yapıyorlar. Duygu analizinde ise kelimelerin çoğu İngilizce ve oyunların ismi olduğu için elde tutulur bir duygu analizi çıkmamıştır. Kelime bulutu ve avatar ile bu kısımda ki açıklık kapatılmıştır. Oyun oynayan kullanıcıların korkulu rüyası olan takılma, lag, ping gibi durumlarda bu sistem nasıl reaksiyon sağladığı yorumlarda tartışılmıştır. Hangi oyunda nasıl performans sağladığı da bunlar içerisindedir. Elde edilen bilgiler, veri görselleştirme teknikleri kullanılarak, grafikler, diyagramlar veya haritalar gibi farklı araçlarla görselleştirildi. Bu sayede, verilerin anlamlı bilgiler veya trendler çıkarılması sonucunda elde edilen bilgilerin kolayca anlaşılması sağlandı.

Kaynakça

https://rpubs.com/tevfik1461/duyguanalizi

https://rpubs.com/GMizuno/tubeR

https://rpubs.com/wangyuy/youtube analysis

https://chat.openai.com

Videolar;

- https://youtu.be/0wP5d050fl0
- https://youtu.be/gg9CbWmhxNw
- https://youtu.be/ok3EYsvLODQ
- https://youtu.be/4GRJNbYQm00
- https://youtu.be/MsMwkz75oJU
- https://youtu.be/AUsg9QSwUps
- https://youtu.be/-oPcBmITLhU
- https://youtu.be/Ken8arVoQgo