



TÜRKİYE CUMHURİYETİ
BİLECİK ŞEYH EDEBALI ÜNİVERSİTESİ
İKTİSADİ VE İDARİ BİLİMLER FAKÜLTESİ
YÖNETİM BİLİŞİM SİSTEMLERİ

HAZIRLAYAN

YUSUF GÜL

16726515384

ÖĞRETİM ÜYESİ

DOÇ. DR. NUR KUBAN TORUN

VERİ MADENCİLİĞİ

YAPAY ZEKA VE GELECEĞİ

BİLECİK OCAK 2024

İçindekiler

ÖNSÖZ	3
ÖZET.....	4
1.GİRİŞ	5
1.1 METİN MADENCİLİĞİ	5
1.2 DUYGU ANALİZİ	5
2.YÖNTEM	6
2.1 KAPSAM VE KISITLAR	6
2.2 PROJENİN AMACI ÖNEMİ	6
2.3 GOOGLE CLOUD APIs & SERVICES	6
2.4 YARARLANILAN YOUTUBE KANALLARI VE VİDEOLAR	7
2.5 R PROGRAMLAMA DİLİ	7
2.6 R PROGRAMLAMA KÜTÜPHANELERİ	8
2.7 R PROGRAMLAMADA KULLANILAN KODLAR VE ANALİZ AŞAMALARI	8
3. BULGULAR	12
3.1 Frekans Tablosu	12
3.2 Kelime Bulutu	12
3.3 Pozitif ve Negatif Kelime Bulutu	13
3.4 Duygu Analizi	14
3.5 DUYGU DAĞILIMI TABLOLARI	16
4. SONUÇ VE TARTIŞMA.....	17
5.KAYNAKÇA.....	18

ÖNSÖZ

Üniversite eğitimim süresince birçok projede aktif olarak yer aldım ve kendimi sürekli olarak geliştirme fırsatı buldum. Bu süreçte, 4. sınıfta aldığımız Veri Madenciliği dersiyle birlikte, oldukça kapsamlı ve nitelikli bir proje üzerinde çalışma şansını elde ettim.

Bu projenin, iş hayatımda benim için değerli bir kaynak olacağına ve bu alandaki yetkinliklerimi daha da geliştirmeme katkı sağlayacağına inancım tam. Aldığımız eğitim ve projelerde edindiğim deneyimler, gelecekteki kariyerimde benim için önemli birer adım olacak.

Özellikle, rehberliği, yönlendirmeleri ve destekleriyle projenin başarılı bir şekilde tamamlanmasında büyük katkı sağlayan Sayın **Doç. Dr. Nur Kuban Torun**'a özel bir teşekkür borçluyum. Verdiğiniz emek ve destekler için minnettarım.

Sizin rehberliğiniz ve desteğinizle edindiğim bilgi birikimi ve deneyimler, benim için çok kıymetli. Bu projede gösterdiğiniz ilgi ve destekleriniz, sadece bu projeye değil, aynı zamanda gelecekteki kariyerimdeki başarılarıma da ışık tutacak.

Proje sürecinde birbirimizle fikir alışverişinde bulunduğumuz **Eren İleri**'ye de ayrıca teşekkür ederim.

ÖZET

Teknoloji sektöründeki kuruluşlar, yapay zeka ve büyük veri analizi konularına odaklanarak, veri analizinden elde edilecek kritik bilgilerin değerini bilmektedirler. Bu veri yığınlarından anlamlı veriler çıkarmak, uzun ve titiz bir süreci gerektirir. Şirketler, veri madenciliği faaliyetleri sonucunda elde ettikleri bilgilerle, sadece kendi iç verilerini değil, dışarıdan gelen veri kaynaklarını da etkili bir şekilde işleyip koruma becerisi kazanmaktadırlar.

Kendi çalışmamda, "R" programlama dilini kullanarak YouTube platformundan elde ettiğim 57000 yorum verisini temizleyip bu yorumları tokenleştirerek 41000 kelime veri seti üzerinde metin analizi ve duygu analizi gerçekleştirdim. Bu analizlerde, yapay zeka teknolojileri temel alınarak YouTube'un önde gelen hesaplarından olan "YouTube Originals", "TEDx Talks", "Venture City", "ColdFusion", "TED-Ed", "DW Documentary", "LastWeekTonight" gibi hesapların verilerini inceledim. API anahtarlarını kullanarak elde ettiğim yorumları detaylı bir şekilde analiz ettim.

Bu çalışma, yapay zeka ve büyük veri analizi alanında önemli bir ilerlemedir. Yapılan analizler, yapay zeka teknolojilerinin sosyal medya platformlarındaki verileri kavrama ve işleme yeteneğinin ne kadar kritik olduğunu vurgulamakta ve gelecekte nasıl bir yer alacağını önizlemektedir. Büyük veri içindeki örüntüleri çıkarmak ve anlamak, günümüz şirketlerinin rekabet avantajı elde etmesinde kilit bir rol oynamaktadır

1.GİRİŞ

1.1 METİN MADENCİLİĞİ

Metin madenciliği, yapılandırılmamış metin verilerinden araştırmaya uygun bilgilerin ortaya çıkarılmasında kullanılan temel bir yöntemdir. Bu yöntem, yapılandırılmamış verileri, daha anlamlı ve kullanılabilir hale getirerek bilgiye dönüştürür. Veri büyüklüğü, işleme sürecini etkilemeden büyük veri setlerini hızlı ve yüksek doğruluk oranlarıyla işleyebilme yeteneği sunar.

Metin madenciliği için kullanılacak veriler çeşitli kaynaklardan toplanabilir. Bu kaynaklar arasında gazeteler, web uygulamaları, akademik makaleler, sosyal medya platformları gibi çeşitli seçenekler bulunmaktadır. Özellikle, sosyal medya platformları, güncel ve geniş veri kaynaklarına erişimin yanı sıra, kullanıcıların duygu, düşünce ve yorumlarını en doğal ve samimi biçimde ifade edebildiği bir alan sunmasıyla önem kazanmaktadır.

Bu çalışmada sosyal medyanın odak noktası olmasının en büyük sebeplerinden biri, kullanıcıların duygusal ve düşünsel açıklamalarını serbestçe yapabildikleri bu platformlardan en güncel verilere ulaşılabilmesidir. Bu veriler, metin madenciliği ile incelendiğinde, toplumun genel eğilimleri, beklentileri ve tepkileri gibi önemli bilgilere erişim sağlamak için büyük bir potansiyel sunar.

1.2 DUYGU ANALİZİ

Duygu analizi, metin madenciliğinin temel alanlarından biridir ve aynı zamanda fikir madenciliği olarak da bilinir. Bu analiz yöntemi, metin madenciliğinde önemli bir rol oynar. İnsanların bir konu hakkındaki duygularını, düşüncelerini ve öznel değerlendirmelerini ölçmek için çeşitli algoritmalar kullanılarak metnin ana temasını belirler.

Verilerdeki metnin ana teması belirlendikten sonra, bu belirlenen kelimelerden kelime bulutları oluşturulur. Bu adım, metnin okuyucuya iletmek istediği duygunun daha anlaşılır hale gelmesini sağlar. Duygu analizi, konu üzerinde duygu skorları hesaplayarak pozitif, negatif veya nötr olarak sınıflandırma yapar. Bu sayede üzerinde çalışılan konunun ağırlıklı duygusal yönü analiz edilebilir.

Duygu analizi, farklı seviyelerde gerçekleştirilebilir ve genellikle üç düzeyde incelenir: doküman düzeyi, cümle düzeyi ve öznitelik düzeyi. Bu çalışmada özellikle cümle düzeyinde yapılan duygu analizi, metnin her bir cümlesinin duygusal yönlerini inceleyerek derinlemesine bir anlayış sağlar.

Bu analiz yöntemi, metnin içsel duygusal zenginliğini anlamak için kapsamlı bir perspektif sunar. Özellikle cümle düzeyinde yapılan analiz, her bir ifadenin taşıdığı duygusal yükü anlamak ve metnin genel tonunu belirlemek açısından büyük önem taşır.

2.YÖNTEM

Yapay Zeka ve Geleceği konusu, sosyal medyada insanların bu konu hakkında yaptığı paylaşımlardan, düşüncelerini objektif bir şekilde ölçmek için “metin madenciliği” yöntemleri kullanılmıştır. Çalışmada içerik analizi ve duygu analizi ile ölçüm yapılmıştır. Analiz yapılırken açık kaynak kodlu bir programlama dili olan R programlama kullanılmıştır. R programlamanın kullanılmasının sebebi büyük veriyi rahatlıkla analiz edebilmesidir.

2.1 KAPSAM VE KISITLAR

İnsanların sosyalleşmek ve dış dünya dışında ihtiyaç duydukları kaynaklara ulaşmak için sosyal medyaya yöneldiği olası bir gerçektir. Kullanıcılarının oldukça önemli bir vaktini alan sosyal medyada, birçok fikir ve düşünce vardır. Bu çalışmanın kapsamı, sosyal medya uygulamaları özelinde YouTube’dur. Bir diğer ifadeyle bu çalışmanın kapsamı, belirtilen uygulamadaki kullanıcı yorumlarıdır. Bu doğrultuda, bu çalışma kapsamında YouTube geliştirici hesabı kullanılarak YouTube yayınından 57 bin 14 yorum çekilmiştir. Belirtilen kapsama ek olarak, proje yapılırken verilerin 19 Temmuz 2016’dan 27 Şubat 2023 tarihine kadar çekilmiş olması da zaman kısıtlılığı sorununu beraberinde getirmiştir.

2.2 PROJENİN AMACI ÖNEMİ

Sosyal medya kullanıcılarının belirli bir konuyla ilgili verdikleri tepkileri ölçmek ve yapay zeka alanındaki geleceğe dair düşünceleri analiz etmek için toplanan veriler üzerinden duygu analizi yapmayı hedefliyoruz. Bu analiz, kullanıcıların yaptığı yorumlardaki duygu ve düşünceleri belirlemeye yönelik olacak.

2.3 GOOGLE CLOUD APIs & SERVICES

Google Cloud, kullanıcılarına sunduğu güçlü bulut platformu hizmetiyle, verilerin hesaplanması, depolanması ve geliştiricilerin uygulamalar oluşturması, test etmesi ve dağıtmasına olanak tanır. Bu kullanıcı dostu platformdan yararlanmak için sadece bir Google hesabına sahip olmanız yeterlidir. Google Cloud'un sağladığı uygulama programlama arayüzü (API) üzerinden adım attıktan sonra, "OAuth 2.0 Client IDs" ve "API Keys" olarak, veri çekme sürecinde kullanacağımız önemli bilgiler olan 'Client Secret' ve 'Client ID' bilgilerine ulaşabiliriz. Bu güvenli erişim bilgileri, veri çekme işlemlerinizi kesintisiz bir deneyim sunar ve Google Cloud'un geniş olanaklarından en iyi şekilde faydalanmanıza olanak sağlar.

2.4 YARARLANILAN YOUTUBE KANALLARI VE VİDEOLAR

KANAL	VİDEO	BAĞLANTI
TEDX TALKS	Understanding Artificial Intelligence and Its Future Neil Nie TEDxDeerfield	Understanding Artificial Intelligence and Its Future Neil Nie TEDxDeerfield (youtube.com)
VENTURE CITY	TIMELAPSE OF ARTIFICIAL INTELLIGENCE (2028 – 3000+)	TIMELAPSE OF ARTIFICIAL INTELLIGENCE (2028 – 3000+) (youtube.com)
COLDFUSION	What is Artificial Intelligence Exactly?	What is Artificial Intelligence Exactly? (youtube.com)
TED-ED	How will AI change the world?	How will AI change the world? (youtube.com)
DW DOCUMENTARY	Artificial intelligence and algorithms: pros and cons DW Documentary (AI documentary)	Artificial intelligence and algorithms: pros and cons DW Documentary (AI documentary) (youtube.com)
LASTWEEKTONIGHT	Artificial Intelligence: Last Week Tonight with John Oliver (HBO)	Artificial Intelligence: Last Week Tonight with John Oliver (HBO) (youtube.com)
YOUTUBE ORIGINALS	How Far is Too Far? The Age of A.I.	How Far is Too Far? The Age of A.I. (youtube.com)

2.5 R PROGRAMLAMA DİLİ

R programlama dili istatistiksel hesaplama, veri analizi ve bilimsel araştırmalarda yaygın olarak kullanılan bir araçtır. İstatistikçiler, veri bilimcileri, analistler, araştırmacılar ve pazarlamacılar tarafından tercih edilir. Bu dil, verileri temizlemek, analiz etmek, görselleştirmek ve anlamlı hale getirmek için güçlü araçlar sunar.

R, dünya çapında birçok analist ve veri bilimcisi tarafından tercih edilen bir dil haline gelmiştir. Özellikle, finanstan üretime, e-ticareten sağlığa, bankacılıktan kapsamlı pazarlamaya kadar geniş bir yelpazedeki sektörlerde kullanılmaktadır. Son zamanlarda güncellenen kütüphaneleriyle, R programlama dili birçok veri işleme alanında kullanılmakta ve farklı sektörlerdeki en karmaşık sorunları çözmek için başvurulmuş bir araç haline gelmiştir.

Birçok büyük kuruluş, R'ı etkin bir şekilde kullanmaktadır. Örnek olarak Twitter, Facebook, Amazon, Mozilla, Microsoft, Google, Bank Of America, Merck, Ford Motor Company, National Weather Service gibi kurumlar, R'ı analitik çözümlerinde temel bir araç olarak benimsemişlerdir.

R, kullanıcılarına güçlü bir programlama arayüzü sunmanın yanı sıra, çeşitli istatistiksel ve veri analizi işlemleri için zengin bir kütüphane ve geniş bir topluluk desteği sağlar. Bu özellikleriyle, farklı endüstrilerdeki karmaşık veri sorunlarını çözmek ve veri odaklı kararlar almak isteyen profesyoneller için önemli bir araç haline gelmiştir.

2.6 R PROGRAMLAMA KÜTÜPHANELERİ

R programlama dilinde YouTube uygulamasından veri çekip analiz etmek için hazır kütüphaneler kullanılmıştır. Bunun için `install.packages()` komutu kullanılmıştır ve `library()` komutu ile aktif edilmiştir, Bu kütüphaneler Şekil 1'deki gibidir.

```
#Kullanacağımız paketleri yükliyoruz
data("stop_words")
library(tuber) #YouTube API ile etkileşim kurmak için kullanılmıştır.
library(httputv) #Elde ettiğimiz çıktıları html olarak görmek için kullanılmıştır.
library(dplyr) #metin verilerini temizlemek ve manipüle etmek için kullanılmıştır.
library(ggplot2) #metin verilerini görselleştirmek için kullanılmıştır.
library(tibble) #metin verilerini temizlemek ve manipüle etmek için kullanılmıştır.
library(magrittr) #%% komutu ile zincirleme için kullanılmıştır.
library(pander) #istatistiksel özetler oluşturmak için kullanılmıştır.
library(pastecs) #çeşitli istatistiksel yöntemler ve özetler sunar
library(sentimentr) #metin verilerinin duygusal yönünü analiz etmek için kullanılmıştır.
library(stopwords) #istenmeyen kelimeleri çıkartmak için kullanılmıştır.
library(hwordcloud) #kelime bulutu oluşturmak için kullanılmıştır.
library(wordcloud2) #kelime bulutu oluşturmak için kullanılmıştır.
library(tidytext) # metin verilerinden kelimeleri çıkarmak için kullanılmıştır.
library(stringr) #metinlerle çalışmayı mümkün olduğunca kolay hale getirmek için kullanılmıştır.
library(tm) #metin verilerini işlemek ve analiz etmek için kullanılmıştır.
```

Şekil 1:Kullanılan Kütüphaneler

2.7 R PROGRAMLAMADA KULLANILAN KODLAR VE ANALİZ AŞAMALARI

R programlama dilinde kütüphanelerin aktif hale getirilmesinden sonra Youtube API'sine erişim sağlanmış ve çekilen veriler birleştirilmiş, veriler Şekil 2 ve Şekil 3' te gösterildiği gibi R' a çağırılmıştır.

```
#YouTube API'sine erişim sağlamak için tuber adlı bir R paketini kullanıyoruz "yt_oauth" fonksiyonu, YouTube API'sine
erişim yetkisi sağlar.
app_id <- "XXXXXXXXXXXXXXXXXXXX.apps.googleusercontent.com"
app_secret <- "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
yt_oauth(app_id, app_secret, token = "")
```

Şekil 2:YouTube API erişimi

```
#"get_all_comments" fonksiyonu, belirli bir YouTube videosunun tüm yorumlarını çekmek için kullanıldı.
video_ids<- c("SN2BZswEUUA","63yr9d1I0cU","kwmX3pd1f10","RzkD_rTEBys","s0dMTAQM4cw","Sqa8Zo2Xwc4","UwsrzCVZAb8")
comments_list <- lapply(video_ids, function(id)
  get_all_comments(video_id = id))
all_comments <- do.call(rbind, comments_list)
```

Şekil 3:YouTube Verilerinin Çekilmesi ve Birleştirilmesi

Yorumlar çekilip birleştirildikten sonra daha sağlıklı bir ön işlem yapılabilmesi için veri seti içinden Şekil 4'de gösterildiği gibi sadece yorumlar ayıklandı.

```
comments <- datas%>%
  dplyr::select(textOriginal)
```

Şekil 4: Yorum Ayıklama

Çekilen yorumların analize uygun olması amacıyla ön işleme tabi tutulmuştur. Bu işlemler şu şekilde gerçekleşmiştir; metindeki harfler küçük harfe dönüştürülmüştür. Noktalama işaretleri ve rakamlar temizlenmiştir. Metinle iç içe geçmiş (örnek: yusuf123) rakam ve sayılar metinden ayıklanmıştır. Alfabetik olmayan karakterler çıkartılmıştır. Uzatılmış anlamsız kelimeler kaldırılmıştır (örnek: aaaaawwww). Bu temizleme işlemleri Şekil 5 ve Şekil 6'daki

kodlar ile yapılmıştır. Temizlenmiş yorumlardan stopwords olarak adlandırılan durak kelimeler çıkartılmıştır. Bu aşamada bağlaçlar, gereksiz tekrarlar, çalışmadan bağımsız ve konu bütünlüğünü bozan kelimeler ortadan kaldırılmıştır. Kişisel verilerin korunması adına isimler ve insan haklarına saldırı anlamı taşıyan kelimeler analizden çıkarılmıştır.

```
#TEMİZLİK İŞLEMLERİ
comment <- comments%>%
  mutate(word=str_to_lower(textOriginal))%>% unnest_tokens(word, textOriginal) # Metindeki harflerin tamamı küçük harfe
  donusturulmuştur.
comment <- comment %>% mutate(word=removePunctuation(word)) # Metindeki noktalama işaretleri kaldırılmıştır.
comment <- comment %>% mutate(word=str_squish(word)) # Metinle iç içe geçmiş (örnek:yusuf123) rakam ve sayılar metinden
  ayıklanmıştır.
comment <- comment %>% mutate(word=removeNumbers(word)) # Metinden rakam ve sayılar çıkarılmıştır.
uzatilmis_kelime_deseni <- "[a-z]\\1{2,}" # Uzatılmış kelimeleri bul
comment$word <- str_replace_all(comment$word, uzatilmis_kelime_deseni, "\\1")# Uzatılmış kelimeleri kaldır
comment <- comment %>% mutate(word = sapply(word, remove_emoji))#emoji temizle
comment <- comment %>%
  mutate(word = str_replace_all(word, "[<.*[>]", ""),#HTML etiketlerini kaldırmak için kullanılır.
    word = gsub("\uFFFF", "", word, fixed = TRUE),#\uFFFF genellikle bilinmeyen veya tanımlanamayan karakterleri
    temsil eder.
    word = gsub("\n", "", word, fixed = TRUE))#her bir kelimenin içindeki yeni satır karakterlerini (\n) siler.

comment <- comment %>% filter(!word %in% custom_stop_words$word) # Durma kelimelerini kaldır
comment <- comment %>% filter(str_length(word)>3) # Karakter sayısı 3'ten büyük kelimeler filtrelenmiştir.
comment$word <- str_replace(comment$word, "[ı]", "i") # Metin istenmeyen formatta kelimeler varsa ve bu kelimeleri çıkarmanız
  analize zarar verecekse yerine kelimeler atayabilirsiniz.
comment <- comment %>% as_tibble()%>%rename(word=value) # Metnin değişiklikler sonra tibble tablo düzenine
  donusturulmuştur.
```

Şekil 5: Verileri Temizleme İşlemleri

```
# ingilizce olan kelimeler alındı.
comment <- filter(comment, str_detect(word, "[a-zA-Z]+"))
```

Şekil 6: İngilizce Olan Kelimeleri Filtreleme

Temizleme işlemi biten veriler içerik analizi yapmak için parçalanmıştır. Parçalama işleminden sonra veriler count edilmiştir. Bu işlem Şekil 7’deki gibi gösterilmiştir.

```
words <- comment %>%
  unnest_tokens(word, word) %>%
  anti_join(custom_stop_words) %>%
  count(word, sort = TRUE)
```

Şekil 7: Verileri Parçalama İşlemi

Hazırlanan veriyi anlamlı hale getirmek ve kelimelerin frekans sayılarını bulmak amacıyla en çok tekrar eden kelimeler grafiği ve kelime bulutu oluşturulmuştur. Bunun için Şekil 8 ve Şekil 9’ da ki kodlar kullanılmıştır.

```
words%>%#ggplot2 kütüphanesinden yararlanarak yatay bir biçimde ilk 20 verimizi görselleştirdik
head(20)%>%
ggplot(aes(reorder(word, n),n))+
geom_col(fill="#5c9a9b", alpha=0.6)+
coord_flip()+
labs(x="Kelimeler",
y="Tekrar Edilme Sayısı",
title = "EN ÇOK TEKRAR EDİLEN 20 KELİME")+
theme_test()
```

Şekil 8: Kelime Frekans Grafiği

```
wordcloud2(data = head(words, 100),
minRotation = -pi/6,
maxRotation = -pi/3,
minSize = 0.1,
rotateRatio = 0,
size = .7,
color = "random-dark",
fontWeight = "bold",
fontFamily = "sans")
```

Şekil 9: Kelime Bulutu

Bing duygu sözlüğü kullanılarak kelimelere pozitif ve negatif duygular verilmiştir. Verilen bu duygular sonucunda pozitif kelimeleri içeren kelime bulutu ve negatif kelimeleri içeren kelime bulutu ayrı ayrı görselleştirilmiştir. Bunun için Şekil 10 ve Şekil 11’ de ki kodlar kullanılmıştır.

```
# Bing duygu sözlüğünü alın
bing_sentiments <- get_sentiments("bing")

# Kelimelerinizi duygu değerleriyle birleştirin
words_sentiments <- words %>%
inner_join(bing_sentiments, by = "word")

# Pozitif ve negatif kelimeleri ayırın
positive_words <- words_sentiments %>%
filter(sentiment == "positive") %>%
head(100)

negative_words <- words_sentiments %>%
filter(sentiment == "negative") %>%
head(100)
```

Şekil 10: Pozitif ve Negatif Kelime Bulutu İçin Duygu Ataması

```
wordcloud2(positive_words)
wordcloud2(negative_words)
```

Şekil 11: Pozitif ve Negatif Kelime Bulutu Çıktısı

Duygu Analizinin görselleştirilmesi için Bing duygu sözlüğü ile yüzde grafiği yapılmıştır. Şekil 12’deki kodlar yazılmıştır.

```
words %>%
  inner_join(get_sentiments("bing")) %>%
  group_by(sentiment) %>%
  summarise(n = mean(n)) %>%
  ggplot(aes(x = sentiment, y = n, fill = sentiment)) +
  geom_col() +
  scale_fill_manual(values = c("positive" = "#00bfc4", "negative" = "#f8766d")) +
  xlab("Duygu") +
  ylab("Frekans")
```

Şekil 12: Yüzde Grafiği

“Sentiment” fonksiyonu ile yapılan duygu dağılımı grafiği pozitif nötr ve negatif olarak yapılmıştır. Şekil 13’deki kodlar yazılmıştır.

```
sentiments<-tablo$sentiment
sentiments<- as.data.frame(sentiments)

sentiments$sentiments[sentiments$sentiments>0] <- "positive"
sentiments$sentiments[sentiments$sentiments==0] <- "neutral"
sentiments$sentiments[sentiments$sentiments<0] <- "negative"

sentiments$sentiments<- as.factor(sentiments$sentiments)

katego <- table(sentiments$sentiments)
colors <- c("red", "grey", "green")
barplot(katego, main = "Duygu Dağılımı", xlab = "Kategori", ylab = "Sıklık", col = colors)
```

Şekil 13: Pozitif Nötr ve Negatif barplot

Duygu sözlüğüne göre pozitif ve negatif en çok tekrar eden ilk 20 kelime Şekil 14’deki gibi görselleştirilmiştir.

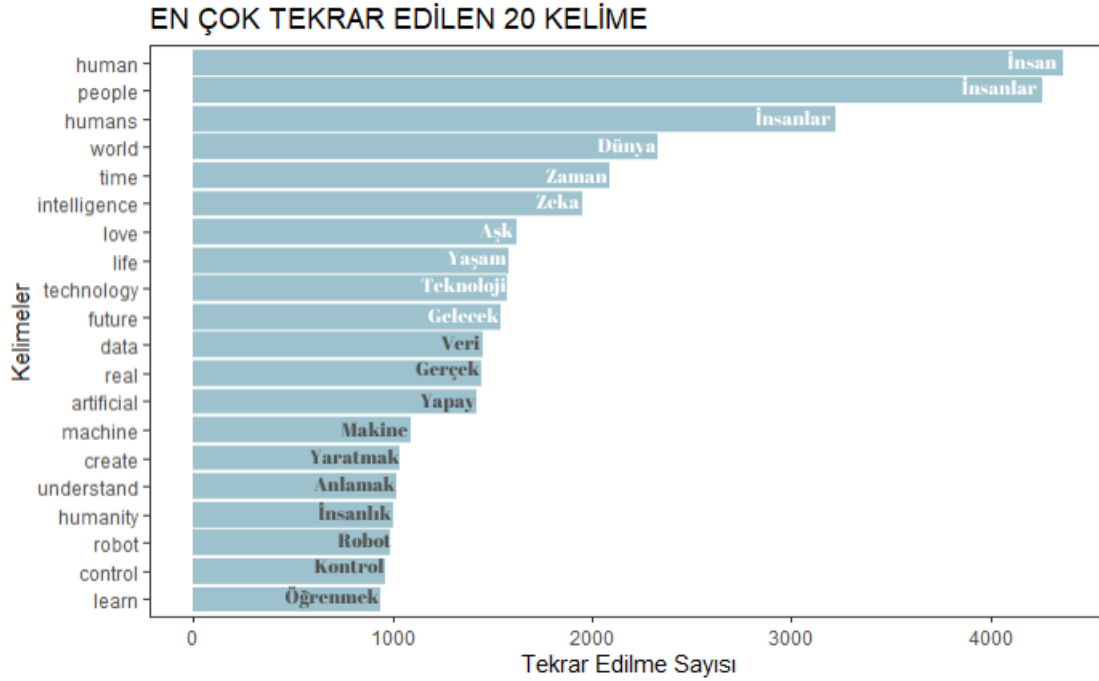
```
words %>%
  inner_join(get_sentiments("bing")) %>%
  head(20) %>%
  arrange(-n) %>%
  group_by(sentiment) %>%
  ggplot(aes(reorder(word, n), n, color = sentiment)) +
  geom_line(aes(group = 1)) +
  geom_point() +
  facet_wrap(~sentiment, scales = "free_y") +
  coord_flip() +
  xlab("Kelime") +
  ylab("Frekans")
```

Şekil 14: Pozitif ve Negatif En Fazla Tekrar Eden Kelimeler

3. BULGULAR

3.1 Frekans Tablosu

Yapılan yorumlar içerisinde en çok tekrar eden kelimenin Şekil 15’te gösterildiği gibi “human” olduğu görülmektedir. Yakın bir farkla “people” ve “humans” kelimeleri “human” kelimesini takip etmektedir.



Şekil 15:En Çok Tekrar Eden 20 Kelime

3.2 Kelime Bulutu

YouTube’den 7 farklı videodan çekilen yorum sayısı 57014 adettir. Veri setinin parçalanmış şekildeki kelime bulutunun gösterimi Şekil 16’daki gibi ortaya çıkmıştır. Oluşturulmuş kelime bulutunda frekansına göre ilk 100 kelime ortaya çıkarılmıştır. Bu kelime bulutuna göre Yapay Zeka ile öne çıkan ilk beş kelime: human (İnsan), people (İnsanlar), humans (İnsanlar), world

(Dünya), time (Zaman) olmuştur.



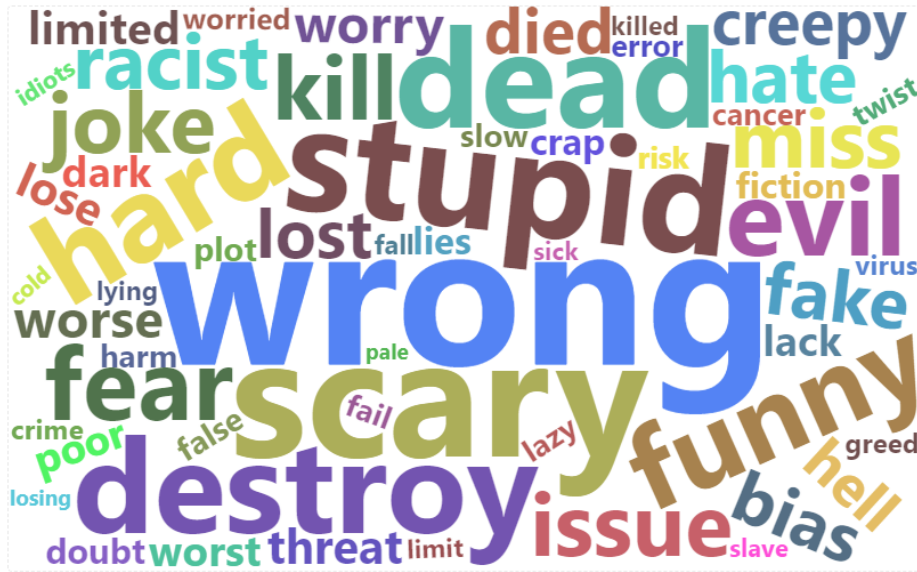
Şekil 16: Kelime Bulutu

3.3 Pozitif ve Negatif Kelime Bulutu

Veri setinin parçalanmış şekildeki kelimelerine duygu sözlüğü ile pozitif ve negatif duygular atayarak pozitif kelime bulutu ve negatif kelime bulutu oluşturulmuştur. Bu kelime bulutlarına göre Pozitif için intelligence (Zeka), love (Aşk), free (Özgür), smart (Akıllı), intelligent (Zeki) kelimeleri öne çıkmıştır. Negatif için ise wrong (Yanlış), scary (Korkutucu), stupid (Aptal), dead (Ölü), destroy (Tahrip Etmek) kelimeleri öne çıkmıştır. Pozitif kelime bulutu Şekil 17'deki gibi ve Negatif kelime bulutu ise Şekil 18'deki gibidir.



Şekil 17: Pozitif Kelime Bulutu

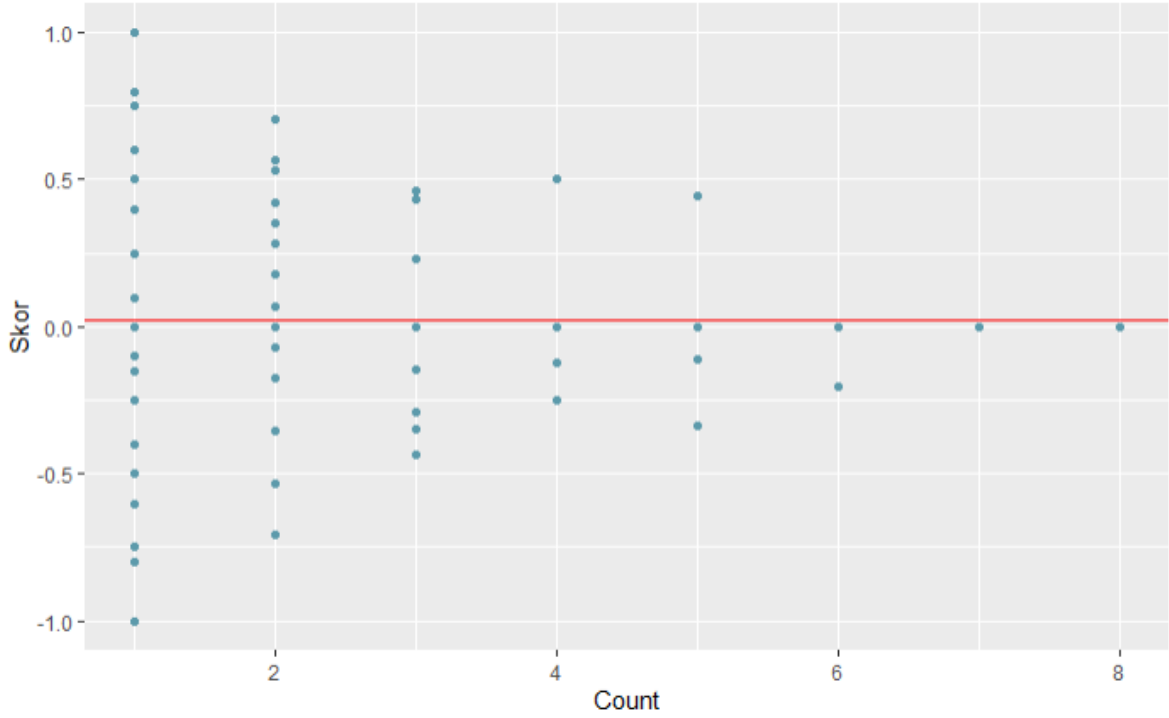


Şekil 18: Negatif Kelime Bulutu

3.4 Duygu Analizi

Duygu durumunun pozitif veya negatif yönlü olmasının gösterimi Şekil 19'daki gibi oluşturulmuştur. +1 değerine yaklaşan değerler pozitif yönlü, 0 nötr ve -1'e yaklaşan değerler negatif duygu yönünü göstermektedir. Analiz sonucuna göre duygu analizi 0.02026 değeri ile 0'a yakınlığından dolayı nötr çıkmış, 0 değerini biraz aşmasından dolayı pozitif yönelimli

çıkarmıştır yorumu yapılabilir. Buradaki kırmızı çizgi duyguyu belirtmektedir. Maviler ise her kelimenin -1 ile 1 arasında ki duygu yönünü belirtmektedir.



Şekil 19: Polarite Analizi

Yapılan duygu analizinin istatistiksel karşılığı Şekil 20’de verilmiştir. Bu tablodaki “mean” kısmı duygu analizimizin matematiksel karşılığına denk gelmektedir. Mean pozitif ve negatif durumların ortalamasını alıp gerçek sonucu veren sütundur.

nbr.val	nbr.null	nbr.na	min	max	range	sum	median	mean
404608	291915	0	-1	1	2	8214	0	0.0203

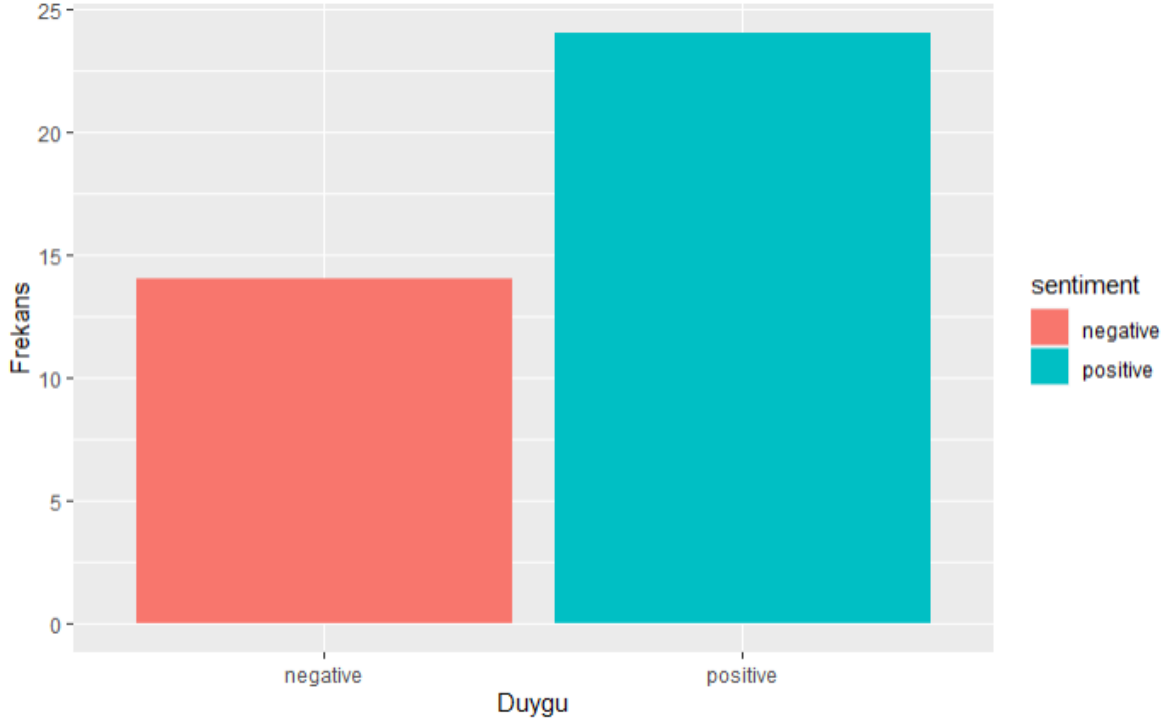
Table: Table continues below

SE.mean	CI.mean.0.95	var	std.dev	coef.var
0.0005307	0.00104	0.114	0.3376	16.63

Şekil 20: İstatistiksel Sonuç

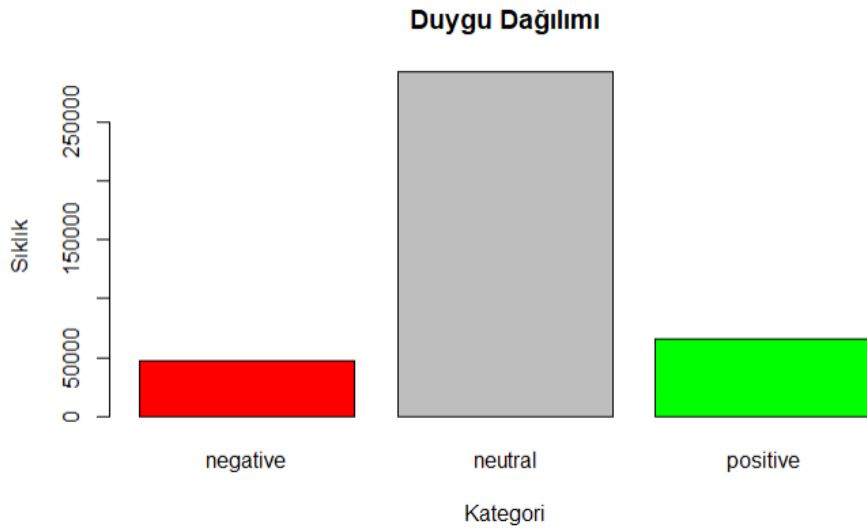
3.5 DUYGU DAĞILIMI TABLOLARI

Bing duygu sözlüğüne göre oluşturulan şekil 21’de pozitif bir baskınlık görülmektedir.



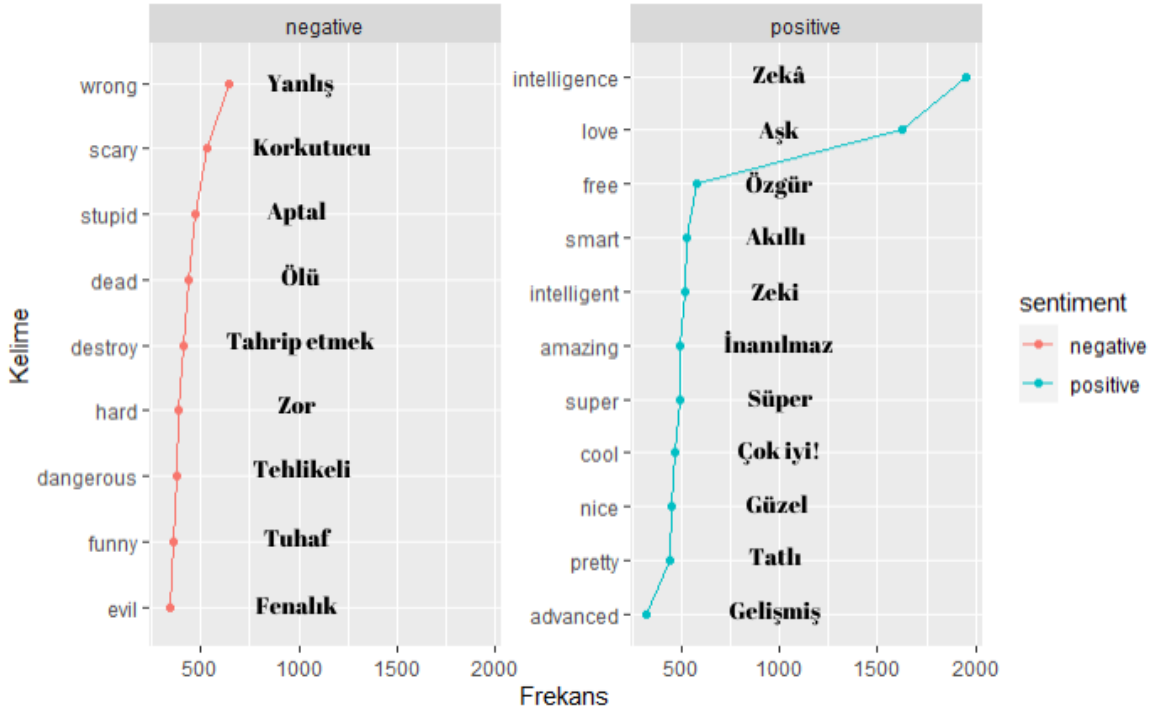
Şekil 21: Bing Sözlüğü ile Duygu Dağılımı

Sentiment fonksiyonu ile oluşturulan şekil 22’de ki bar grafiğinde ise pozitif yönelimli nötr bir baskınlık görülmektedir.



Şekil 22: Sentiment Fonksiyonu ile Duygu Dağılımı

Yorumlarda insanların duygularını ifade etmek amacı ile en çok kullandıkları kelimelerin Pozitif ve Negatif dağılımı Şekil 23'deki gibidir.



Şekil 23: Duygu sözlüğüne göre en çok kullanılan kelimeler

4. SONUÇ VE TARTIŞMA

YouTube platformu, kullanıcılar tarafından yoğun bir şekilde kullanılarak birçok içerik üretilmiş ve bu içeriklere yapılan yorumlar aracılığıyla kullanıcıların düşüncelerini ifade etmelerine olanak tanımıştır. Bu çalışmada, kullanıcıların Yapay Zeka ve Geleceği konusundaki bakış açıları, duyguları ve düşünceleri analiz yöntemleri kullanılarak incelenmiştir.

Metin içinde en fazla kullanılan 20 kelimenin sıralaması, frekans grafiği ve kelime bulutu ile görselleştirilerek sunulmuştur. "human" ve "people" kelimelerinin yorumlarda öne çıkması, yorum yapanların insanlığın önemine ve değerine vurgu yapmayı amaçladığını göstermektedir. Bu kelimeler, insanların birlikteliğini, dayanışmasını ve ortak değerlerini ifade etmek için kullanılmış olabilir. Ayrıca, bu kelimelerin insanlığın karşı karşıya olduğu sorunlara ve zorluklara dikkat çekmek amacıyla da kullanıldığı düşünülebilir.

Yapılan kelime bazlı duygu analizi, Yapay Zeka konusuna karşı genel bir nötr duygu skoru ortaya koymuştur. Frekans ve duygu analizi sonuçlarına göre, insanların Yapay Zeka konusuna nötr bir bakış açısına sahip olduğunu söyleyebiliriz. Ancak, şekil 20'de görüldüğü üzere ortalama duygu skoru pozitif bir nötrlük içermektedir.

Bu bulgular, kullanıcıların Yapay Zeka ve gelecekteki rolü konusunda genel olarak dengeli bir perspektife sahip olduğunu göstermektedir. Yapay Zeka konusundaki nötr bakış açısı, hem

olumlu hem de olumsuz yönleri değerlendirmeye açık bir tutumu yansıtabilir. Araştırma sonuçları, Yapay Zeka alanındaki tartışmaların geniş bir yelpazede gerçekleştiğini ve kullanıcıların çeşitli görüşlere sahip olduğunu göstermektedir.

5.KAYNAKÇA

- Metin Madenciliğinde Makine Öğrenmesinin Uygulanması | by Onur Dayıbaşı | Programming / Algorithms & Data Structures | Medium
- https://youtube.com/playlist?list=PLkbyF4y0IkKADZ74l4Y9S1ZlZr4sq7H_L&si=_j_M6SF2aT7UMkHV
- R-Studio-Twitter-Text-Mining-and-Sentiment-Analysis/sentiment.R at main · sametturgut/R-Studio-Twitter-Text-Mining-and-Sentiment-Analysis · GitHub
- ESTUStat_2022Guz_VeriGorsellestirme/Alıřtırmalar/R_textvis.R at 0909ca0a167ce76f68b966b8d9a93359c719fc07 · mcavs/ESTUStat_2022Guz_VeriGorsellestirme · GitHub
- 727452 (dergipark.org.tr)
- Feeling Analysis for Sadness and Happiness using Googlen-gram Database Googlen-gram Veritabanı ile Üzüntü ve Mutluluk Üzerine Duygu Analizi | IEEE Conference Publication | IEEE Xplore
- 2317288 (dergipark.org.tr)
- Uluslararası Yönetim Biliřim Sistemleri ve Bilgisayar Bilimleri Dergisi » Makale » Kripto Para Birimlerinin Twitter Verileri ile Metin Madencilięi Kapsamında İncelenmesi (dergipark.org.tr)
- Microsoft Word - SON OKUMA.docx (dergipark.org.tr)
- RPubS - Duygu Analizi Üzerine Bir Vaka Çalışması
- R-Ladies (Türkiye) Webinar Serisi 5: tidytext ile Metin Madencilięi ve Duygu Analizi (youtube.com)
- 2025797 (dergipark.org.tr)