**T.C.**

**BİLECİK ŞEYH EDEBALİ ÜNİVERSİTESİ**

**İKTİSADİ VE İDARİ BİLİMLER FAKÜLTESİ**

**YÖNETİM BİLİŞİM SİSTEMLERİ**

**VERİ MADENCİLİĞİ**

**2024 AMERİKA SEÇİMİ**

Hazırlayan

FERİT TALHA GÜN

GÖREV RAPORU

**ÖZET**

**VERİ MADENCİLİĞİ 2024 AMERİKA SEÇİMİ**

Görev özeti, 2024 amerika seçimlerinin sonuçlarının olası tahminini içermektedir. Yapılan görev hazırlayanın yetenek ve istatistiksel bakış açısına göre hazırlanmıştır. Projede R programlama dili kullanılmıştır. Yeterli deneyim ile daha iyi sonuçlar elde edilir.

1. GİRİŞ

Veri, bilgilerin depolandığı, işlendiği ve analiz edildiği temel bir öğedir. Veri, genellikle sayısal veya sembolik formatta bilgileri temsil eder ve organizasyonlar, işletmeler ve bireyler tarafından bir dizi farklı kaynaktan elde edilir. Veri üretimi, bu bilgilerin toplanması, oluşturulması veya kaydedilmesi sürecini ifade eder. Bu süreç, sensörler, kullanıcı etkileşimleri, yazılım kayıtları, anketler veya diğer veri toplama yöntemleri aracılığıyla gerçekleşebilir. Veri üretimi, ardından veri depolama, temizleme, analiz ve yorumlama adımlarıyla devam eder. Veri üretimi ve yönetimi, bilgi tabanlı kararlar almak, desenleri belirlemek, trendleri takip etmek ve iş süreçlerini iyileştirmek gibi bir dizi hedefe hizmet eder. Bu süreç, modern bilgi çağında, işletmelerin rekabet avantajı elde etmelerine ve daha bilinçli kararlar almalarına olanak tanır.

2024 Amerika seçimleri, demokrasinin önemli bir unsuru olup, seçim süreçlerinin anlaşılması ve değerlendirilmesi için veri madenciliği teknikleri ile analiz edilmektedir. Bu çalışma, R programlama dili ve veri madenciliği araçları kullanılarak, 2024 Amerika seçimleriyle ilgili kapsamlı bir veri analizi gerçekleştirmeyi amaçlamaktadır. İlk olarak, seçimle ilgili çeşitli veri kaynaklarından elde edilen veriler toplanacak ve bir veri seti oluşturulacaktır. Bu veri seti, adayların seçim kampanyaları, halkın duygusal tepkileri, sosyal medya aktiviteleri ve anket sonuçları gibi çeşitli faktörleri içerecek şekilde tasarlanacaktır. Daha sonra, R programlama dili ve veri madenciliği kütüphaneleri kullanılarak veri seti analiz edilecek. İstatistiksel analizler, veri görselleştirmeleri ve duygu analizi gibi teknikler, seçim sürecinin dinamiklerini anlamak için uygulanacaktır. Adayların popülerliği, seçmen davranışları, kampanya stratejileri ve diğer önemli faktörler üzerinde derinlemesine bir anlayış sağlamak için çeşitli analizler gerçekleştirilecektir. Sonuçlar, veri madenciliği yöntemleri kullanılarak elde edilen bulguların ve trendlerin seçim sürecinin anlaşılmasına katkıda bulunup bulunmadığını değerlendirecek ve bu analizlerin gelecekteki seçim stratejileri üzerindeki etkisini değerlendirecektir. Bu çalışma, veri madenciliği ve R programlama dilinin seçim analizi alanında nasıl etkili bir şekilde kullanılabileceğini göstererek, politika yapıcıları, analistler ve araştırmacılar için değerli bir kaynak sağlamayı amaçlamaktadır.

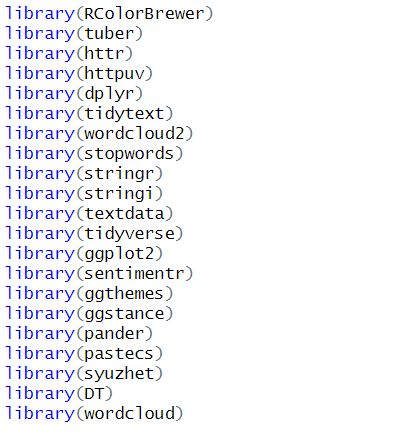
1. R Programlama Dili

R, istatistiksel hesaplamalar, veri analizi ve grafik çizimleri gibi veri bilimi alanlarında kullanılan açık kaynaklı bir programlama dilidir. Belli bir vektör tabanlı dil olan R, özellikle veri setlerinin manipülasyonu ve istatistiksel analiz için güçlü bir araçtır. İstatistikçiler, veri bilimciler ve araştırmacılar, R'ı geniş bir paket koleksiyonu ve istatistiksel fonksiyonları ile kullanarak, veri madenciliği ve analitik projelerde etkili bir şekilde çalışabilirler. Ayrıca, zengin grafik yetenekleri sayesinde sonuçları görselleştirmek için de yaygın olarak tercih edilmektedir. R, açık kaynaklı olması nedeniyle geniş bir topluluk tarafından desteklenir ve sürekli olarak geliştirilmektedir

* 1. Youtube Developer

YouTube Developer Hesabı, YouTube API'lerini kullanarak YouTube platformuyla entegrasyon yapmak isteyen geliştiricilerin ihtiyaç duyduğu bir hesaptır. YouTube API'leri, geliştiricilere YouTube'un çeşitli özelliklerine programatik olarak erişme ve uygulamalarını YouTube platformuyla entegre etme imkanı sağlar.

* 1. Kullanılan Kütüphaneler

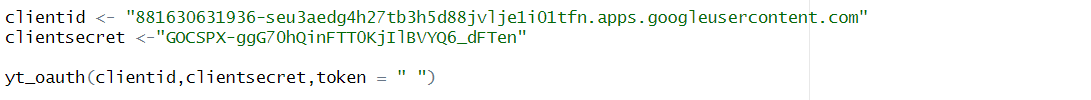


**Şekil 1**: R dilinde kullanılan kütüphaneler

.

* 1. R Programlamada Kullanılan Kodlar

R programlama dilinde aktif hale getirilen kütüphanelerden sonra R programlama dili ile Youtube Developer hesabını birbirlerine bağlanıp Youtube’dan veri çekmek için uygun ortam Şekil 2’ deki gibi hazırlamıştır.



**Şekil 2**: Youtube Developer Hesabı ile R bağlantısı

Hangi videodan veri çekeceği ve çekilen yorumların csv formatına dönüştürülmesi, Şekil 3’teki gibi belirtilmiştir.



**Şekil 3**: Video anahtarının ve yorum verilerinin csv dönüşümü

Çekilen yorumların analize uygun olması amacıyla temizleme işlemi yapılmıştır.



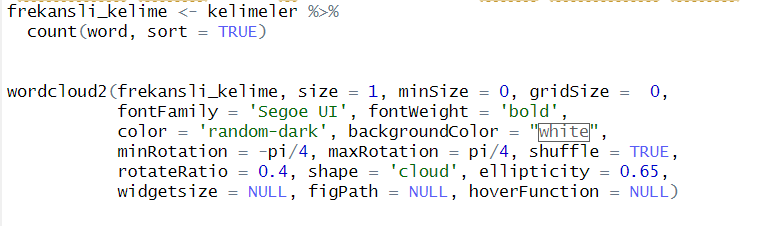
**Şekil 4**: Yorumların Temizlenmesi

Temizlenmiş yorumları ayrıştırarak kelimeler isimli tabloya dönüştürülerek durak kelimeleri ve son temizlik işlemlerinin yapılması.



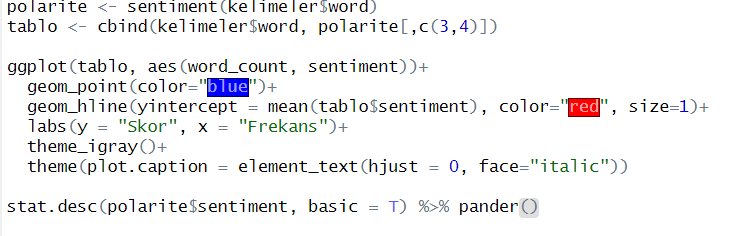
**Şekil 5**: Kelime tablosuna dönüştürülüp temizlenmesi

Hazırlanan veriyi anlamlı bir hale getirmek ve en çok tekrar eden kelimeleri bulmak için Şekil 6’ daki kelime bulutunu oluşturmak için komut yazılmıştır.



Şekil 6: Kelime frekanslarıyla kelime bulutunun oluşturulması

Yorumlar sonucunda elde edilen kelimelerin polarite skoru şekil 7’deki gibidir.



Şekil 7: Polarite Analizi

1. Bulgular

3.1 Kelime Bulutu

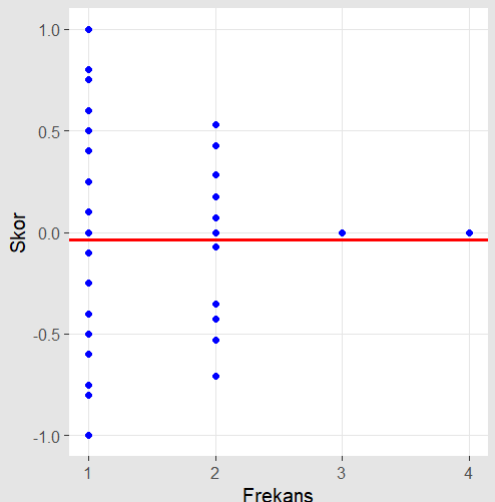
Konuyu güncel olarak ele alan videodan alınan verilerle ortaya çıkan kelime bulutu Şekil 8’deki gibidir. Elde edilen yorum sayısı 20021’dir. Bu kelime bulutuna göre en çok tekrar eden kelime Trump’tır. Ardından Biden gelmektedir.



Şekil 8: Kelime Bulutu

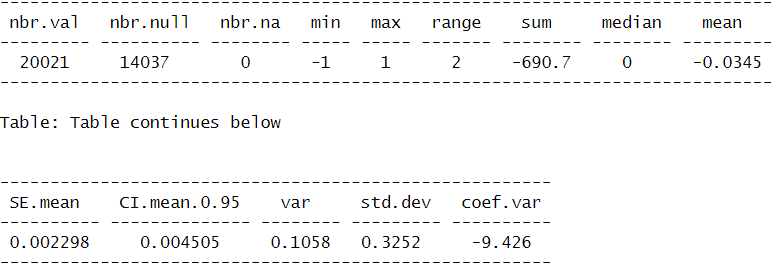
3.2 Duygu Analizi

Seçim sonuçlarının duygu durumu Şekil 9’da oluşturulmuştur. +1 değerine yaklaşan değerler pozitif yönlü, 0 nötr ve -1’e yaklaşan değerler negatif duygu yönünü göstermektedir. Analiz sonucuna göre duygu analizi -0.0345 değeri ile 0’a yakınlığından dolayı nötr çıkmış, 0 değerini biraz aşmasından dolayı negatif yönelimli çıkmıştır yorumu yapılabilir. Buradaki kırmızı çizgi duyguyu belirtmektedir. Maviler ise her kelimenin -1 ile 1 arasında ki duygu yönünü belirtmektedir.



Şekil 9: Duygu Analizi Sonucu

Analizin matematiksel karşılığı Şekil 10’da verilmiştir.



Şekil 10: Matematiksel Sonuç

4. SONUÇ VE TARTIŞMA

Seçimdeki popüler taraflar zamanla değişmekte olup, adayların medya yüzleriyle fikirler ve düşünceler değişime uğramıştır. Zamanında Trump devrini bitiren Biden sosyal medya duruşu sebebiyle trend olma özelliğini yitirerek yavaş yavaş insanların Trump’a yönelmesine sebep olmuştur. Bu sebeple en çok tekrar eden kelime Trump olmuştur. Fakat sırasıyla çok tekrar eden kelimelere bakıldığında insanların demokrat bir yönetim istediği çıkarımını yapabiliriz. Bu nedenle Cumhuriyetçi olan Trump için her ne kadar kazanıyor gibi gözüksede zor bir seçim beklediğini söylemek mümkündür. Çıkan ortalama sonucuna göre negatif yönlü bir duygu gelişimi olduğunu söyleyebiliriz. Bu da Trump’a olan güvenin yavaşça azaldığını söyleyebiliriz.

Çalışma en çok yorum alan ve iki adayında çalkantılı durumlarda olduğu dönemin video yorumlarından alınmıştır. 2542 yorum ve 20021 kelime kullanılmıştır. Daha güncel durumlar ele alındığında çıkan sonuç büyük değişiklikler gösterebilir. Araştırma Global bir haber sitesinden alındığı için iki tarafı sevmeyen farklı ülke vatandaşlarıda tepkilerini göstermişlerdir. Daha yerel anketlerle daha iyi sonuçlar elde edinebilinir.