

T.C BİLECİK ŞEYH EDEBALI ÜNİVERSİTESİ

İKTİSADİ VE İDARİ BİLİMLER FAKÜLTESİ

YÖNETİM BİLİŞİM SİSTEMLERİ



VERİ MADENCİLİĞİ

SPOTİFY DUYGU ANALİZİ

ÖZGE YILMAZ

İÇİNDEKİLER

1. Giriş
2. Veri Madenciliği Nedir?
 - 2.1. Veri Madenciliği Süreci
3. Yöntem
 - 3.1. Metin Madenciliği
 - 3.2. Seçilen Konu ve Kullanılan Veri Seti
4. Projeye Başlarken Kullanılan R Kütüphaneleri
5. Veri Analizi
 - 5.1. Yapılan Çıkarımlar
 - 5.2. Duygu Analizi
6. Sonuç
7. Kaynakça

GİRİŞ

Günümüzde müzik, insanların duygularını ifade etmeleri ve başkalarına ulaşmaları için önemli bir araçtır. Spotify gibi müzik platformları, kullanıcıların binlerce şarkıya erişmesini sağlar. Bu proje, Spotify şarkılarının sözlerinden şarkıların genel duygu durumlarını (pozitif, negatif, nötr) çıkartmak amacıyla yapılacaktır. Şarkı sözleri, bir şarkının duygusal tonunu anlamak için önemli bir kaynaktır. Bu projede, Spotify verisinden elde edilen şarkı sözleri üzerinden metin madenciliği ve duygu analizi yapılacaktır. Proje, veri temizleme, kelime tokenizasyonu, duygu analizi ve sonuçların görselleştirilmesi adımlarından oluşmaktadır. Sonuç olarak, şarkı sözlerinin duygusal yapısı analiz edilerek, şarkıların genel ruh hali hakkında çıkarımlar yapılacaktır.

VERİ MADENCİLİĞİ NEDİR?



Veri madenciliği, büyük ve karmaşık veri setlerinden değerli bilgiler elde etme, desenler keşfetme ve anlamlı sonuçlar çıkarma sürecidir. Bu süreç, istatistiksel analiz, makine öğrenimi, veritabanı yönetimi ve veri görselleştirme gibi alanları kapsar. Veri madenciliği, verilerdeki gizli ilişkileri, eğilimleri ve kalıpları tespit etmek için kullanılır. Elde edilen bu bilgiler, organizasyonların daha bilinçli kararlar almasına, operasyonel verimliliği artırmaya ve stratejik planlamalar yapmasına olanak sağlar.

Veri Madenciliği Süreci:

Veri Toplama (Data Collection): Bu aşama, analiz için gerekli verilerin toplanmasını içerir. Veriler, şirket içi veritabanlarından, açık veri kaynaklarından, internetten veya sensörlerden toplanabilir. Toplanan veriler, doğru analizler yapabilmek için kullanılabilir hâle getirilir.

Veri Temizleme (Data Cleaning): Ham veriler genellikle eksik, hatalı ya da tutarsız bilgiler içerebilir. Bu nedenle, veri temizleme, hatalı verilerin düzeltilmesi, eksikliklerin giderilmesi ve gereksiz verilerin çıkarılması sürecidir.

Veri Dönüştürme (Data Transformation): Veriler, analiz için uygun formatlara dönüştürülür. Veri dönüştürme, verinin işlenebilir hale getirilmesi için gereklidir. Bu aşama, farklı veri türlerinin uyumlu hale getirilmesini ve verilerin analiz için optimize edilmesini sağlar.

Veri Depolama (Data Storage): Veri madenciliği sürecinde, veriler güvenli ve erişilebilir bir şekilde depolanır. Veriler genellikle büyük veri platformlarında, bulut sistemlerinde ya da yerel veri depolama alanlarında saklanır.

Veri Analizi (Data Analysis): Veri analizinin temel amacı, verilerdeki kalıpları ve ilişkileri tespit etmektir. İstatistiksel analizler, makine öğrenimi algoritmaları ve modelleme teknikleri kullanılarak veriler analiz edilir.

Veri Görselleştirme (Data Visualization): Sonuçlar, grafikler, tablolar ve haritalar aracılığıyla görselleştirilir. Veri görselleştirme, analiz edilen verilerin daha kolay anlaşılmasını ve içgörülerin hızlı bir şekilde elde edilmesini sağlar.

YÖNTEM

METİN MADENCİLİĞİ

Metin madenciliği, büyük miktarda metin verisinden anlamlı bilgiler çıkarma ve ilişkileri keşfetme sürecidir. Metin madenciliği, doğal dil işleme (NLP), makine öğrenimi ve istatistiksel analiz tekniklerini birleştirerek, metin tabanlı verilerdeki gizli bilgileri ortaya çıkarır. Bu süreç, metinlerdeki kalıpları, trendleri ve önemli noktaları analiz etmek için kullanılır. Günümüzde metin madenciliği, müşteri geri bildirimlerini analiz etme, sosyal medya yorumlarını inceleme, duygu analizi yapma, belge sınıflandırması ve konu modelleme gibi geniş bir uygulama alanına sahiptir.

SEÇİLEN KONU VE KULLANILAN VERİ SETİ

Müzik, duygusal ifadeyi güçlü bir şekilde aktarabilen ve toplumsal etkileri derin olan bir sanat dalıdır. Dijital platformlar, sanatçılara eserlerini geniş kitlelere ulaştırma imkânı sunar. Bu proje, Spotify veritabanındaki şarkı sözleri üzerinden yapılan metin madenciliği ile sanatçının şarkılarının duygu durumlarını (pozitif, negatif, nötr) analiz etmeyi amaçlamaktadır. Bu analiz, şarkı sözlerindeki anahtar kelimeler ve duygusal tonların belirlenmesini içerir. Bu şekilde, sanatçının müziği ile dinleyici kitlesi arasındaki etkileşim ve şarkılarının toplumsal ve kültürel yankıları derinlemesine incelenerek, müziğin dinleyici üzerindeki etkisi değerlendirilecektir.



MANAS_1245_AGRawal · UPDATED 9 MONTHS AGO



0

New Notebook



Download



spotify



2015-2024 yılları arasında yayınlanan pop şarkılarının şarkı sözleri verisi

Toplam veri sayısı: 6000

Çalışılan veri sayısı: 922

PROJEYE BAŞLARKEN KULLANILAN R KÜTÜPHANELERİ

```
# Gerekli kutuphaneleri yukleyin
install.packages("tm")
install.packages("textclean")
install.packages("tidyverse")
install.packages("wordcloud")
install.packages("syuzhet")
install.packages("quanteda")
install.packages("stringr")
install.packages("tidytext")
install.packages("text2vec")
install.packages("ggplot2")
install.packages("writexl")
install.packages("sentimentr")
install.packages("lubridate")
install.packages("igraph")
install.packages("topicmodels")
install.packages("RTextTools")
install.packages("text")
install.packages("tidyr")
install.packages("caret")
install.packages("dplyr")
install.packages("RColorBrewer")
install.packages("wordcloud2")
install.packages("openxlsx")

library(tm)
library(textclean)
library(tidyverse)
library(wordcloud)
library(syuzhet)
library(quanteda)
library(stringr)
library(tidytext)
library(text2vec)

# Metin madenciligi icin temel paket
# Metin temizleme islemleri icin
# Veri isleme ve gorsellestirme icin
# Kelime bulutlari olusturma
# Duygu analizi
# Metin analizi icin gelismis araclar
# Metin manipulasyonu
# Metin analizi icin tidy format
# Derin metin analizi
# Gorsellestirme ic'in
# Excel dosyasD1 yazma
# Duygu analizi
# Zaman verisi manipulasyonu
# grafikteki aracları gorsellestirme
# Konu modelleme
# Metin sD1nD1flandD1rma
# NLP icin modern araclar
# Veri bicimleme
# Makine ogrenimi araclarD1

# Metin madenciligi paketini yukle
# Metin temizleme islemi icin
# Veri isleme ve gorsellestirme
# Kelime bulutlari
# Duygu analizi yapmak icin
# Metin analizi ve modelleme
# Metin manipulasyonu
# Tidy text islemleri
# Derin metin analizi
```

R PROGRAMLAMA DA KULLANILAN KODLAR

- İlk olarak hazır veri dosyamızı R 'a yüklüyoruz. Gerekli olan sütunları seçip yeni olan dosyayı bilgisayarımıza kaydediyoruz.

```
data <- read.csv("C:/Users/Asus/Desktop/train.csv")

data_filtered <- data %>%
  mutate(release_year = year(ymd(track_album_release_date))) %>%
  filter(release_year >= 2015 & release_year <= 2024)

train_selected <- data_filtered %>%
  select(lyrics, playlist_genre, track_album_release_date)

write.xlsx(train_selected, "selected_train.xlsx")
```

- Daha sonra temizleme işlemlerini yapıp analiz yapacağımız şarkı türünü seçip tekrar temiz veriyi kaydediyoruz.

```
train_selected$lyrics <- trimws(train_selected$lyrics)
train_selected$lyrics <- tolower(train_selected$lyrics)
train_selected$lyrics <- gsub("[[:punct:]]", "", train_selected$lyrics)
train_selected$lyrics <- gsub("[0-9]", "", train_selected$lyrics)
train_selected$lyrics <- removewords(train_selected$lyrics, stopwords("en"))

write.xlsx(train_selected, "cleaned_train.xlsx")

train_pop <- train_selected %>% filter(playlist_genre == "pop")
write.xlsx(train_pop, "train_pop.xlsx")
```

- Veri setinden istenmeyen ve gereksiz kelimelerin temizlenme işlemini yapıyoruz.

```
custom_stop_words <- c("yeah", "oh", "na", "and", "but", "or", "the", "im", "que", "just", "got", "la",
  "de", "te", "youre", "wanna", "aint", "dont", "ooh", "el", "yo", "en", "ya", "lo", "hey",
  "ddi", "ddim", "ddits", "ddif", "if", "ah", "es", "tu", "ddive", "se", "si", "mi", "bitch", "uh",
  "em", "shit", "fuck", "ddit", "ddin", "don", "version", "taylor", "ddis", "song", "ll", "una", "dif",
  "LL", "if", "im", "dont", "yeah", "ooh", "go", "can", "cause", "know", "now", "una", "ive", "us", "one", "take", "let", "way", "much", "litte")
stop_words_extended <- bind_rows(stop_words, tibble(word = custom_stop_words, lexicon = "custom"))
```

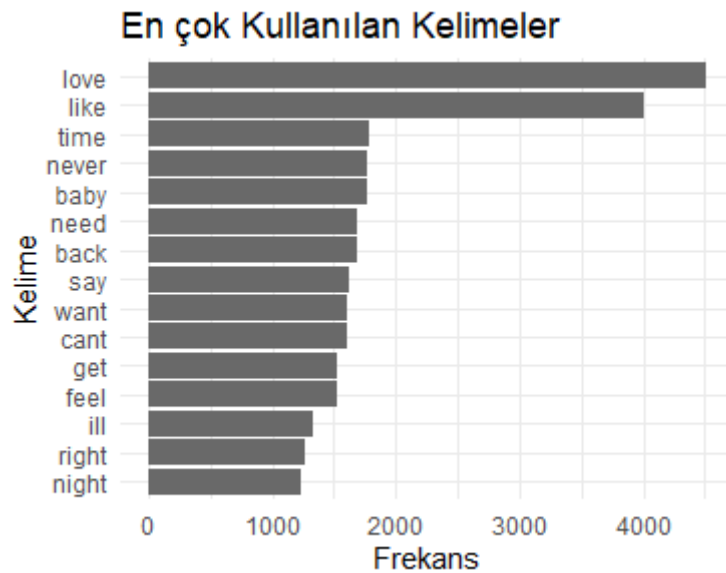
- Aşağıdaki komut ile veri setimizin kelime frekansı analizini yapıyoruz.

```

filtered_word_counts <- train_pop %>%
  unnest_tokens(word, lyrics) %>%
  mutate(word = tolower(word)) %>%
  anti_join(tibble(word = custom_stop_words), by = "word") %>%
  count(word, sort = TRUE)

filtered_word_counts %>%
  slice_max(n, n = 15) %>%
  ggplot(aes(reorder(word, n), n)) +
  geom_bar(stat = "identity", fill = "dimgray") +
  coord_flip() +
  labs(
    title = "En çok kullanılan kelimeler",
    x = "kelime",
    y = "Frekans"
  ) +
  theme_minimal()

```



YAPILAN ÇIKARIMLAR:

Aşk ve İlişkiler Teması Öne Çıkıyor:

- "Love", "like", "baby", "need", "want" gibi kelimelerin sıklıkla kullanılması, pop şarkılarında aşk ve ilişkiler temasının çok baskın olduğunu gösteriyor.
- Bu kelimeler, duygusal bir bağ kurma, romantizm ya da özlem gibi duygusal temaları yansıtır.

Zaman ve Anlam Arayışı:

- "Time" ve "never" kelimelerinin popülerliği, şarkı sözlerinde zamanın akışı, geçmişe ya da geleceğe yönelik düşüncelerin önemli bir yer kapladığını gösteriyor.

İçsel ve Duygusal Durumlar:

- "Feel" ve "ill" kelimeleri, bireylerin içsel duygularını, sağlık veya ruh haliyle ilgili ifadelerini anlatmak için kullanıldığını gösteriyor. Pop müzik, genellikle bireyin hislerini ifade eden sözlere sahiptir.

Geceler ve Anlık Duygular:

- "Night" kelimesinin sık kullanılması, pop şarkılarında gecenin romantizm, eğlence ya da melankoli ile ilişkilendirildiğini düşündürüyor.

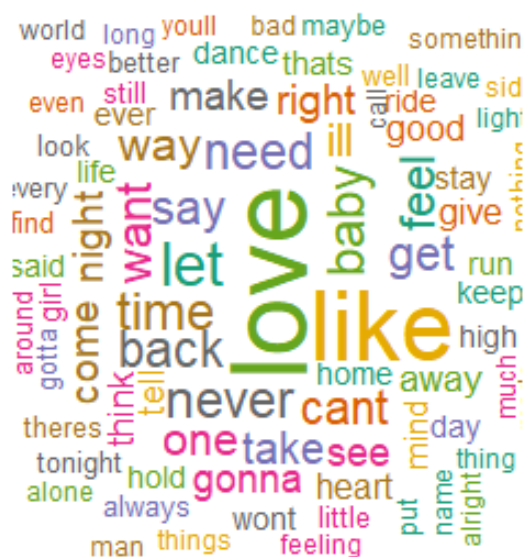
Geri Dönüş ve Bağlantı Arayışı:

- "Back" kelimesi, bir şeye ya da birine dönmeyi ifade edebilir. Bu, pişmanlık, geçmişe özlem veya yeniden bağlanma temasını yansıtıyor olabilir.

- Daha sonra kelime bulutumuzu oluşturuyoruz.

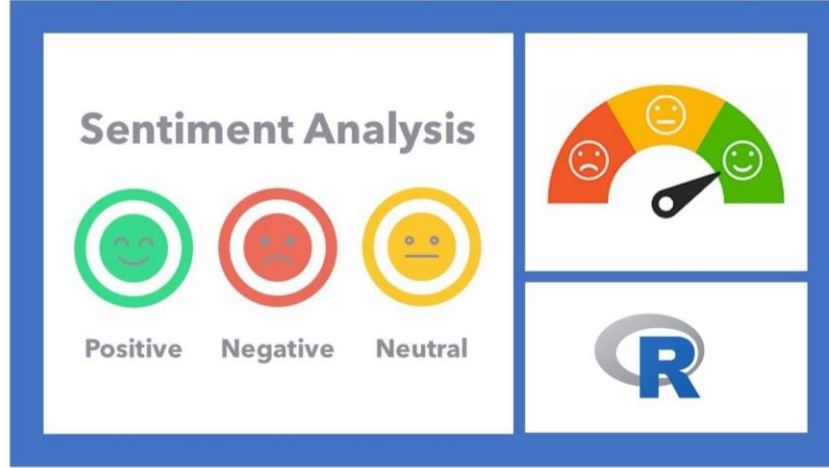
```
par(mar = c(1, 1, 1, 1))

wordcloud(
  words = filtered_word_counts$word,
  freq = filtered_word_counts$n,
  min.freq = 10,
  max.words = 80,
  scale = c(3, 0.5),
  colors = brewer.pal(8, "Dark2"),
  random.color = TRUE,
  random.order = FALSE,
  rot.per = 0.3
)
```



DUYGU ANALİZİ

Duygu analizi, bir metindeki duygusal tonları veya hissiyatları (örneğin, pozitif, negatif veya nötr duygular) belirlemek için kullanılan bir doğal dil işleme (NLP) tekniğidir. Bu analiz, kullanıcı yorumları, sosyal medya gönderileri, ürün incelemeleri ve şarkı sözleri gibi metin tabanlı veriler üzerinde yapılır.



- Duygu analizi yapıp bu analiz sonucunu pozitif, negatif ve nötr duyguların oranlarını görselleştiriyoruz.

```
nrc <- get_sentiments("nrc")

sonuc <- train_pop %>%
  mutate(linenummer = row_number()) %>%
  unnest_tokens(word, lyrics) %>%
  inner_join(nrc, by = c("word" = "word")) %>%
  group_by(linenummer) %>%
  summarise(sentiment_score = sum(ifelse(sentiment == "positive", 1,
                                         ifelse(sentiment == "negative", -1, 0)))) %>%
  left_join(train_pop %>% mutate(linenummer = row_number()), by = "linenummer")

neutral <- length(which(sonuc$sentiment_score == 0))
positive <- length(which(sonuc$sentiment_score > 0))
negative <- length(which(sonuc$sentiment_score < 0))

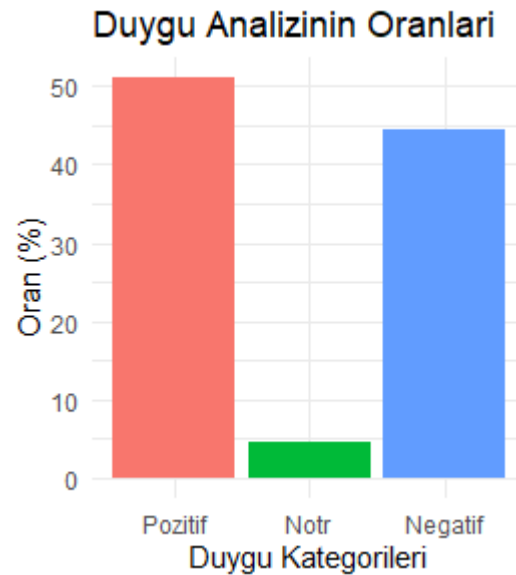
toplamlam <- positive + neutral + negative

Duygular <- c("Pozitif", "Nötr", "Negatif")
Degerler <- c((positive / toplamlam) * 100, (neutral / toplamlam) * 100, (negative / toplamlam) * 100)

output <- data.frame(Duygular, Degerler)

output$Duygular <- factor(output$Duygular, levels = Duygular)
```

```
ggplot(output, aes(x = Duygular, y = Degerler)) +
  geom_bar(stat = "identity", aes(fill = Duygular), show.legend = FALSE) +
  ggtitle("Duygu Analizinin Oranlari") +
  theme_minimal() +
  xlab("Duygu Kategorileri") +
  ylab("Oran (%)")
```

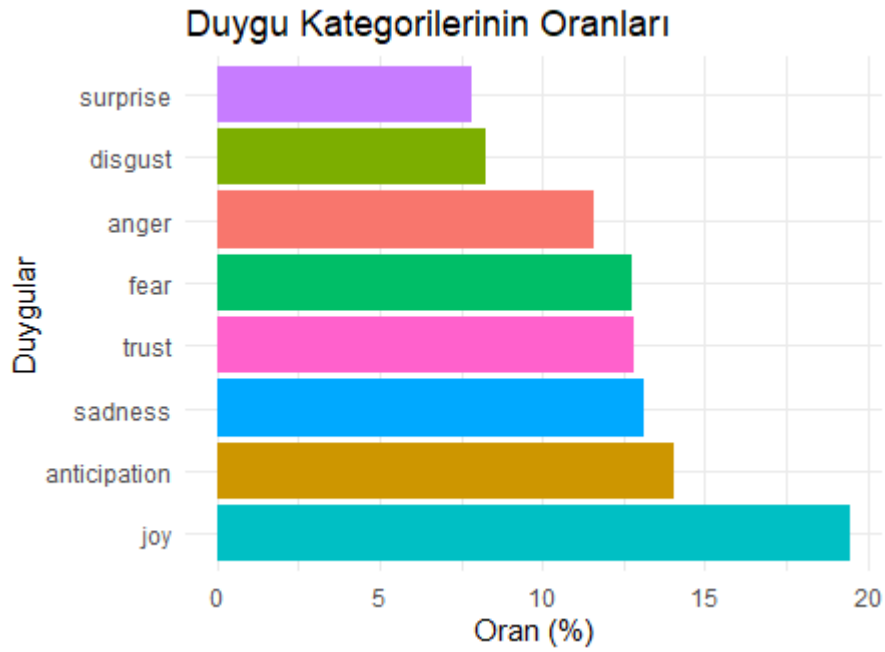


- Daha sonra belirli duyguları (neşe, öfke, şaşkınlık, korku, üzüntü, tikslenme, güven, beklenti) çıkarıp, her bir duygunun şarkı sözlerinde ne kadar yer aldığını hesaplayıp oranlarını görselleştiren bir bar grafiği oluşturuyoruz.

```
nrc_8 <- get_sentiments("nrc") %>%
  filter(sentiment %in% c("joy", "anger", "surprise", "fear", "sadness", "disgust", "trust", "anticipation"))

sonuc_8 <- train_pop %>%
  unnest_tokens(word, lyrics) %>%
  inner_join(nrc_8, by = "word") %>%
  count(sentiment) %>%
  mutate(Oran = n / sum(n) * 100)

ggplot(sonuc_8, aes(y = reorder(sentiment, -Oran), x = Oran, fill = sentiment)) +
  geom_bar(stat = "identity", show.legend = FALSE) +
  labs(title = "Duygu Kategorilerinin Oranları", y = "Duygular", x = "Oran (%)") +
  theme_minimal()
```

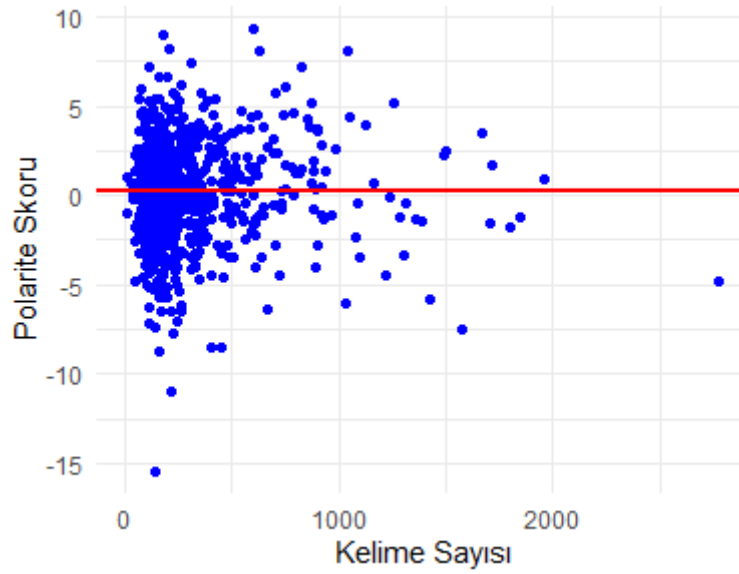


- Polarite dağılımını hesaplayıp grafikleştiriyoruz.

```
polarite <- get_sentiment(train_pop$lyrics, method = "syuzhet")
train_pop$word_count <- sapply(train_pop$lyrics, function(x) str_count(x, "\\w+"))

tablo <- data.frame(word_count = train_pop$word_count, polarite = polarite)

ggplot(tablo, aes(word_count, polarite)) +
  geom_point(color="blue") +
  geom_hline(yintercept = mean(tablo$polarite), color="red", linewidth=1) +
  labs(y= "Polarite Skoru", x= "Kelime Sayısı") +
  theme_minimal() +
  labs() +
  theme(plot.caption = element_text(hjust = 0, face = "italic"))
```



- Yapılan duygu analizinin matematiksel sonucu aşağıda belirtilmiştir. Bu tablodaki “mean” kısmı duygu analizimizin matematiksel karşılığıdır. “mean” pozitif ve negatif sonuçlarının ortalamasını belirtir.

```
mean_polarite <- mean(tablo$polarite, na.rm = TRUE)
se_mean <- sd(tablo$polarite, na.rm = TRUE) / sqrt(length(tablo$polarite))
CI_mean_0.95 <- mean_polarite + c(-1, 1) * qt(0.975, df = length(tablo$polarite) - 1) * se_mean
var_polarite <- var(tablo$polarite, na.rm = TRUE)
std_dev_polarite <- sd(tablo$polarite, na.rm = TRUE)
coef_var <- (std_dev_polarite / mean_polarite) * 100

stat_summary <- data.frame(
  mean = mean_polarite,
  se.mean = se_mean,
  CI.mean.0.95 = CI_mean_0.95[2] - CI_mean_0.95[1], #
  var = var_polarite,
  std.dev = std_dev_polarite,
  coef.var = coef_var
)

print(stat_summary)
```

mean	se.mean	CI.mean.0.95	var	std.dev	coef.var
0.2626898	0.09022073	0.3541241	7.504877	2.739503	1042.866

POLARİTE GRAFİĞİ ÇIKARIMLARI:

1. Polarite Eğilimi:

- Kırmızı çizgi (ortalama polarite), genelde pozitif bir polariteye yakınsa, bu durum şarkıların çoğunun olumlu duygular taşıdığını doğrular. Ancak, grafikte mavi noktaların negatif polariteye doğru bir kısmının olduğunu da görüyoruz. Bu da negatif duygular taşıyan şarkıların az olmadığını, ancak pozitif duyguların baskın olduğunu gösterir.

2. Yoğunluk Alanı:

- Grafik üzerindeki mavi noktaların büyük çoğunluğu polarite skorlarının **0 ile 5 arasında** yoğunlaştığını gösteriyor. Bu, pozitif polaritenin baskın olduğunu açıklar.
- Ayrıca, az sayıda şarkının polarite skorlarının sıfır civarında (nötr) olduğunu fark edebilirsiniz, çünkü bu bölgede yoğunluk daha düşük.

3. Negatif Polarite:

- Negatif polarite skorlarının (-5 ile -10 arasında) bulunduğu mavi noktalar da dikkate değer. Bu, negatif duygular taşıyan şarkıların varlığını doğrular, ancak sayılarının pozitif duygulara kıyasla daha az olduğunu gösterir.

SONUÇ

Bu projede, Spotify şarkı sözlerinden elde edilen veriler üzerinden metin madenciliği ve duygu analizi gerçekleştirilmiştir. Veri madenciliği süreci, veri toplama, temizleme, dönüştürme, depolama, analiz ve görselleştirme gibi temel aşamaları kapsayacak şekilde yapılandırılmıştır. Çalışmada, toplam 6000 şarkı verisinden elde edilen şarkı sözlerinden 922'si analiz için uygun hale getirilmiş ve bu veri seti üzerinden detaylı incelemeler yapılmıştır.

Analiz sonucunda, şarkı sözlerinde öne çıkan temaların başında aşk, ilişkiler, zaman ve duygusal durumlar gelmektedir. Özellikle sık kullanılan kelimelerin frekans analizi, pop müzik şarkılarının dinleyicileriyle nasıl bir bağ kurduğunu anlamak için önemli bilgiler sunmuştur. "Love", "baby", "time" ve "feel" gibi kelimelerin yüksek frekansta yer alması, pop şarkılarının duygusal bir bağ kurma, romantizm ve özlem gibi temaları yoğun bir şekilde işlediğini göstermiştir. Bununla birlikte, zaman kavramını ifade eden "time" ve "never" kelimelerinin popülerliği, şarkıların geçmişe veya geleceğe yönelik düşünceler içerdiğini ortaya koymaktadır.

Duygu analizi sonuçlarına göre, şarkıların genel olarak pozitif bir polariteye sahip olduğu, yani olumlu duygular içerdiği tespit edilmiştir. Ancak analizler, belirli bir oranda nötr ve negatif duyguların da şarkı sözlerinde yer aldığını göstermektedir. Bu durum, pop müziğin yalnızca olumlu duyguları yansıtmaktan öte, daha karmaşık ve çeşitlilik içeren bir duygusal yapıya sahip olduğunu ortaya koymaktadır. Polarite grafikleri, pozitif duyguların şarkılarda baskın

olduğunu; bununla birlikte, melankoli, özlem ve pişmanlık gibi daha karanlık temaların da azımsanmayacak bir yer tuttuğunu göstermiştir.

Grafik analizleri ayrıca, kelimelerin uzunluğu ile frekansları arasındaki ilişkiyi de ortaya koymuştur. Daha uzun kelimelerin frekansının daha düşük olduğu ve bu nedenle daha az kullanıldığı gözlemlenmiştir. Bu durum, uzun kelimelerin daha teknik veya karmaşık yapıya sahip olmasından kaynaklanabilir. Analiz edilen şarkı sözlerinde kullanılan kelimelerin genel olarak kısa ve anlaşılır olduğu, bu durumun şarkıların daha geniş bir dinleyici kitlesine hitap etme amacıyla bağlantılı olabileceği değerlendirilmiştir.

Bu projede elde edilen bulgular, metin madenciliği tekniklerinin müzik analizinde ne denli etkili ve faydalı olduğunu ortaya koymuştur. Şarkı sözlerinin duygusal yapısı, sanatçılar ve müzik platformları için stratejik karar alma süreçlerinde kullanılabilecek değerli içgörüler sunmaktadır. Örneğin, şarkı öneri sistemlerinin geliştirilmesi, dinleyici alışkanlıklarının analiz edilmesi veya yeni albüm çalışmalarında hedef kitleye yönelik içerik oluşturulması gibi pek çok farklı alanda bu analizlerden faydalanılabilir.

Proje kapsamında yapılan analizler, Spotify gibi dijital müzik platformlarının kullanıcı davranışlarını daha iyi anlamalarına ve kullanıcı deneyimini iyileştirmelerine de katkı sağlayabilir. Ayrıca sanatçılar, şarkı sözlerindeki temalar ve duygular üzerinden dinleyicileriyle nasıl bir bağ kurduklarını anlamak ve gelecekteki projelerinde bu bağları güçlendirmek için bu tür analizlerden yararlanabilirler.

Sonuç olarak, bu çalışma, metin madenciliği ve duygu analizi tekniklerinin yalnızca büyük veri setlerinde değil, aynı zamanda daha odaklı ve tematik çalışmalar için de etkili olduğunu göstermektedir. İlerleyen dönemlerde daha geniş veri setleri, gelişmiş algoritmalar ve farklı analiz teknikleri kullanılarak, müziğin duygusal etkileri ve toplumsal yansımaları üzerine daha kapsamlı araştırmalar yapılabilir. Bu tür çalışmalar, müzik endüstrisinin geleceğini şekillendirebilecek stratejik bilgiler sunmanın yanı sıra, dinleyici deneyimini daha kişiselleştirilmiş ve anlamlı hale getirebilir.

KAYNAKÇA

<https://dergipark.org.tr/tr/pub/uybisbbd/issue/71032/1089670>

<https://aws.amazon.com/tr/what-is/data-mining/>

<https://mis.sadievrenseker.com/2014/06/metin-madenciligi-text-mining/>

<https://artiwise.com/tr/2024/04/04/duygu-analizi-nedir-kullanim-alanlari-ve-zorluklari/>

<https://www.veribilimiokulu.com/r-ile-metin-madenciligi-bolum-1/>

<https://chatgpt.com/>