**T.C**

**BİLECİK ŞEYH EDEBALİ ÜNİVERSİTESİ**

**FEN FAKÜLTESİ**

**İSTATİSTİK ve BİLGİSAYAR BİLİMLERİ BÖLÜMÜ**



**2022-2023 BAHAR DÖNEMİ**

**VERİ MADENCİLİĞİNE RAPORU**

**Dünya'daki Kuraklık Sorunu YouTube Yorumları**

**Veri Madenciliği Analizi**

**KADİR GÜRKAN ALTIPARMAK**

**BİLECİK - 2023**

**İÇİNDEKİLER**

A) Giriş 3

1) Kullanılan Kütüphaneler4

B) YouTube’dan R Programlama Dili ile Veri Madenciliği 4

C) Sonuç13

**D) Kaynakça14**

**BİLECİK – 2023**

**ÖZET**

Bu rapor, Dünya'daki kuraklık sorunuyla ilgili YouTube yorumlarının R programıyla çekilmesi ve veri madenciliği yöntemleriyle analiz edilmesinin özetini sunmaktadır. Çalışmada, belirli bir YouTube videosuna yapılan toplam 11556 yorumun analizi gerçekleştirilmiştir. Amacımız, insanların kuraklık hakkında düşüncelerini bazı analiz yöntemleriyle değerlendirmek ve çıkarımlar elde etmektir.

1. **Giriş**

Günümüzde dünya genelinde giderek artan bir tehdit olan kuraklık, su kaynaklarının azalması ve iklim değişikliği gibi faktörlerle birlikte insanlığın karşılaştığı önemli bir sorundur. Kuraklık, tarımsal üretimi etkileyerek gıda güvenliği sorunlarına yol açmakta, ekosistemleri tehdit etmekte ve su kaynaklarının sürdürülebilirliğini tehlikeye sokmaktadır. Bu nedenle, kuraklıkla mücadele stratejilerinin geliştirilmesi ve farkındalık oluşturulması büyük bir önem taşımaktadır.

Bu bağlamda, sosyal medya platformları, insanların düşüncelerini, duygusal tepkilerini ve bilgi paylaşımlarını ifade etmek için önemli bir kanal haline gelmiştir. YouTube gibi popüler video paylaşım platformları, milyonlarca kullanıcının görüşlerini paylaşabileceği bir ortam sunmaktadır. Bu kullanıcı yorumları, geniş bir veri kaynağı oluşturarak kuraklık hakkındaki insan düşüncelerini anlamamızı sağlar.

Bu rapor, YouTube'da paylaşılan bir kuraklık videosuna yapılan yorumları kullanarak veri madenciliği yöntemlerini uygulamayı hedeflemektedir. Yorumları toplamak ve analiz etmek için R programlama dili ve ilgili paketlerden faydalanacağız. Bu çalışmanın amacı, insanların kuraklık konusundaki düşüncelerini belirlemek, duygusal tepkilerini analiz etmek ve farklı temaları ortaya çıkarmaktır.

Raporumuzda, veri temizleme adımlarından başlayarak metin madenciliği ve duyarlılık analizi gibi teknikleri kullanarak yorumları derinlemesine inceleyeceğiz. En sık kullanılan kelimeleri belirleyerek, kuraklıkla ilgili anahtar kavramları ve vurgulanması gereken noktaları ortaya çıkaracağız. Ayrıca, kelime bulutları ve grafikler aracılığıyla verileri görselleştirerek raporumuza daha fazla etkileyicilik katacağız.

Bu çalışmanın sonuçları, kuraklıkla mücadele eden topluluklar, bilim insanları, politika yapıcılar ve ilgili paydaşlar için değerli bilgiler sunabilir. Elde edilecek veriler, kuraklıkla ilgili politika kararları alınırken ve kaynakların etkin bir şekilde kullanılması için stratejiler geliştirilirken rehberlik edebilir. Ayrıca, insanların duygusal tepkilerini anlamak, kuraklık konusunda farkındalık oluşturma çalışmalarının başarısını artırabilir ve halkın bu konuda daha bilinçli bir şekilde hare ket etmesine katkı sağlayabilir. Veri madenciliği yöntemleriyle elde edilen bulgular, kuraklıkla mücadelede öncelikli alanları belirlemek, kaynakları doğru bir şekilde yönlendirmek ve stratejileri etkin bir şekilde uygulamak için önemli bir rehberlik sağlayabilir.

Ayrıca, raporumuzda kullanılan metin madenciliği teknikleri, büyük veri analizi alanında da uygulanabilir bir yöntem sunmaktadır. Sosyal medya platformlarından elde edilen veriler, çeşitli konularda kamuoyu görüşlerini anlamak, trendleri belirlemek ve kullanıcı davranışlarını analiz etmek için değerli bir kaynak olabilir. Bu çalışma, metin madenciliği yöntemlerinin sosyal medya verileri üzerinde nasıl uygulanabileceğine dair bir örnek sunmaktadır.

Sonuç olarak, bu çalışma kuraklıkla ilgili YouTube yorumlarından elde edilen verileri kullanarak metin madenciliği yöntemlerini uygulamayı ve bu verilerden çıkarılan bilgilerle kuraklıkla mücadelede rehberlik edecek sonuçlar elde etmeyi hedeflemektedir. Bu çalışmanın, kuraklıkla mücadele eden topluluklara, bilim insanlarına, politika yapıcılara ve ilgili paydaşlara önemli bir perspektif sunacağına inanıyoruz. Ayrıca, metin madenciliği tekniklerinin sosyal medya verileri üzerindeki uygulamalarının daha geniş bir alanda kullanılabileceğine dikkat çekmek istiyoruz.

1. **Kullanılan Kütüphaneler**

* tuber: YouTube API'sı ile etkileşim kurmak için kullanılır.
* tm: Metin madenciliği işlemleri için kullanılır.
* SnowballC: Metin kökü çıkarma işlemi yapmak için kullanılır.
* wordcloud: Kelime bulutu oluşturmak için kullanılır.
* RColorBrewer: Renk paletleri oluşturmak için kullanılır.
* syuzhet: Duygusal analiz yapmak için kullanılır.
* ggplot2: Veri görselleştirme için kullanılır.
* stringr: Metin işleme işlemleri için kullanılır.

1. **YouTube’dan R Programlama Dili ile Veri Madenciliği**

YouTube APİ Key tanımlanması;

api <- "1039146937954-v02c7svvtsdml396ajsi8iptopljr95l.apps.googleusercontent.com"

apsecret <- "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"

yt\_oauth(api,apsecret,token="")

Yukarıdaki kod parçacığıyla YouTube API kimlik bilgileri yazılır ve yetkilendirme yapılır. Böylece yt\_oauth() fonksiyonuyla YouTube'a bağlanabilirsiniz.

YouTube API anahtarlarını almak için aşağıdaki adımları takip edebilirsiniz:

Google Cloud Platform (GCP) hesabı oluşturun veya mevcut hesabınıza giriş yapın. GCP, YouTube API'lerine erişim sağlamak için kullanılacaktır.

* GCP Console'a erişmek için şu adrese gidin: <https://console.cloud.google.com/>
* Sol üst köşede bulunan "Proje Oluştur" düğmesine tıklayarak yeni bir proje oluşturun. Proje adını ve proje kimliğini (Project ID) belirleyin.
* GCP Console'un üst kısmındaki "API ve Hizmetler" bölümüne gidin.
* "API ve Hizmetler'e Git" düğmesine tıklayarak "API ve Hizmetler Kütüphanesi" sayfasına gidin.
* "API Kütüphanesini Keşfet" düğmesine tıklayın ve "YouTube Data API v3"ü arayın.
* "YouTube Data API v3" öğesini bulun ve etkinleştirmek için üzerine tıklayın.
* Sol taraftaki menüden "Kimlik Bilgileri" bölümüne gidin.
* "Kimlik Bilgileri Oluştur" düğmesine tıklayarak bir kimlik bilgileri seti oluşturun.
* Kimlik bilgileri oluştururken, API anahtarlarından "OAuth 2.0 İstemci Kimlik Bilgileri"ni seçin.

Kimlik bilgileri oluşturduktan sonra, API anahtarınızı ve gizli anahtarınızı (secret key) elde edeceksiniz. API anahtarınız, YouTube API'larına erişmek için kullanılırken, gizli anahtarınız, kimlik doğrulama işlemleri için gereklidir.

Bu adımları izleyerek YouTube API anahtarlarınızı elde edebilir ve kullanabilirsiniz. Ancak, API anahtarlarının kullanımı ve güvenliği konusunda dikkatli olmanız önemlidir. Anahtarlarınızı güvende tutmak ve yetkisiz erişime karşı korumak için gerekli önlemleri almanız gerekmektedir.

YOUTUBE Video Yorumları Çekme:

comments <- get\_all\_comments("HStCv8ixyWg")

kod parçacağında Noktalı işaretler arasına; ([https://www.youtube.com/watch?v=**HStCv8ixyWg**](https://www.youtube.com/watch?v=HStCv8ixyWg))

YouTube videosu linkinin kalın harflerle gösterilen kısmını yazıp, yorumları bu sayede çekmeye başlayabiliriz.

Çekilen Yorumları CSV Formatında Kaydetme:

getwd()

write.csv(comments, "comments.csv", row.names = FALSE)

kod parçacağıyla burada getwd() fonksiyonu ile hangi dosya konumunun açık olduğuna bakıp yorumlarımızı kaydedebiliriz.

Veri Setini İşleme;

yorumlar <- read.table(file.choose(),header=T,sep=",")

head(yorumlar$textOriginal)

Öncelikle csv formatında kaydettiğimiz verimizi çağırıp ilk 6 sırasına bakalım.

Veri Setini Temizleme;

yorumlar <- yorumlar[!( duplicated(yorumlar$textOriginal)),]

Tekrar eden yorumları temizlemek için bu fonksiyonu kullanabiliriz.

emoji\_pattern <- "[\U{1F600}-\U{1F6FF}]|[\U{2600}-\U{26FF}]|[\U{2700}-\U{27BF}]|[\U{1F300}-\U{1F5FF}]|[\U{1F680}-\U{1F6FF}]|[\U{1F1E0}-\U{1F1FF}]|[\U{1F900}-\U{1F9FF}]|[\U{1F200}-\U{1F2FF}]|[\U{1F4C0}-\U{1F4FF}]|[\U{1F600}-\U{1F64F}]|[\U{1F680}-\U{1F6FF}]|[\U{1F1E6}-\U{1F1FF}]"

yorumlar$temiz\_metin <- str\_replace\_all(yorumlar, emoji\_pattern, "")

temiz\_metin adlı bir sütun oluşturup yorumlardaki emojileri temizledik.

yorumlar$temiz\_metin <- tolower(yorumlar$textOriginal)

Yorumları küçük harfe çevirdik.

yorumlar$temiz\_metin <- str\_replace\_all(yorumlar$temiz\_metin, "[^[:alnum:] ]", "")

Yorumlardaki özel karakterlerin ve noktalama işaretlerinin analiz üzerindeki etkisi azaltılır ve daha temiz bir metin elde edilir.

#yorumlar$temiz\_metin <- removeNumbers(yorumlar$temiz\_metin)

Eğer çektiğiniz yorumlardaki sayılar analizinizi etkilemeyecekse bu fonksiyon sayesinde sayıları kaldırabilirsiniz.

Noktalama işaretlerini kaldıralım.

yorumlar$temiz\_metin <- removePunctuation(yorumlar$temiz\_metin)

Boşlukları temizleyelim.

yorumlar$temiz\_metin <- stripWhitespace(yorumlar$temiz\_metin)

Durak kelimeleri temizleyelim.

stop\_kelimeler <- c("acaba", "ama", "ancak", "artık", "aslında", "aynen", "az", "bana", "bazen", "belki", "ben", "beni", "benim", "bile", "bir", "biraz", "birçoğu", "birçok", "birkaç", "biz", "bizden", "bize", "bizi", "bizim", "bu", "buna", "bunda", "bundan", "bunu", "bunun", "burada", "böyle", "çoğu", "çoğuna", "çoğunu", "çok", "çünkü", "da", "daha", "de", "defa", "diğer", "diye", "eğer", "en", "gibi", "hem", "hep", "hepsi", "hepsine", "hepsini", "her", "her biri", "herhangi", "herkes", "herkese", "herkesi", "herkesin", "hiç", "hiçbir", "için", "ile", "ise", "kez", "ki", "kim", "kime", "kimi", "kimin", "mu", "mü", "nasıl", "ne", "neden", "nerde", "nerede", "nereden", "niye", "niçin", "o", "on", "ona", "ondan", "onlar", "onlara", "onlardan", "onları", "onların", "onu", "oysa", "oysaki", "önce", "önceki", "rağmen", "sadece", "sanki", "sen", "siz", "sizden", "size", "sizi", "sizin", "şey", "şeyden", "şeye", "şeyi", "şeyler", "şunu", "tabii", "tamam", "tüm", "tümü", "ve", "veya", "ya", "yalnız", "yada", "yani", "yerine", "yine", "zaten", "abi", "kadar", "barış", "video", "olarak", "şu", "bence", "olur", "olacak", "yil", "olan", "mı", "videoyu", "lan", "özcan", "öyle", "bişey", "bey", "yada", "birşey", "göre", "işte", "fakat", "olduğu", "tabi", "fakat", "bunlar", "olduğu")

yorumlar$temiz\_metin <- removeWords(yorumlar$temiz\_metin, words = stop\_kelimeler)

yorumlar$temiz\_metin <- gsub("\\s+", " ", yorumlar$temiz\_metin)

Yorumlardaki ardışık boşluk karakterlerini tek bir boşluk karakteriyle değiştirelim.

#yorumlar$temiz\_metin <- iconv(yorumlar$temiz\_metin, from = "UTF-8", to = "ASCII//TRANSLIT")

Yorumlarda eğer Türkçe karakter istemiyorsak bu fonksiyonu kullanabiliriz.

En Çok Tekrar Eden Kelimeler;

Metin Belgelerinin Oluşturulması

m\_belge <- Corpus(VectorSource(yorumlar$temiz\_metin))

Terim-Belge Matrisi Oluşturulması

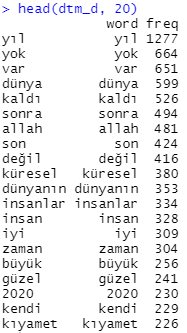
m\_belge\_dtm <- TermDocumentMatrix(m\_belge)

dtm\_m <- as.matrix(m\_belge\_dtm

Frekansın azalan değerine göre sıralayıp ardından tekrar büyükten küçüğe çıktı almak için terse sıralayalım ve en çok tekrar eden 20 kelimeyi gösterelim.

dtm\_v <- sort(rowSums(dtm\_m),decreasing=TRUE)

dtm\_d <- data.frame(word = names(dtm\_v),freq=dtm\_v)

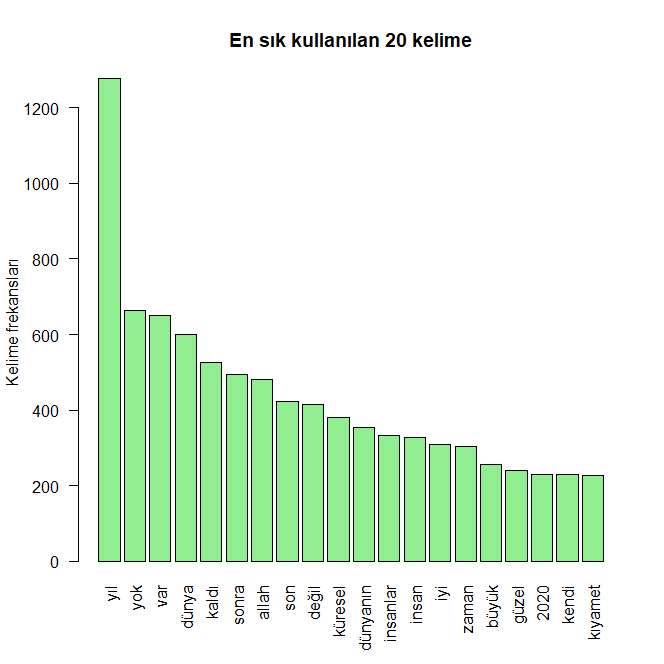


En çok tekrar eden 20 kelimenin bir de grafiğine bakalım.

barplot(dtm\_d[1:20,]$freq, las = 2, names.arg = dtm\_d[1:20,]$word,

col ="lightgreen", main ="En sık kullanılan 20 kelime",

ylab = "Kelime frekansları")



Kelime Bulutu;

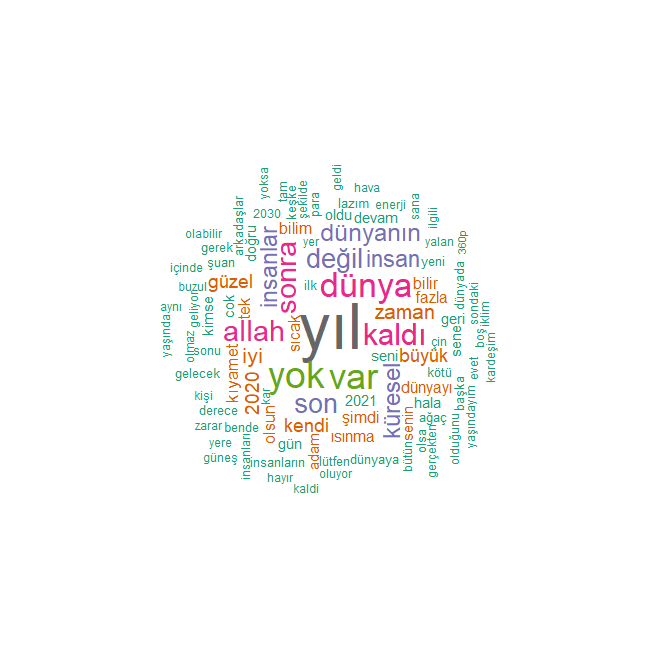
set.seed(1234)

wordcloud(words = dtm\_d$word, freq = dtm\_d$freq, min.freq = 5,

max.words=100, random.order=FALSE, rot.per=0.40,

colors=brewer.pal(8, "Dark2"))

Çıktı olarak da;

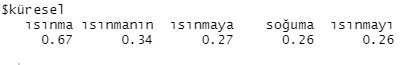


Terimleri Çağrıştıran Kelimeler;

İstediğimiz terimlerin diğer terimlerle olan ilişkisini araştıralım.

findAssocs(m\_belge\_dtm, terms = c("dünya", "küresel"), corlimit = 0.25)

Çıktı olarak da;





Duygu Analizi;

sentiment <- get\_nrc\_sentiment(yorumlar$temiz\_metin)

head(sentiment,20)

# transpose

td<-data.frame(t(sentiment))

# RowSums işlevi, bir gruplama değişkeninin her düzeyi için satırlar boyunca sütun toplamlarını hesaplar.

td\_new <- data.frame(rowSums(td[2:10723]))

# Dönüşüm ve temizlik yapalım

names(td\_new)[1] <- "count"

td\_new <- cbind("sentiment" = rownames(td\_new), td\_new)

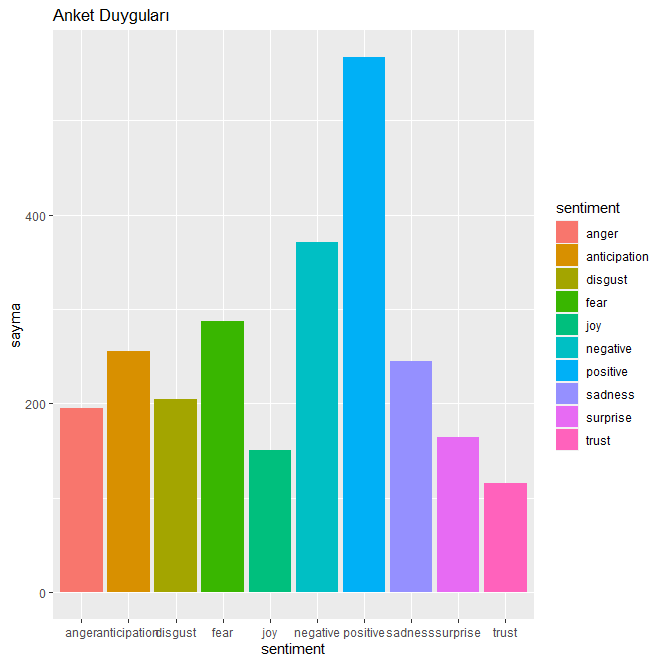
rownames(td\_new) <- NULL

td\_new2<-td\_new[1:10,]

Duygu analizi grafiği için;

quickplot(sentiment, data=td\_new2, weight=count, geom="bar", fill=sentiment, ylab="sayma")+ggtitle("Anket Duyguları")

Duygu analizini sütun grafiğiyle inceleyelim.



Bir de yüzde olarak sütun grafiğinde inceleyelim.

barplot(

sort(colSums(prop.table(sentiment[, 1:10]))),

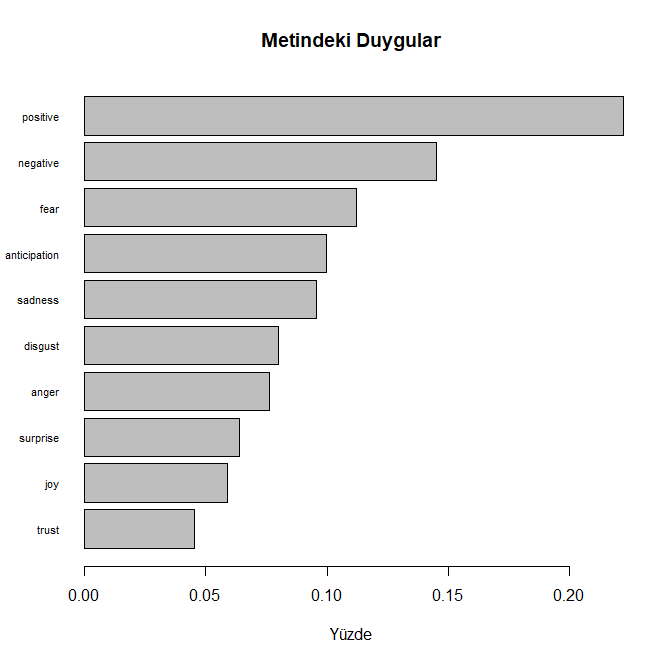
horiz = TRUE,

cex.names = 0.7,

las = 1,

main = "Metindeki Duygular", xlab="Yüzde"

)



1. **Sonuç**

Bu raporda, YouTube'da kuraklık hakkında yapılan yorumları analiz ettik. İlk olarak, yorumları temizleyerek gereksiz karakterleri ve kelimeleri çıkardık. Ardından, en sık kullanılan kelimeleri ve kelime bulutunu görselleştirdik. Daha sonra, yorumlardaki duygusal kategorileri analiz ederek duygusal skorları çıkardık ve bu skorları çubuk grafiğiyle gösterdik. Sonuçlarımız, yorumların genel olarak olumsuz eğilimli olduğunu, kaygı ve endişe duygularının yaygın olduğunu gösterdi. Bu analiz, YouTube kullanıcılarının kuraklık konusunda endişelerini dile getirdikleri bir platform olduğunu göstermektedir. Kuraklıkla ilgili yapılan yorumları anlamak, ilgili taraflar için değerli bilgiler sunabilir ve ilgili politika ve stratejilerin oluşturulmasına katkıda bulunabilir.

Bu çalışma, kuraklıkla mücadelede önemli bir adım olarak karşımıza çıkmaktadır. Elde ettiğimiz veriler ve analiz sonuçları, kuraklıkla mücadele eden topluluklar, bilim insanları, politika yapıcılar ve ilgili paydaşlar için değerli bir kaynak sunmaktadır. Bu çalışma, kuraklıkla mücadele stratejilerinin geliştirilmesi ve kaynakların etkin bir şekilde yönetilmesi konusunda rehberlik edebilir.

Sonuçlarımız, kuraklıkla mücadelede önleyici tedbirlerin ve sürdürülebilir su kaynaklarının yönetiminin önemini vurgulamaktadır. Ayrıca, kuraklık olaylarının sosyal ve ekonomik etkilerini anlamak ve insanların duygusal tepkilerini değerlendirmek, bilinçlendirme çalışmalarının etkinliğini artırabilir. Bu çalışma, halkın kuraklık konusunda daha bilinçli bir şekilde hareket etmesini sağlamak için iletişim stratejilerinin ve eğitim programlarının geliştirilmesine katkıda bulunabilir.

Ancak, bu çalışma sınırlı bir örnekleme ve belirli bir coğrafi alana odaklanmıştır. Genellemeler yapmadan önce daha geniş kapsamlı araştırmaların yapılması önemlidir. Ayrıca, kuraklıkla mücadele konusunda politika kararları alınırken ve stratejiler geliştirilirken, bilimsel verilere dayalı yaklaşımların ve uzman görüşlerinin dikkate alınması önemlidir.

Sonuç olarak, bu çalışma, kuraklıkla mücadelede bilimsel araştırmanın önemini vurgulayan ve ilgili paydaşlara değerli bir kaynak sunan bir adımdır. Bu çalışmanın sonuçları, gelecekteki çalışmalara, politika yapıcılarına ve topluluklara yönelik daha kapsamlı kuraklık stratejileri ve çözümleri geliştirmek için bir temel sağlamaktadır.

En son olarak raporu birkaç sloganla bitirmek istiyorum;

* Kuraklık, dünyamızın karşı karşıya olduğu büyük bir tehdittir. Toprakların susuz kalması, tarım alanlarının verimsizleşmesi, su kaynaklarının azalması... Bu sorunlarla baş etmek için hepimize görev düşüyor.
* Su kaynaklarımızı korumak için hepimizin birlikte hareket etmesi önemli. Bilinçli tüketim alışkanlıkları, suyu verimli kullanan teknolojilerin tercih edilmesi, toplumsal farkındalığın artırılması...
* Unutmayalım, gelecek nesillerin suya ihtiyacı var. Onlara temiz su ve sağlıklı bir çevre bırakabilmek için bugünden adım atalım.

1. **Kaynakça**

[**https://www.youtube.com/watch?v=HStCv8ixyWg&list=WL&index=1**](https://www.youtube.com/watch?v=HStCv8ixyWg&list=WL&index=1)

[**https://medium.com/@tumuhimbisemoses/youtube-data-mining-experiences-of-using-r-1419e9aefcc5**](https://medium.com/@tumuhimbisemoses/youtube-data-mining-experiences-of-using-r-1419e9aefcc5)

[**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/)

[**https://chat.openai.com/**](https://chat.openai.com/)

[**https://www.youtube.com/watch?v=mfD0FY9oHGc&list=WL&index=2&t=360s**](https://www.youtube.com/watch?v=mfD0FY9oHGc&list=WL&index=2&t=360s)