



**İKTİSADİ İDARİ VE SOSYAL
BİLİMLER FAKÜLTESİ**

MAYIS 2018
TURKİYE POLİTİK GÖRÜŞÜ
VERİ İNCELEMESİ
DÖNEM ÖDEVİ FİNAL PROJESİ

Kaan HACİÖMEROĞLU
Furkan KÜTÜK
Halil İbrahim ALTUNTAŞ

Dr. Öğr. Üyesi Nesibe MANAV

İstanbul, Ocak 2022

İÇİNDEKİLER

Özet

Giriş

1.Bölüm Veri Anlama

- 1.1 Veri Setini Okutma
- 1.2 Veri Setindeki Soruların Karşılıkları
- 1.3 Veri Setini Factor Cinsine Çevirme

2.Bölüm Veri Hazırlama

- 2.1 Parti İsimlerinin Anonime Çevrilmesi
- 2.2 Veri Setinin Özeti
- 2.3 Partilerin Yüzdelik Olarak Değerleri
- 2.4 Veri Seti Eksik Değer Sorgulaması
- 2.5 En Çok Tekrar Eden Veri
 - 2.5.1 En Çok Katılım Sağlayan Bölge
 - 2.5.2 En Çok Katılım Sağlayan Yaş Aralığı
 - 2.5.3 En Çok Katılım Sağlayan Eğitimin Durumu
 - 2.5.4 Anket Sonucu En Çok Desteklenen Parti

3.Bölüm Veri Görselleştirme

3.1 Soruların Grafikleri

3.1.1 Sizce Ekonomik Durumumuz İyi Mi?

3.1.2 Eğitimde Reforma İhtiyaç Var Mı?

3.1.3 Destek Kurumlarının Özelleştirilmesini Destekliyor Musunuz?

3.1.4 Devlet Belli Suçlar İçin İdam Cezası Getirmeli Mi?

3.1.5 Gazetecileri Yeterince Tarafsız Buluyor Musunuz?

3.1.6 Saat 22:00'den Sonra İçki Alma Yasağını Destekliyor Musunuz?

3.1.7 Laik Bir Devlette Yaşamak İster Misiniz?

3.1.8 Kürtaj Yasağını Destekliyor Musunuz?

3.1.9 Olağanüstü Halin (Ohal) Özgürlükleri Kısıtladığını Düşünüyor Musunuz?

3.1.10 Meclise Yeni Bir Partinin Katılmasını İster Misiniz?

3.2 Karşılaştırma Grafikleri

3.2.1 Bölgelere Göre Eğitim Durumu Grafiği

3.2.2 Bölgelere Göre Partilere Verilen Destek Grafiği

3.2.3 Partilerin Destek Aldığı Yas Grupları

3.2.4 Eğitim Durumuna Göre Partilere Verilen Destek Grafiği

3.3 Partilere Verilen Destek Sayısı Grafiği

4.Bölüm Veri Modelleme

4.1 Eğitim ve Test Seti Oluşturma

4.2 Basit (Naive) Bayes Algoritması

4.2.1 Naive Bayes Modeli Oluşturma

4.2.2 Performans Değerlendirmesi

4.3 Karar Ağacı Algoritması

4.3.1 Karar Ağacı Modeli Oluşturma

4.3.2 Karar Ağacı Grafiği

4.3.3 Performans Değerlendirmesi

4.4 K-Means Algoritması

4.4.1 K-Means İçin Veri Hazırlama

4.4.2 K-Means İçin Veriyi Anlama

4.4.3 K-Means İçin Optimal Küme Sayısı Belirlenmesi

4.4.4 K-Means Modeli Oluşturma

4.4.5 K-Means Modelinin Grafiği

4.4.5.1 K-Means Modelinin Grafiğinin İncelenmesi

Sonuç

Değerlendirme

KAYNAKÇA

Özet

Veri setimizde sorular, cinsiyet, bölgeler ve yaş aralığına bağlı olarak desteklenen partileri görebilmekteyiz. Bu kategorilere göre veriyi sırasıyla Veriyi Anlama, Veriyi Hazırlama, Veriyi Görselleştirme ve Veriyi Modelleme başlıkları içerisinde çalışmalarını yaptık.

Veriyi Anlama evresinde kullanacağımız veri setini okutup ilerideki evreler için verinin anlamını yükseltmeye çalıştık.

Veriyi Hazırlama evresinde görselleştirme ve modelleri daha rahat uygulayabilmek adına çalışmaları gerçekleştirdik.

Veriyi Görselleştirme evresinde veri başlıkları arasında grafiklerle karşılaştırma ve incelemesini yaptık.

Veriyi Modelleme evresinde Naive Bayes, Karar Ağacı ve K-Means modelleriyle inceledik.

Giriş

Büyük veri, geleneksel veri tabanı tekniklerinin kullanılması suretiyle işlenmesi mümkün olmayan, farklı hacimlerdeki heterojen veriyi tanımlayan yeni bir kavramdır ve çeşitli dijital içeriklerden oluşmaktadır.(AKTAN,2018,s3)

Kullandığımız “Türkiye Politik Görüşü” veri seti 2018 Mayıs ayında dijital anket yoluyla oluşturulmuştur.

Rstudio üzerinden seçtiğimiz veri setini incelemeye başlayacağız. Gerekli kütüphaneler kullanılarak Veriyi Anlama ve Veriyi Hazırlama başlıklarından geçirdikten sonra veri setini görselleştirip gerekli karşılaştırmaları yaparak anlatımını kolaylaştıracğız. Veri seti üzerinde grafik yardımı ile anlatım ve modelleme sonucunda modelimizin doğruluk oranını tespit etmeye çalışacağız. Modelleme için Naive Bayes, Karar Ağacı ve K-Means modellerini kullanacağız modellemeler sonucunda model performans değerlendirmesini inceleyeceğiz.

1.Bölüm Veri Anlama

1.1 Veri Setini Okutma

```
library(readr)
data <- read_csv("yonelimfinal.csv")
view(data)
data <- data [-1] # VERİDEN TARİH SUTUNUNU SİLMEK
```

Şekil 1 Veri Okuma

	Cinsiyet	Yas	Bolge	Egitim	soru1	soru2	soru3	soru4	soru5	soru6	soru7	soru8	soru9	soru10	parti
1	Erkek	18-30	Marmara	Lisans	Hayır	Evet	Evet	Hayır	Hayır	Hayır	Evet	Hayır	Evet	Evet	Parti5
2	Erkek	18-30	Marmara	Lisans	Evet	Evet	Hayır	Evet	Hayır	Evet	Hayır	Evet	Hayır	Hayır	Parti1
3	Kadın	18-30	Marmara	Lisans	Hayır	Evet	Hayır	Hayır	Hayır	Hayır	Evet	Hayır	Evet	Evet	Parti3
4	Erkek	0-18	Marmara	Lise	Evet	Evet	Hayır	Evet	Hayır	Evet	Evet	Evet	Evet	Evet	Parti5
5	Erkek	18-30	Marmara	Ön Lisans	Hayır	Evet	Evet	Evet	Hayır	Hayır	Evet	Evet	Hayır	Hayır	Parti5
6	Erkek	60+	Marmara	Lisans Üstü	Hayır	Hayır	Hayır	Hayır	Hayır	Hayır	Hayır	Hayır	Hayır	Hayır	Parti4
7	Kadın	18-30	Marmara	Lisans	Evet	Evet	Evet	Hayır	Hayır	Evet	Evet	Evet	Evet	Evet	Parti3
8	Erkek	0-18	Güneydoğu	Lise	Evet	Evet	Evet	Evet	Hayır	Hayır	Evet	Evet	Evet	Evet	Parti2
9	Erkek	0-18	Akdeniz	Lise	Hayır	Evet	Evet	Hayır	Evet	Hayır	Evet	Hayır	Evet	Evet	Parti5
10	Erkek	0-18	Doğu Anadolu	Lise	Evet	Evet	Hayır	Hayır	Hayır	Hayır	Evet	Hayır	Hayır	Hayır	Parti2
11	Erkek	18-30	İç Anadolu	Lisans	Hayır	Evet	Evet	Evet	Hayır	Hayır	Evet	Hayır	Evet	Evet	Parti2
12	Erkek	18-30	Marmara	Lise	Hayır	Evet	Evet	Hayır	Hayır	Hayır	Evet	Hayır	Evet	Hayır	Parti3
13	Erkek	18-30	İç Anadolu	Lisans	Hayır	Evet	Evet	Evet	Hayır	Evet	Evet	Hayır	Evet	Hayır	Parti1
14	Erkek	0-18	Akdeniz	Lise	Evet	Evet	Hayır	Evet	Hayır	Evet	Hayır	Evet	Hayır	Hayır	Parti5
15	Erkek	18-30	Akdeniz	Lisans	Hayır	Evet	Evet	Evet	Hayır	Evet	Evet	Hayır	Evet	Evet	Parti2
16	Erkek	18-30	Güneydoğu	Ön Lisans	Hayır	Evet	Evet	Hayır	Hayır	Hayır	Evet	Hayır	Evet	Evet	Parti2
17	Erkek	18-30	Karadeniz	Lisans	Hayır	Evet	Evet	Evet	Hayır	Evet	Evet	Hayır	Evet	Hayır	Parti1
18	Kadın	18-30	Marmara	Lisans	Evet	Evet	Hayır	Evet	Hayır	Hayır	Evet	Hayır	Evet	Evet	Parti2

Şekil 2 Veri Seti

Veri setimiz 15 sütun 885 satırdan oluşmaktadır. Veri seti içerisinde Cinsiyet, Yaş, Eğitim durumu, Sorular ve Parti isimleri gibi başlıklardan oluşmaktadır.

1.2 Veri Setindeki Soruların Karşılıkları

Soru 1: Sizce Ekonomik Durumumuz İyi Mi?

Soru 2: Eğitimde Reforma İhtiyaç Var Mı?

Soru 3: Özelleştirmeyi Destekliyor Musunuz?

Soru 4: Devlet Belli Suçlar İçin İdam Cezası Getirmeli Mi?

Soru 5: Gazetecilerimizi Yeterince Tarafsız Buluyor Musunuz?

Soru 6: Saat 22:00'den Sonra İçki Alma Yasağını Destekliyor Musunuz?

Soru 7: Laik Bir Devlette Yaşamak İster Misiniz?

Soru 8: Kürtaj Yasağını Destekliyor Musunuz?

Soru 9: Olağanüstü Halin (OHAL) Özgürlükleri Kısıtladığını Düşünüyor Musunuz?

Soru 10: Meclise Yeni Bir Bölümün Girmesini İster Misiniz?

1.3 Veri Setini Factor Cinsine Çevirme

```
install.packages("dplyr")
library(dplyr)
data <- mutate_if(data, is.character, as.factor) # TÜM VERİ SETİNİ FACTORE ÇEVİRMEK
str(data) # VERİ SETİNİN TIPLERİNİN KONTROLÜ
```

Şekil 3 Veri Tipi Değiştirme

```
> str(data) # VERİ SETİNİN TIPLERİNİN KONTROLÜ
tibble [885 x 15] (S3: tbl_df/tbl/data.frame)
 $ cinsiyet: Factor w/ 2 levels "Erkek","Kadın": 1 1 2 1 1 1 2 1 1 1 ...
 $ Yas     : Factor w/ 5 levels "0-18","18-30",...: 2 2 2 1 2 5 2 1 1 1 ...
 $ Bolge   : Factor w/ 7 levels "Akdeniz","Doğu Anadolu",...: 7 7 7 7 7 7 7 4 1 2 ...
 $ Egitim  : Factor w/ 6 levels "İlkokul","Lisans",...: 2 2 2 4 6 3 2 4 4 4 ...
 $ soru1   : Factor w/ 2 levels "Evet","Hayır": 2 1 2 1 2 2 1 1 2 1 ...
 $ soru2   : Factor w/ 2 levels "Evet","Hayır": 1 1 1 1 1 2 1 1 1 1 ...
 $ soru3   : Factor w/ 2 levels "Evet","Hayır": 1 2 2 2 1 2 1 1 1 2 ...
 $ soru4   : Factor w/ 2 levels "Evet","Hayır": 2 1 2 1 1 2 2 1 2 2 ...
 $ soru5   : Factor w/ 2 levels "Evet","Hayır": 2 2 2 2 2 2 2 2 1 2 ...
 $ soru6   : Factor w/ 2 levels "Evet","Hayır": 2 1 2 1 2 2 1 2 2 2 ...
 $ soru7   : Factor w/ 2 levels "Evet","Hayır": 1 2 1 1 1 2 1 1 1 1 ...
 $ soru8   : Factor w/ 2 levels "Evet","Hayır": 2 1 2 1 1 2 1 1 2 2 ...
 $ soru9   : Factor w/ 2 levels "Evet","Hayır": 1 2 1 1 2 2 1 1 1 2 ...
 $ soru10  : Factor w/ 2 levels "Evet","Hayır": 1 2 1 1 2 2 1 1 1 2 ...
 $ parti   : Factor w/ 6 levels "Parti1","Parti2",...: 5 1 3 5 5 4 3 2 5 2 ...
```

Şekil 4 Veri Tiplerinin Kontrolü

Dplyr kütüphanesini kullanarak “mutate_if” komutu ile tüm veri setimizi Factor formatına çevirdik.

2.Bölüm Veri Hazırlama

2.1 Parti İsimlerinin Anonime Çevrilmesi

```
str(data$parti) # PARTİ SUTUNUNUN LEVEL DEĞERİİ KONTROLU
# PARTİ İSİMLERİNİN ANONİME ÇEVİRİLMESİ
levels(data$parti) <- c("Parti1","Parti2","Parti3","Parti4","Parti5","Parti6")
```

Şekil 5 Parti İsimlerinin Anonim Yapılması

Veri setimizin içindeki Parti sütunun değerlerinin isimlerini parti isimlerinin anonim olması bakımından değiştirdik.

2.2 Veri Setinin Özeti

```
summary(data) #VERİLERİN ÖZETİ
```

Şekil 6 Veri Seti Özeti Almak

Cinsiyet	Yaş	Bölge	Eğitim	soru1	soru2	soru3	soru4	soru5
Erkek:807	0-18 :389	Akdeniz :101	İlkokul : 12	Evet :325	Evet :837	Evet :485	Evet :515	Evet : 49
Kadın: 78	18-30:372	Doğu Anadolu: 18	Lisans :209	Hayır:560	Hayır: 48	Hayır:400	Hayır:370	Hayır:836
	30-50: 85	Ege :130	Lisans Üstü: 35					
	50-60: 25	Güneydoğu : 21	Lise :532					
	60+ : 14	İç Anadolu :116	Ortaokul : 25					
		Karadeniz : 62	Ön Lisans : 72					
		Marmara :437						
soru6	soru7	soru8	soru9	soru10	parti			
Evet :252	Evet :746	Evet :285	Evet :617	Evet :687	Parti1:126			
Hayır:633	Hayır:139	Hayır:600	Hayır:268	Hayır:198	Parti2:255			
					Parti3:193			
					Parti4: 20			
					Parti5:264			
					Parti6: 27			

Şekil 7 Veri Seti Özeti

Summary fonksiyonu veri setimizin özetini vermektedir. Veri setimizin tüm değerleri sütunların isimleri ve içerikleri yukarıdaki Şekil 7’de gösterildiği gibidir.

2.3 Partilerin Yüzdelik Olarak Değerleri

```
round(prop.table(table(data$parti)) * 100, digits = 1) #YÜZDELİK OLARAK PARTİ OYLARI
```

Şekil 8 Parti Oylarını Yüzdesel Olarak Almak

Parti1	Parti2	Parti3	Parti4	Parti5	Parti6
14.2	28.8	21.8	2.3	29.8	3.1

Şekil 9 Parti Oylarının Yüzdesel Olarak Sonucu

Round fonksiyonu sayesinde partilerin aldıkları oyları yüzdelik değer olarak gösterilmesini sağladık. Parti oyları yüzdelik değer biçiminde Şekil 9’da gösterilmektedir.

2.4 Veri Seti Eksik Değer Sorgulaması

```
is.na(data) #VERİ EKSİK DEĞER SORGULAMASI
summary(is.na(data))
```

Şekil 10 Eksik Veri Kontrolü

```
Cinsiyet      Yas      Bolge      Egitim      soru1      soru2      soru3
Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical
FALSE:885     FALSE:885     FALSE:885     FALSE:885     FALSE:885     FALSE:885     FALSE:885
soru4        soru5        soru6        soru7        soru8        soru9        soru10
Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical
FALSE:885     FALSE:885     FALSE:885     FALSE:885     FALSE:885     FALSE:885     FALSE:885
parti
Mode :logical
FALSE:885
```

Şekil 11 Eksik Veri Kontrolünün Özeti

İs.na fonksiyonu sayesinde veri setimizin eksik verilerini sorguluyoruz, çıkan sonuçlara bakıldığında zaman veri setinde eksik bir değer olmadığını gördüğümüz için eksik değerlere bir atama yapmıyoruz. Summary fonksiyonu sayesinde aldığımız sonucun özetini çıkartıyoruz. Çıkarılan özet Şekil 11’de gösterilmektedir.

2.5 En Çok Tekrar Eden Veri

Veri setimizde sütunları temel alarak en çok tekrar eden değerleri çıkardık.

2.5.1 En Çok Katılım Sağlayan Bölge

```
encok_bolge <- table(data$Bolge)
encok_deger_bolge <- names(encok_bolge[which.max(encok_bolge)])
summary(data$Bolge)
encok_deger_bolge #HANGI BOLGENIN COGUNLUK OLDUGUNU GOSTERIR
```

Şekil 12 En Çok Katılım Sağlayan Bölge

```
> summary(data$Bolge)
Akdeniz Doğu Anadolu      Ege      Güneydoğu      İç Anadolu      Karadeniz      Marmara
101      18      130      21      116      62      437
> encok_deger_bolge #HANGI BOLGENIN COGUNLUK OLDUGUNU GOSTERIR
[1] "Marmara"
```

Şekil 13 En Çok Katılım Sağlayan Bölge Sonucu

Veri setini incelediğimizde yapılan ankete katılım en fazla Marmara bölgesinden olmuştur. Sonuçlar Şekil 13’de gösterilmektedir.

2.5.2 En Çok Katılım Sağlayan Yaş Aralığı

```
encok_yas <- table(data$Yas)
encok_deger_yas <- names(encok_yas[which.max(encok_yas)])
summary(data$Yas)
encok_deger_yas #HANGI YAS GRUBUNUN COGUNLUK OLDUGUNU GOSTERIR
```

Şekil 14 En Çok Katılım Sağlayan Yaş Aralığı

```
> summary(data$Yas)
0-18 18-30 30-50 50-60 60+
 389   372    85    25    14
> encok_deger_yas #HANGI YAS GRUBUNUN COGUNLUK OLDUGUNU GOSTERIR
[1] "0-18"
```

Şekil 15 En Çok Katılım Sağlayan Yaş Aralığı Sonucu

Veri setini incelediğimizde yapılan ankete katılım en fazla 0-18 yaş grubu aralığındadır. Sonuçlar Şekil 15’de gösterilmektedir.

2.5.3 En Çok Katılım Sağlayan Eğitimin Durumu

```
encok_egitim <- table(data$Egitim)
encok_deger_egitim <- names(encok_egitim[which.max(encok_egitim)])
summary(data$Egitim)
encok_deger_egitim #HANGI EGITIM DURUMUNUN COGUNLUK OLDUGUNU GOSTERIR
```

Şekil 16 En Çok Katılım Sağlayan Eğitimin Durumu

```
> summary(data$Egitim)
ilkokul  Lisans Lisans Üstü  Lise  Ortaokul  Ön Lisans
    12    209     35    532     25     72
> encok_deger_egitim #HANGI EGITIM DURUMUNUN COGUNLUK OLDUGUNU GOSTERIR
[1] "Lise"
```

Şekil 17 En Çok Katılım Sağlayan Eğitimin Durumu Sonucu

Veri setini incelediğimizde yapılan ankete katılım en fazla Lise mezunları tarafından yapılmıştır. Sonuçlar Şekil 17’de gösterilmektedir.

2.5.4 Anket Sonucu En Çok Desteklenen Parti

```
parti <- table(data$parti)
encok_deger_parti <- names(parti[which.max(parti)])
summary(data$parti)
encok_deger_parti #HANGI PARTININ COGUNLUK OLDUGUNU GOSTERIR
```

Şekil 18 Anket Sonucu En Çok Desteklenen Parti

```
> summary(data$parti)
Parti1 Parti2 Parti3 Parti4 Parti5 Parti6
  126    255    193     20    264     27
> encok_deger_parti #HANGI PARTININ COGUNLUK OLDUGUNU GOSTERIR
[1] "Parti5"
```

Şekil 19 Anket Sonucu En Çok Desteklenen Parti Sonucu

Veri setini incelediğimizde yapılan anketin sonucuna göre Parti 5 en çok desteği almıştır. Sonuçlar Şekil 19’da gösterilmektedir.

3. Bölüm Veri Görselleştirme

3.1 Soruların Grafikleri

```
install.packages("ggplot")  
library(ggplot2)
```

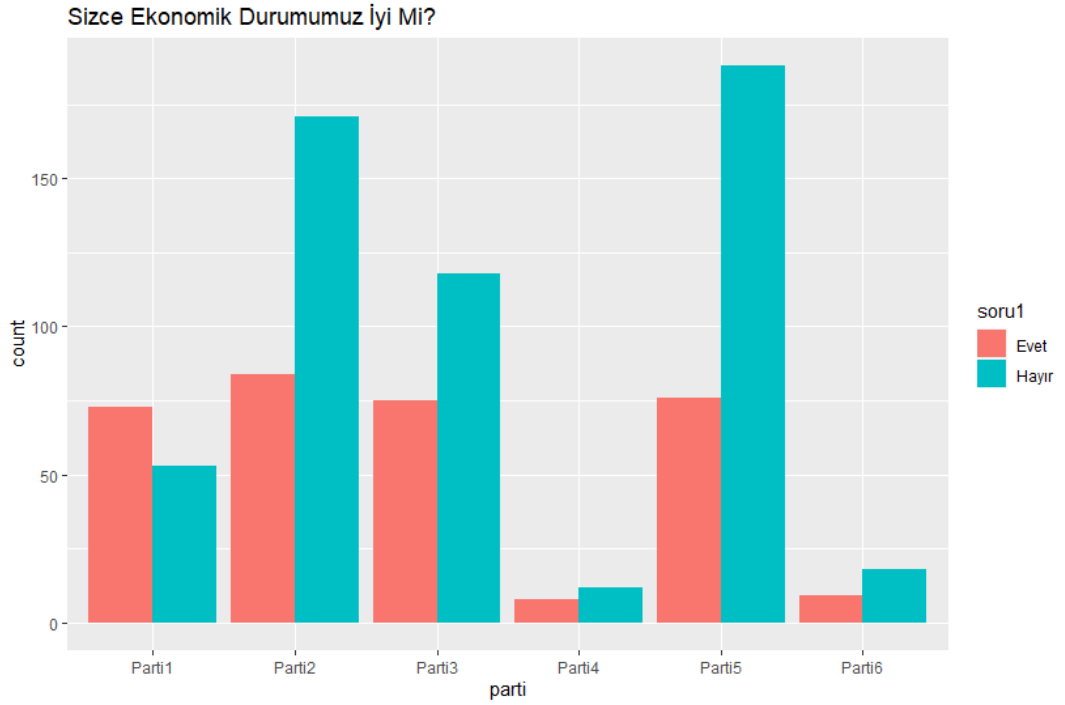
Şekil 20 ggplot kütüphanesi okutmak

Ggplot kütüphanesini kullanarak ankette sorulan soruların cevapları ve verilen cevaplara göre en çok destek alan partilerin grafiklerini çıkardık.

3.1.1 Sizce Ekonomik Durumumuz İyi Mi?

```
ggplot(data) +  
  geom_bar(mapping = aes(x = parti , fill = soru1, ), position = "dodge") +  
  labs(title = "Sizce Ekonomik Durumumuz İyi Mi?")
```

Şekil 21 Sizce Ekonomik Durumumuz İyi Mi?



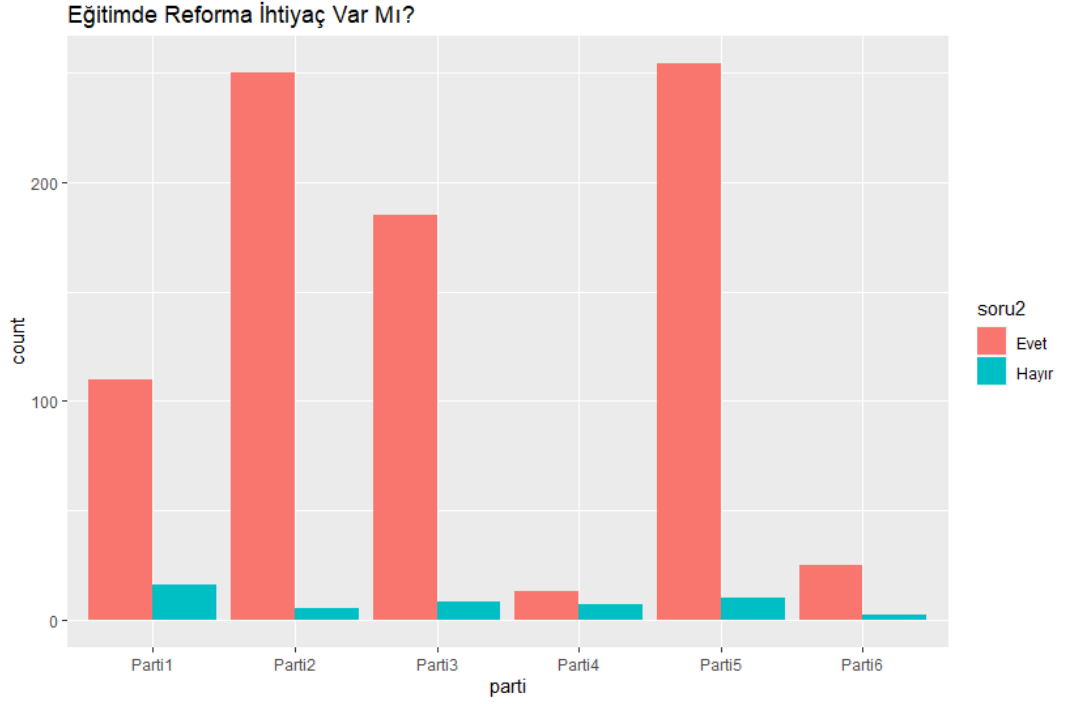
Şekil 22 Sizce Ekonomik Durumumuz İyi Mi? Grafiği

Şekil 22’de görülen sonuçlara göre “Sizce Ekonomik Durumumuz İyi Mi?” sorusuna en çok evet veren katılımcılar Parti 2’yi, hayır cevabını veren katılımcılar ise Parti 5’i desteklemiştir.

3.1.2 Eğitimde Reforma İhtiyaç Var Mı?

```
ggplot(data) +  
  geom_bar(mapping = aes(x =parti , fill = soru2, ), position = "dodge") +  
  labs(title = "Eğitimde Reforma İhtiyaç Var Mı?")
```

Şekil 23 Eğitimde Reforma İhtiyaç Var Mı?



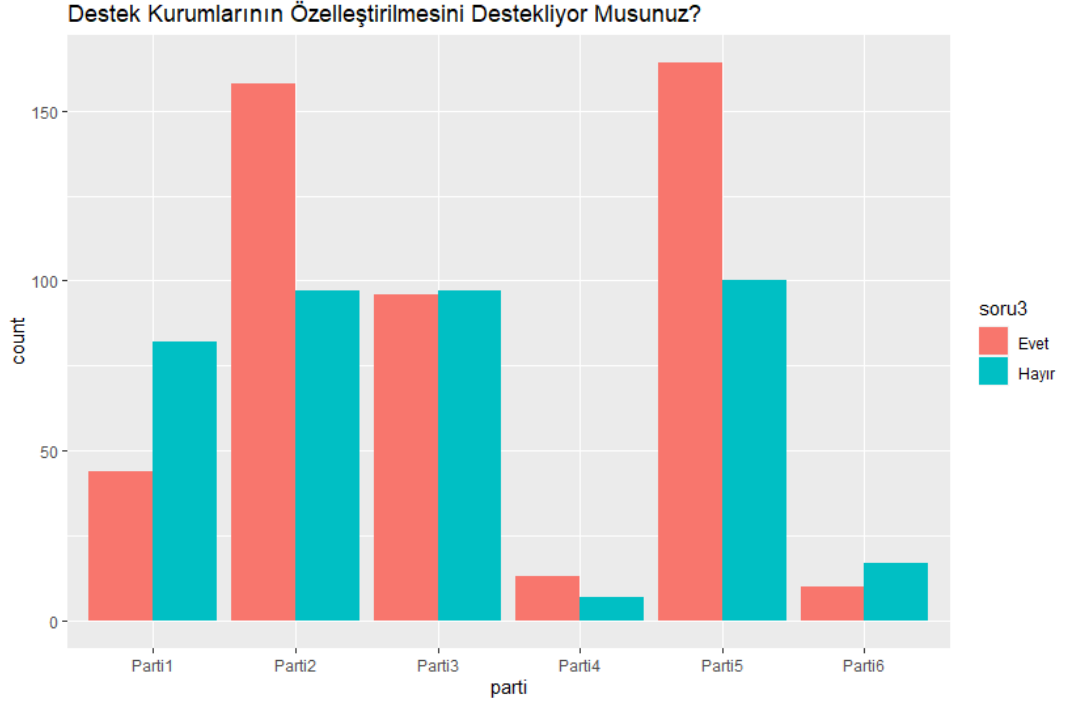
Şekil 24 Eğitimde Reforma İhtiyaç Var Mı? Grafiği

Şekil 24’de görülen sonuçlara göre “Eğitimde Reforma (düzenlemeye) İhtiyaç Var Mı?” sorusuna en çok evet veren katılımcılar Parti 5’i, hayır cevabını veren katılımcılar ise Parti 1’i desteklemiştir.

3.1.3 Destek Kurumlarının Özelleştirilmesini Destekliyor Musunuz?

```
ggplot(data) +  
  geom_bar(mapping = aes(x =parti , fill = soru3, ), position = "dodge") +  
  labs(title = "Destek Kurumlarının Özelleştirilmesini Destekliyor Musunuz?")
```

Şekil 25 Destek Kurumlarının Özelleştirilmesini Destekliyor Musunuz?



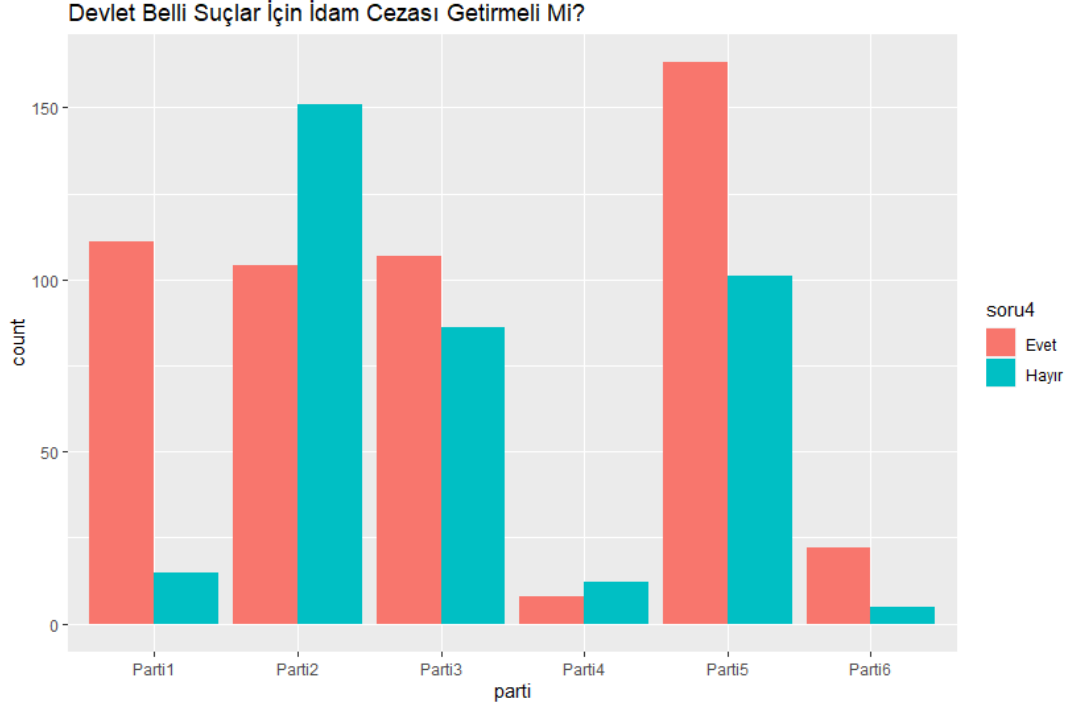
Şekil 26 Destek Kurumlarının Özelleştirilmesini Destekliyor Musunuz? Grafiği

Şekil 26’da görülen sonuçlara göre “Destek Kurumlarının Özelleştirilmesini Destekliyor Musunuz?” sorusuna en çok evet veren katılımcılar Parti 5’i, hayır cevabını veren katılımcılar ise Parti 5’i desteklemiştir.

3.1.4 Devlet Belli Suçlar İçin İdam Cezası Getirmeli Mi?

```
ggplot(data) +  
  geom_bar(mapping = aes(x =parti , fill = soru4, ), position = "dodge") +  
  labs(title = "Devlet Belli Suçlar İçin İdam Cezası Getirmeli Mi?")
```

Şekil 27 Devlet Belli Suçlar İçin İdam Cezası Getirmeli Mi?



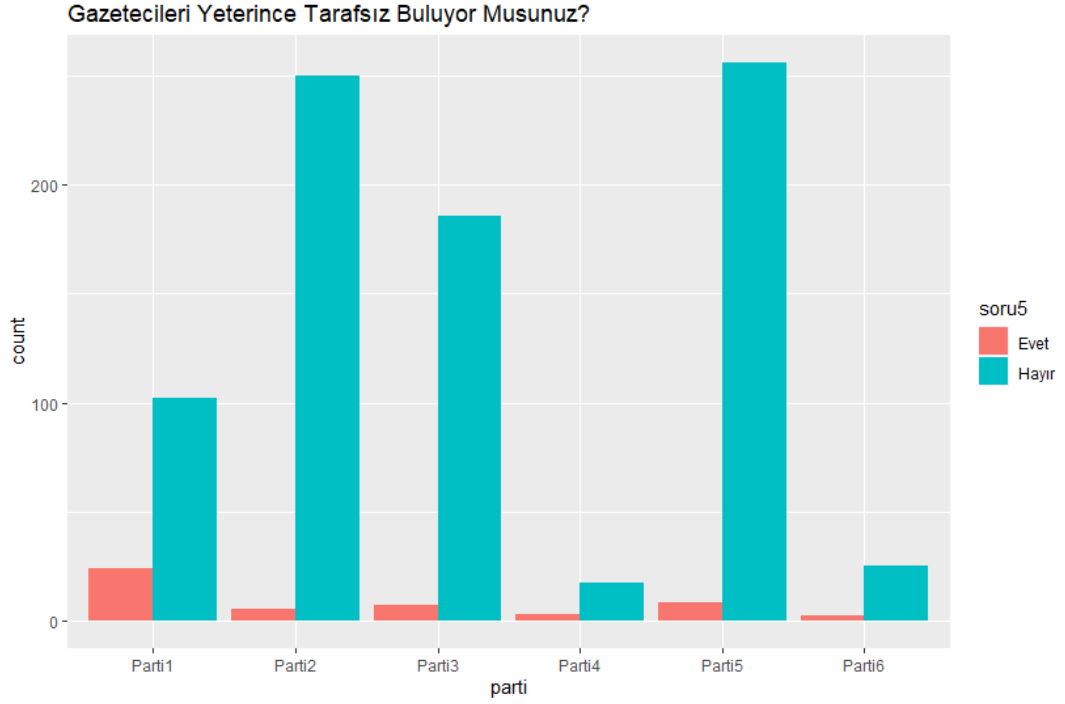
Şekil 28 Devlet Belli Suçlar İçin İdam Cezası Getirmeli Mi? Grafiği

Şekil 28’de görülen sonuçlara göre “Devlet Belli Suçlar İçin İdam Cezası Getirmeli Mi?” sorusuna en çok evet veren katılımcılar Parti 5’i, hayır cevabını veren katılımcılar ise Parti 2’yi desteklemiştir.

3.1.5 Gazetecileri Yeterince Tarafsız Buluyor Musunuz?

```
ggplot(data) +  
  geom_bar(mapping = aes(x =parti , fill = soru5, ), position = "dodge") +  
  labs(title = "Gazetecileri Yeterince Tarafsız Buluyor Musunuz?")
```

Şekil 29 Gazetecileri Yeterince Tarafsız Buluyor Musunuz?



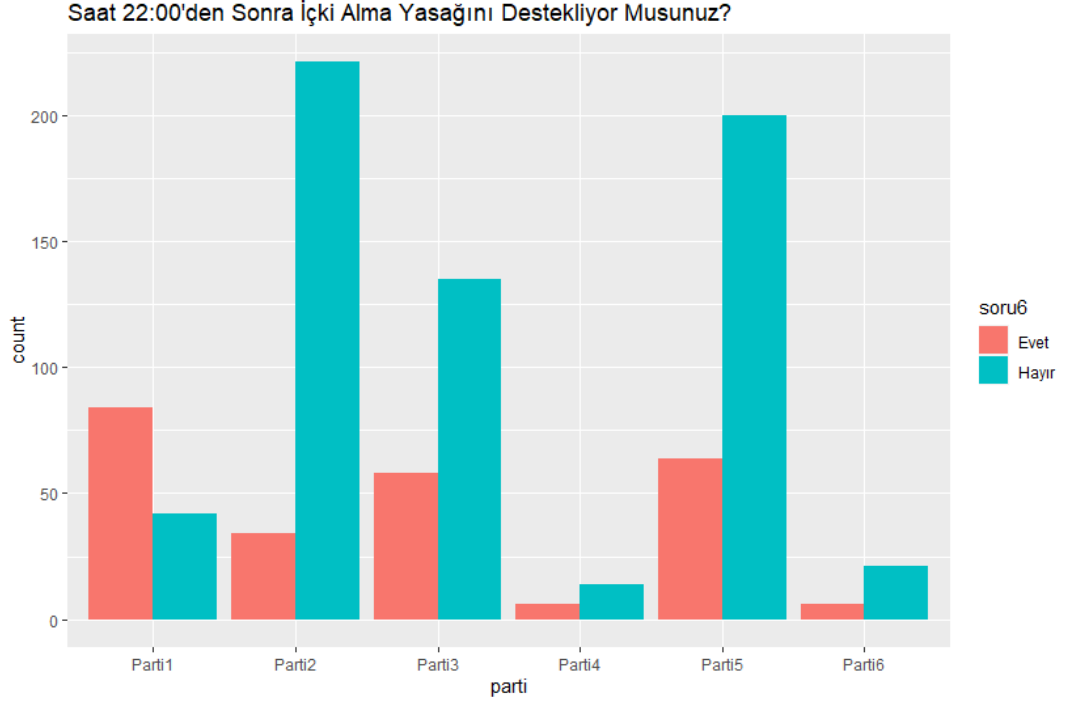
Şekil 30 Gazetecileri Yeterince Tarafsız Buluyor Musunuz? Grafiği

Şekil 30’da görülen sonuçlara göre “Gazetecileri Yeterince Tarafsız Buluyor Musunuz?” sorusuna en çok evet veren katılımcılar Parti 1’i, hayır cevabını veren katılımcılar ise Parti 5’i desteklemiştir.

3.1.6 Saat 22:00'den Sonra İçki Alma Yasağını Destekliyor Musunuz?

```
ggplot(data) +  
  geom_bar(mapping = aes(x =parti , fill = soru6, ), position = "dodge") +  
  labs(title = "Saat 22:00'den Sonra İçki Alma Yasağını Destekliyor Musunuz?")
```

Şekil 31 Saat 22:00'den Sonra İçki Alma Yasağını Destekliyor Musunuz?



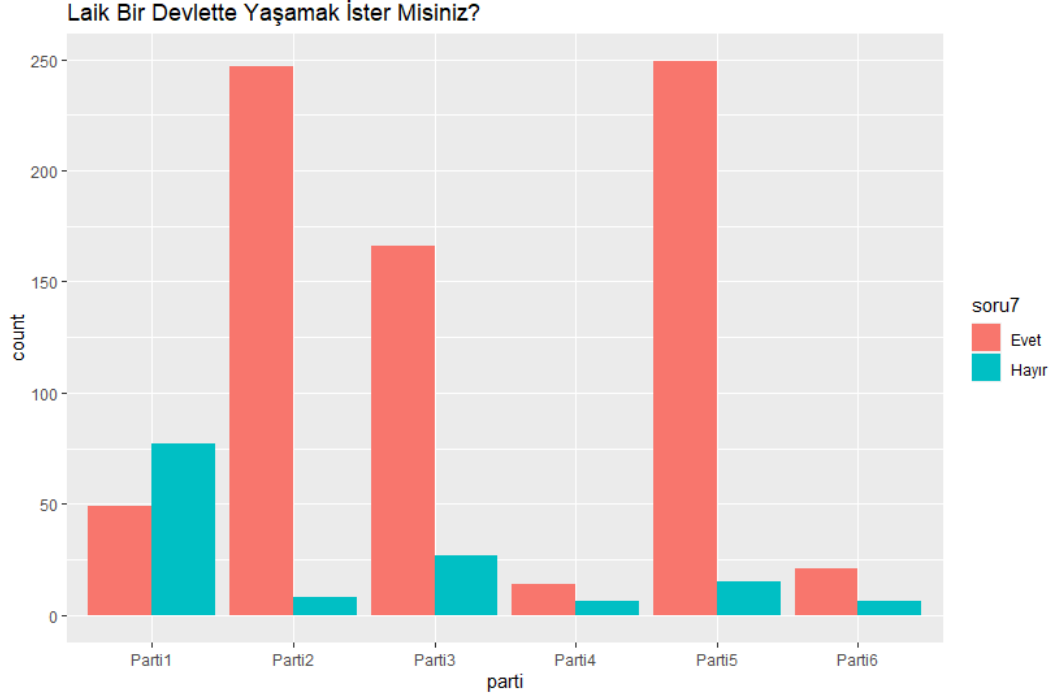
Şekil 32 Saat 22:00'den Sonra İçki Alma Yasağını Destekliyor Musunuz? Grafiği

Şekil 32’de görülen sonuçlara göre “Saat 22:00'den Sonra İçki Alma Yasağını Destekliyor Musunuz?” sorusuna en çok evet veren katılımcılar Parti 1’i, hayır cevabını veren katılımcılar ise Parti 2’yi desteklemiştir.

3.1.7 Laik Bir Devlette Yaşamak İster Misiniz?

```
ggplot(data) +  
  geom_bar(mapping = aes(x =parti , fill = soru7, ), position = "dodge") +  
  labs(title = "Laik Bir Devlette Yaşamak İster Misiniz?")
```

Şekil 33 Laik Bir Devlette Yaşamak İster Misiniz?



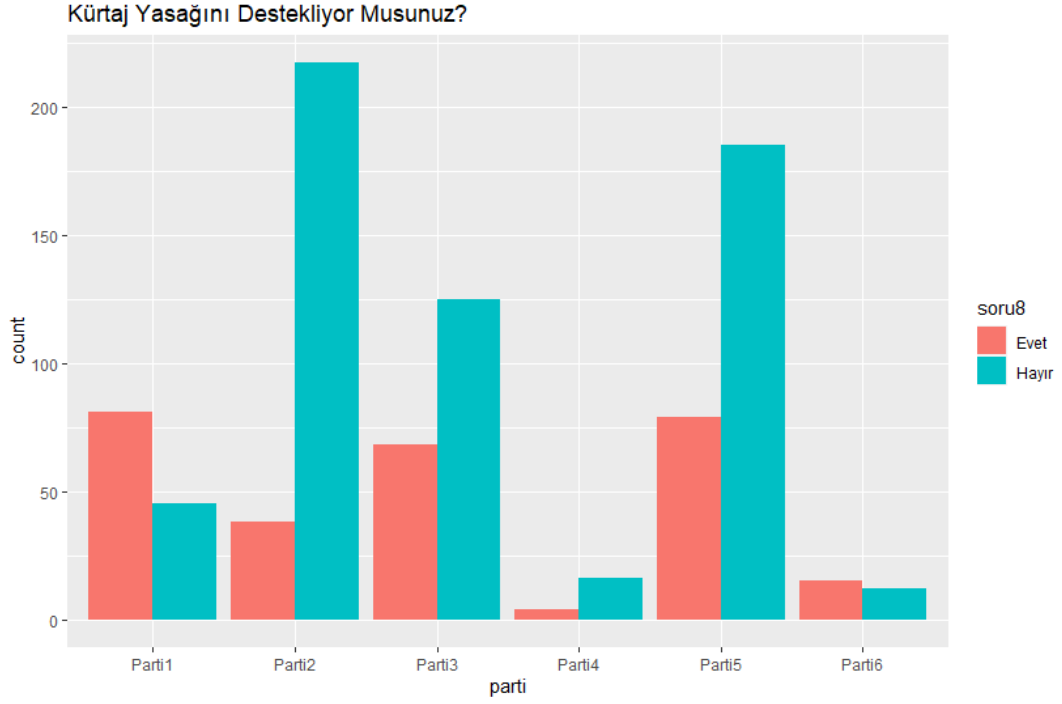
Şekil 34 Laik Bir Devlette Yaşamak İster Misiniz? Grafiği

Şekil 34’de görülen sonuçlara göre “Laik Bir Devlette Yaşamak İster Misiniz?” sorusuna en çok evet veren katılımcılar Parti 5’i, hayır cevabını veren katılımcılar ise Parti 1’i desteklemiştir.

3.1.8 Kürtaj Yasağını Destekliyor Musunuz?

```
ggplot(data) +  
  geom_bar(mapping = aes(x =parti , fill = soru8, ), position = "dodge") +  
  labs(title = "Kürtaj Yasağını Destekliyor Musunuz?")
```

Şekil 35 Kürtaj Yasağını Destekliyor Musunuz?



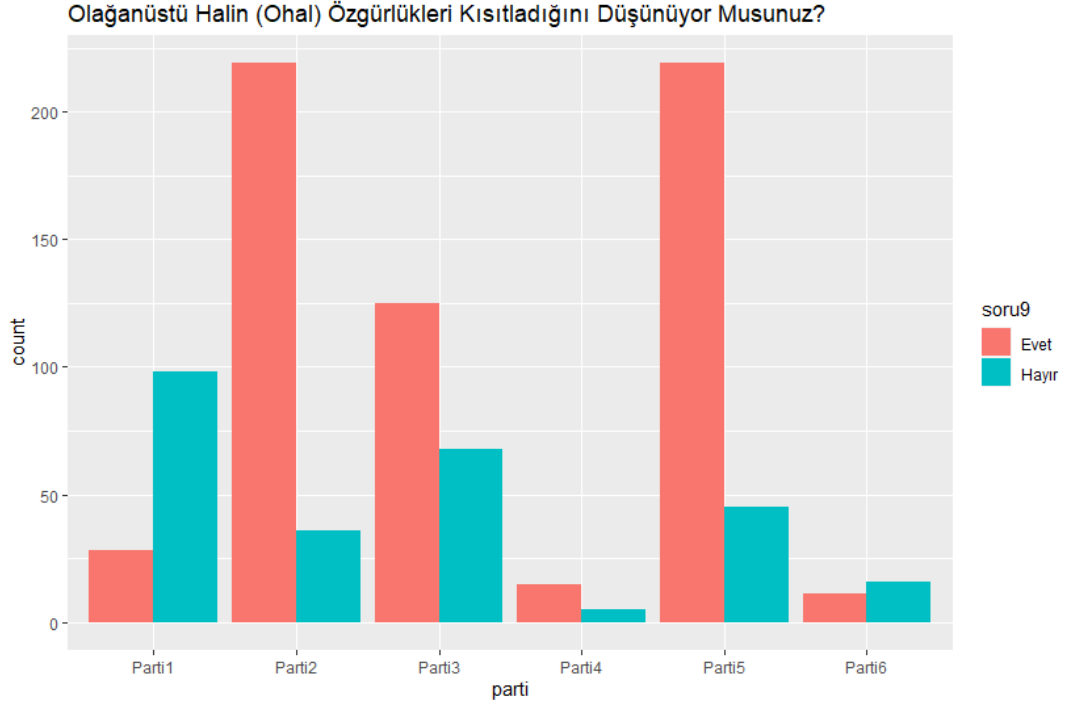
Şekil 36 Kürtaj Yasağını Destekliyor Musunuz? Grafiği

Şekil 36’da görülen sonuçlara göre “Kürtaj Yasağını Destekliyor Musunuz?” sorusuna en çok evet veren katılımcılar Parti 1’i, hayır cevabını veren katılımcılar ise Parti 2’yi desteklemiştir.

3.1.9 Olağanüstü Halin (Ohal) Özgürlükleri Kısıtladığını Düşünüyor Musunuz?

```
ggplot(data) +  
  geom_bar(mapping = aes(x = parti , fill = soru9, ), position = "dodge") +  
  labs(title = "Olağanüstü Halin (Ohal) Özgürlükleri Kısıtladığını Düşünüyor Musunuz?")
```

Şekil 37 Olağanüstü Halin (Ohal) Özgürlükleri Kısıtladığını Düşünüyor Musunuz?



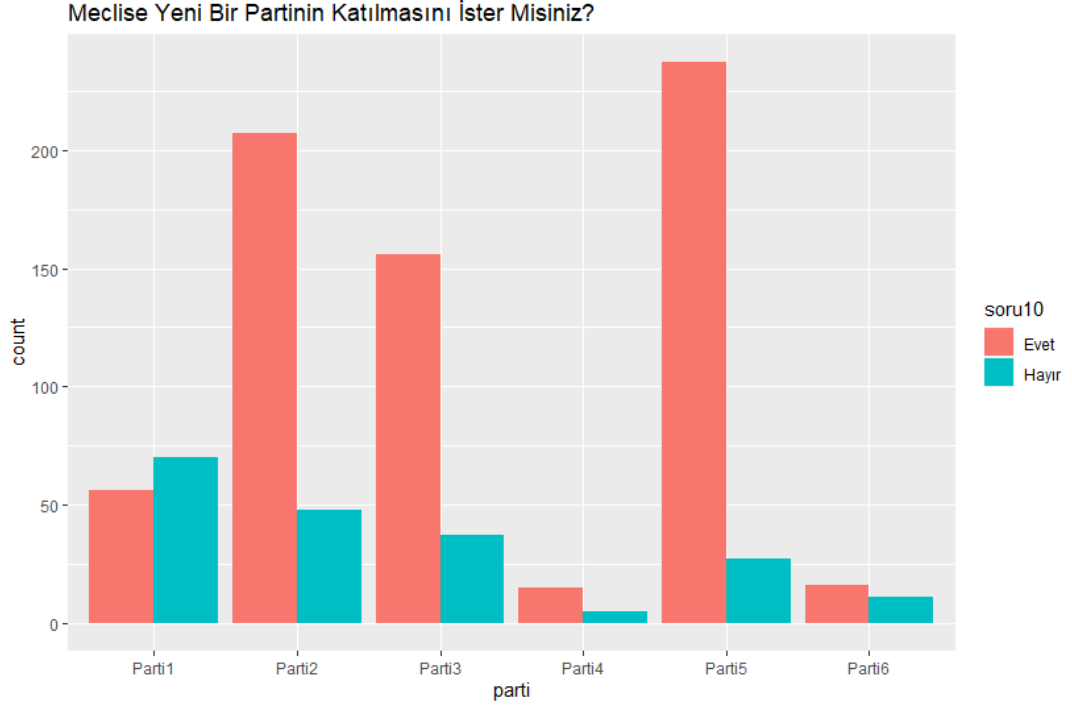
Şekil 38 Olağanüstü Halin (Ohal) Özgürlükleri Kısıtladığını Düşünüyor Musunuz? Grafiği

Şekil 38’de görülen sonuçlara göre “Olağanüstü Halin (Ohal) Özgürlükleri Kısıtladığını Düşünüyor Musunuz?” sorusuna en çok evet veren katılımcılar Parti 2’yi ve Parti5’i, hayır cevabını veren katılımcılar ise Parti 1’i desteklemiştir.

3.1.10 Meclise Yeni Bir Partinin Katılmasını İster Misiniz?

```
ggplot(data) +  
  geom_bar(mapping = aes(x =parti , fill = soru10, ), position = "dodge") +  
  labs(title = "Meclise Yeni Bir Partinin Katılmasını İster Misiniz?")
```

Şekil 39 Meclise Yeni Bir Partinin Katılmasını İster Misiniz?



Şekil 40 Meclise Yeni Bir Partinin Katılmasını İster Misiniz? Grafiği

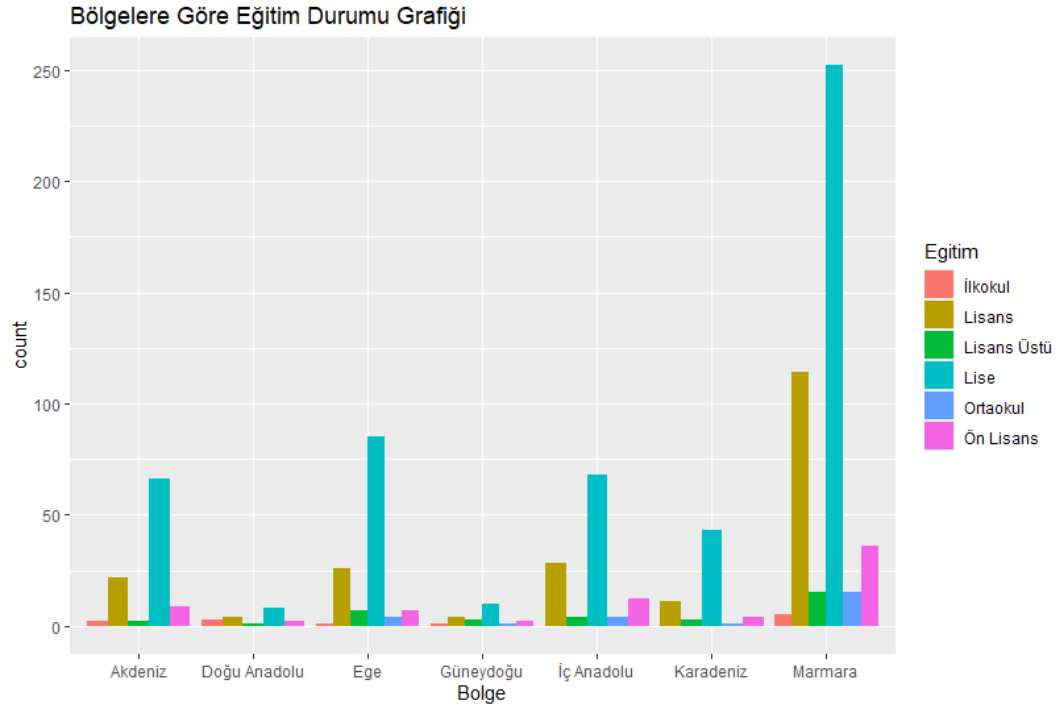
Şekil 40’da görülen sonuçlara göre “Meclise Yeni Bir Partinin Katılmasını İster Misiniz?” sorusuna en çok evet veren katılımcılar Parti 5’i, hayır cevabını veren katılımcılar ise Parti 1’i desteklemiştir.

3.2 Karşılaştırma Grafikleri

3.2.1 Bölgelere Göre Eğitim Durumu Grafiği

```
egitim_bolge <- ggplot(data, aes(Bolge, fill = Egitim))  
egitim_bolge + geom_bar(position = "dodge") +  
  labs(title = "Bölgelere Göre Eğitim Durumu Grafiği")
```

Şekil 41 Bölgelere Göre Eğitim Durumu



Şekil 42 Bölgelere Göre Eğitim Durumu Grafiği

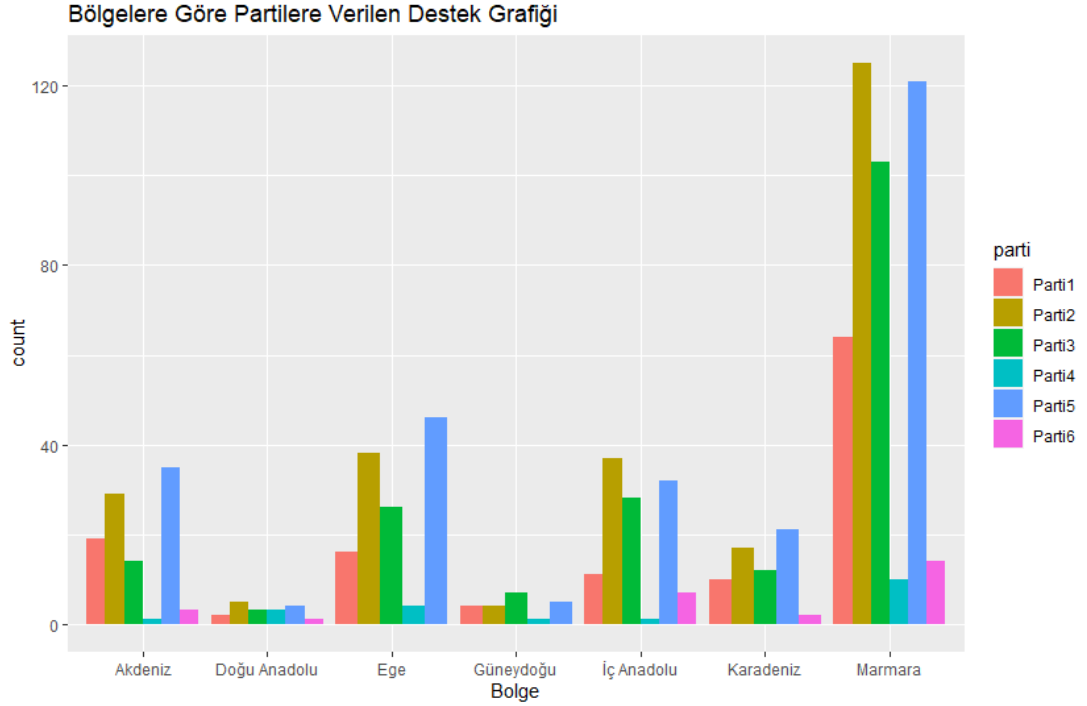
Şekil 42’de gördüğümüz grafikte bölgelere göre eğitim durumu incelenmiştir. Grafiğe bağlı olarak eğitim durumları Marmara bölgesinde en fazla, en azı ise Doğu Anadolu’da gözükmemektedir.

Genel grafiğe baktığımız zaman eğitim seviyesi en fazla Lise Mezunları olduğunu görebiliyoruz.

3.2.2 Bölgelere Göre Partilere Verilen Destek Grafiği

```
bolge <- ggplot(data, aes(Bolge, fill = parti))  
bolge + geom_bar(position = "dodge") +  
  labs(title = "Bölgelere Göre Partilere Verilen Destek Grafiği")
```

Şekil 43 Bölgelere Göre Partilere Verilen Destek



Şekil 44 Bölgelere Göre Partilere Verilen Destek Grafiği

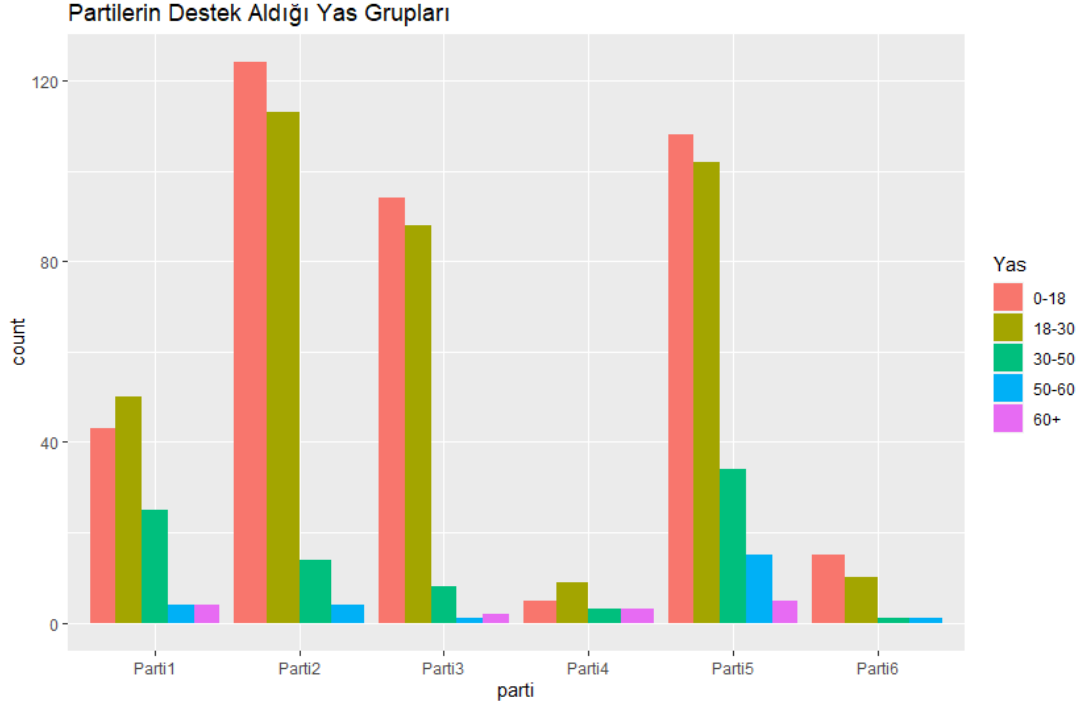
Şekil 44’de gördüğümüz grafikte bölgelere göre partilere verilen destek grafiği verilmiştir. Grafiğe bağlı olarak partilere en çok destek verilen bölge Marmara’dır.

Genel grafiğe baktığımız zaman bölgelere göre verilen en çok destek Parti 2 ve Parti 5 arasındadır.

3.2.3 Partilerin Destek Aldığı Yas Grupları

```
yas <- ggplot(data, aes(parti, fill = yas))  
yas + geom_bar(position = "dodge") +  
  labs(title = "Partilerin Destek Aldığı Yas Grupları")
```

Şekil 45 Partilerin Destek Aldığı Yas Grupları



Şekil 46 Partilerin Destek Aldığı Yas Grupları Grafiği

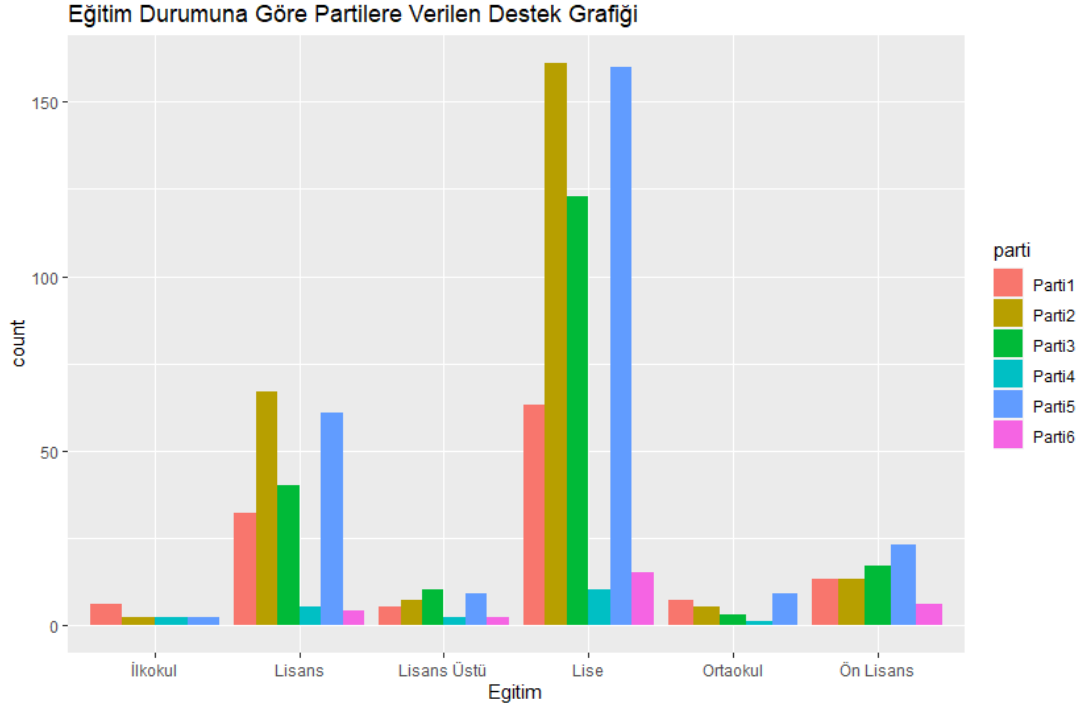
Şekil 46’da gördüğümüz grafikte partilerin destek aldığı yaş grupları verilmiştir. Grafiğe baktığımızda katılımcıların çoğunun 0-18 yaş grubu arasında olduğunu görebilmekteyiz. Grafiğe bağlı olarak 0-18 yaş grubu arası en çok parti 2’yi, 18-30 yaş grubu arası en çok parti 2’yi, 30-50 yaş grubu arası en çok parti 5’i 50-60 yaş grubu arası en çok parti 5’i, 60+ yaş grubu arası ise parti 5’i desteklemiştir.

Sonuç olarak çoğunlukla parti 5’in desteklendiği görülmüştür.

3.2.4 Eğitim Durumuna Göre Partilere Verilen Destek Grafiği

```
egitim <- ggplot(data, aes(Egitim, fill = parti))  
egitim + geom_bar(position = "dodge") +  
  labs(title = "Eğitim Durumuna Göre Partilere Verilen Destek Grafiği")
```

Şekil 47 Eğitim Durumuna Göre Partilere Verilen Destek



Şekil 48 Eğitim Durumuna Göre Partilere Verilen Destek Grafiği

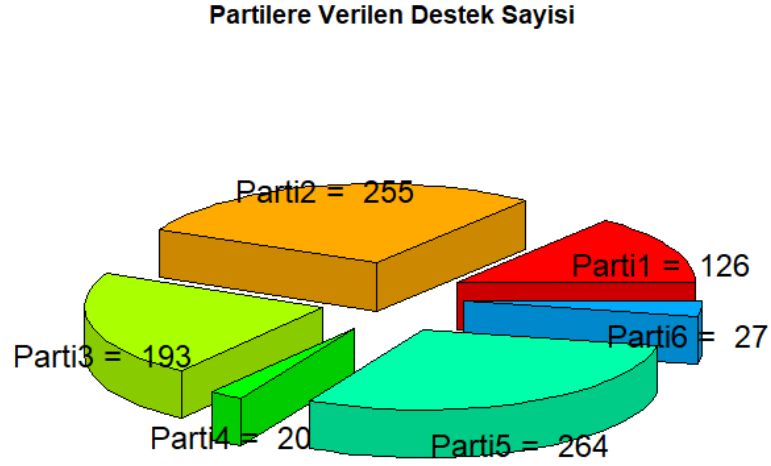
Şekil 48’de gördüğümüz grafikte eğitim durumuna göre partilere verilen destek grafiği verilmiştir. Grafiğe bağlı olarak ankete katılım sağlayan kişilerin çoğunun eğitim durumu Lise olarak gözükmektedir.

Sonuç olarak Parti 2 ve Parti 5’in destekleyenlerin çoğunluk olduğunu görmekteyiz.

3.3 Partilere Verilen Destek Sayısı Grafiđi

```
#install.packages("plotrix")  
library(plotrix)  
frekanslar2 <- table(data$parti)  
etiketler <- paste(names(frekanslar2), frekanslar2, sep=" = ")  
pie3D(frekanslar2, labels = etiketler, main="Partilere Verilen Destek Sayısı",  
      col = rainbow(9), explode = 0.3)
```

Şekil 49 Partilere Verilen Destek Sayısı



Şekil 50 Partilere Verilen Destek Sayısı Grafiđi

Şekil 50’de gördüğümüz pasta grafiđinde partilere verilen destek sayıları verilmiştir. Grafiđi incelediğimizde partiler arasında en çok desteklenen Parti 5 olduđu görölmektedir.

Grafiđe göre en çoktan en aza verilen destek sayısı şu şekildedir;

Parti 5: 264

Parti 2: 255

Parti 3: 193

Parti 1: 126

Parti 6: 27

Parti 4: 20

4.Bölüm Veri Modelleme

4.1 Eğitim ve Test Seti Oluşturma

```
library(caret)
set.seed(27)
egitimIndisleri <- createDataPartition(y = data$parti, p = .75, list = FALSE)
egitim <- data[egitimIndisleri,]
test <- data[-egitimIndisleri,]
```

Şekil 51 Eğitim ve Test Seti Oluşturma

Bu aşamada, veri seti iki bölüme ayrılır:

Eğitim Seti: Veri setinin bu kısmı, Makine Öğrenimi modelini oluşturmak ve eğitmek için kullanılır.

Test Seti: Veri setinin bu kısmı modelin verimliliğini değerlendirmek için kullanılır.

Eğitim ve Test setini oluştururken elle sayı ataması yapmamak için “Caret” kütüphanesinden kullanıldı. Caret sayesinde veri setini %75 eğitim, %25 test seti olarak ayarlandı.

Değerler her okunduğunda farklı setler vermemesi için set seed değeri 27 olarak kullanıldı.

```
testNitelikleri <- test[, -15]
testHedefNitelik <- test[[15]]

egitimNitelikleri <- egitim[, -15]
egitimHedefNitelik <- egitim[[15]]
```

Şekil 52 Nitelik Ve Hedef Seti Oluşturma

Eğitim ve test setlerinde tahminde kullanılacak nitelikler ve hedef niteliğin ayrı nesnelere atanması yapıldı.

4.2 Basit (Naive) Bayes Algoritması

İsmi matematikçi Thomas Bayes'den alan bir sınıflandırma algoritmasıdır. Bayes sınıflandırması olasılık ilkelerine göre tanımlanmış bir dizi hesaplama ile sisteme sunulan verilerin sınıfını yani kategorisini tespit etmeyi amaçlamaktadır. Bayes sınıflandırmasında sisteme belirli bir oranda öğretilmiş veri sunulmaktadır. Bu veriler kategorik veriler olup, öğretilmiş veriler üzerinden yeni bir veri setinin hangi kategoride olduğu tespit edilmeye çalışılmaktadır. Bu tespit başarısı, öğretilmiş verilerin fazlalığından beslenmektedir. (EROĞLU,2009,s.44)

4.2.1 Naive Bayes Modeli Oluşturma

```
#install.packages("e1071")
library(e1071)
naiveBayes_modeli <- naiveBayes(egitimNitelikleri, egitimHedefNitelik)
naiveBayes_modeli
```

Şekil 53 Naive Bayes Modeli

e1071 kütüphanesi sayesinde Naive Bayes modelini oluşturuyoruz, modeli oluştururken “egitimNitelikleri” ve “egitimHedefNitelikleri” adındaki setlerimizi kullanıyoruz.

```
(tahminisinfilar <-
predict(naiveBayes_modeli,testNitelikleri))
```

Şekil 54 Tahmin Sınıfı

Naive Bayes modeli ile test niteliklerinden tahmini sınıfı oluşturduk.

```
> summary(tahminisinfilar)
Parti1 Parti2 Parti3 Parti4 Parti5 Parti6
   36    103    25     1    50     4
```

Şekil 55 Tahmin Sınıfı Özeti

tahminSinfilar setinin özeti Şekil 55’de gösterildiği gibidir.

4.2.2 Performans Değerlendirmesi

```
# Performans degerlendirmesi
confusionMatrix(data = tahminisiniflar, reference = testHedefnitelik)
```

Şekil 56 Performans Değerlendirmesi

Oluşturduğum tahmin sınıfını “confusionMatrix” komutu ile performans değerlerini alıyoruz.

Confusion Matrix and Statistics						
Reference						
Prediction	Parti1	Parti2	Parti3	Parti4	Parti5	Parti6
Parti1	23	0	5	2	4	2
Parti2	2	43	22	3	32	1
Parti3	1	5	11	0	7	1
Parti4	0	0	1	0	0	0
Parti5	3	15	9	0	21	2
Parti6	2	0	0	0	2	0
Overall Statistics						
Accuracy : 0.4475						
95% CI : (0.3805, 0.516)						
No Information Rate : 0.3014						
P-Value [Acc > NIR] : 3.597e-06						
Kappa : 0.2604						
McNemar's Test P-Value : NA						
Statistics by Class:						
	Class: Parti1	Class: Parti2	Class: Parti3	Class: Parti4	Class: Parti5	Class: Parti6
Sensitivity	0.7419	0.6825	0.22917	0.000000	0.31818	0.00000
Specificity	0.9309	0.6154	0.91813	0.995327	0.81046	0.98122
Pos Pred Value	0.6389	0.4175	0.44000	0.000000	0.42000	0.00000
Neg Pred Value	0.9563	0.8276	0.80928	0.977064	0.73373	0.97209
Prevalence	0.1416	0.2877	0.21918	0.022831	0.30137	0.02740
Detection Rate	0.1050	0.1963	0.05023	0.000000	0.09589	0.00000
Detection Prevalence	0.1644	0.4703	0.11416	0.004566	0.22831	0.01826
Balanced Accuracy	0.8364	0.6490	0.57365	0.497664	0.56432	0.49061

Şekil 57 Naive Bayes Modelinin Performans Değeri Sonuçları

Performans değerlendirme sonuçlarına göre doğruluk oranı 0.45 olarak görülmektedir.

4.3 Karar Ağacı Algoritması

Karar ağacı, çok sayıda kayıt içeren bir veri kümesini, bazı kuralları uygulayarak daha küçük kümelerle bölmek için kullanılan bir yapıdır. Karar ağacı kökten yapraklara doğru yinelemeli olarak veriyi bölerek kazanma yöntemine göre inşa edilir. Başlangıçta bütün veriler ağacın kökünde toplanır. Değişkenlerin seçimi bilgi kazanımı değerine göre belirlenir. Yinelemeli olan algoritmanın döngüden çıkması için o düğümdeki tüm öğelerin aynı sınıfa dahil olması şartı vardır. Eğer kalan değerler sadece bir sınıfa aitse veya sınıflandırılacak değer kalmadıysa döngüsel algoritma sonlanır ve karar ağacı oluşturulmuş olur. (SEZER,2008,s26)

4.3.1 Karar Ağacı Modeli Oluşturma

```
library(rpart)
fit <- rpart(parti ~., data = training, method = 'class')
```

Şekil 58 Karar Ağacı Modeli Oluşturma

Rpart kütüphanesi sayesinde Karar Ağacı modelini oluşturuyoruz.

```
(tahmin<-
predict(fit,test_,type = "class"))
```

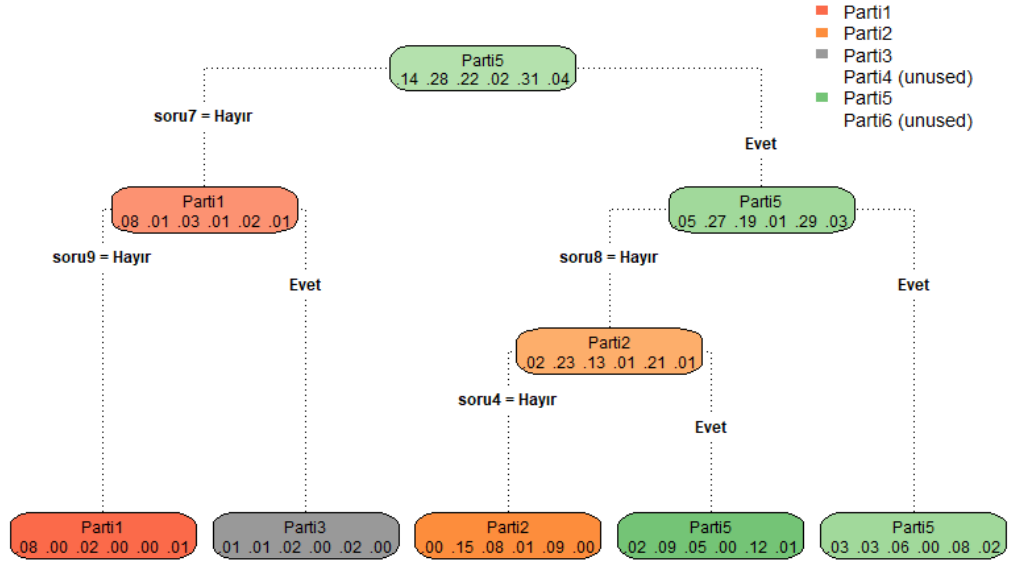
Şekil 59 Tahmin Sınıfı Oluşturma

Karar Ağacı modeli ile test setinden tahmin veri seti oluşturduk.

```
> summary(tahmin)
Parti1 Parti2 Parti3 Parti4 Parti5 Parti6
28      64      4      0     125      0
```

Şekil 60 Tahmin Sınıfı Özeti

4.3.2 Karar Ağacı Grafiği



Şekil 61 Karar Ağacı Grafiği

Veri setimizde Parti5 en çok destek alan parti olduğundan dolayı Karar Ağacı modeli Parti5 ile başlıyor. Parti5'i destekleyen bir kişi soru 7'ye Hayır cevabı verir ise model Parti1'de olmasını gösteriyor buna göre Şekil 61'de ki Kara Ağacı yorumlanabilir.

4.3.3 Performans Değerlendirmesi

```
# Performans degerlendirmesi
library(caret)
confusionMatrix(data = tahmin, reference = test_$parti)
```

Şekil 62 Performans Değerlendirmesi

Confusion Matrix and Statistics						
Reference						
Prediction	Parti1	Parti2	Parti3	Parti4	Parti5	Parti6
Parti1	21	1	4	1	1	0
Parti2	1	32	13	1	17	0
Parti3	0	0	2	1	1	0
Parti4	0	0	0	0	0	0
Parti5	14	37	26	4	41	3
Parti6	0	0	0	0	0	0
Overall Statistics						
Accuracy : 0.4344						
95% CI : (0.3681, 0.5025)						
No Information Rate : 0.3167						
P-Value [Acc > NIR] : 0.0001596						
Kappa : 0.2256						
McNemar's Test P-Value : NA						
Statistics by Class:						
	Class: Parti1	Class: Parti2	Class: Parti3	Class: Parti4	Class: Parti5	Class: Parti6
Sensitivity	0.58333	0.4571	0.04444	0.00000	0.6833	0.00000
Specificity	0.96216	0.7881	0.98864	1.00000	0.4783	1.00000
Pos Pred Value	0.75000	0.5000	0.50000	NaN	0.3280	NaN
Neg Pred Value	0.92228	0.7580	0.80184	0.96833	0.8021	0.98643
Prevalence	0.16290	0.3167	0.20362	0.03167	0.2715	0.01357
Detection Rate	0.09502	0.1448	0.00905	0.00000	0.1855	0.00000
Detection Prevalence	0.12670	0.2896	0.01810	0.00000	0.5656	0.00000
Balanced Accuracy	0.77275	0.6226	0.51654	0.50000	0.5808	0.50000

Şekil 63 Karar Ağacı Modelinin Performans Değeri Sonucu

Performans değerlendirme sonuçlarına göre doğruluk oranı 0.43 olarak görülmektedir.

4.4 K-Means Algoritması

Kümeleme (Clustering) veri setleri içindeki gizli grup veya ilişki şemalarını belirlemeyi amaçlayan bir veri analizi tekniğidir. İki temel kümeleme tekniği söz konusudur: 1- Partisyon Yönetimi ve 2- Hiyerarşik Kümeleme. Algoritmik zekanın gelişimi ile çok çeşitli kümeleme tekniği geliştirilmiş ve geliştirilmeye devam edilmektedir.

Bu çalışma ile en popüler kümeleme teknikleri olan K-Means'i kullanacağız.

4.4.1 K-Means İçin Veri Hazırlama

```
data_k <- as.data.frame(data)
data_k <- data.matrix(data_k)
head(data_k)
```

Şekil 64 Veri Seti Hazırlama Kodları

Veri seti içindeki kategorik değerleri K-Means algoritmasında kullanmak için data.matrix komutu ile numeric değerle çevirdik.

	Cinsiyet	Yas	Bölge	Eğitim	soru1	soru2	soru3	soru4	soru5	soru6	soru7	soru8	soru9	soru10	parti
[1,]	1	2	7	2	2	1	1	2	2	2	1	2	1	1	5
[2,]	1	2	7	2	1	1	2	1	2	1	2	1	2	2	1
[3,]	2	2	7	2	2	1	2	2	2	2	1	2	1	1	3
[4,]	1	1	7	4	1	1	2	1	2	1	1	1	1	1	5
[5,]	1	2	7	6	2	1	1	1	2	2	1	1	2	2	5
[6,]	1	5	7	3	2	2	2	2	2	2	2	2	2	2	4

Şekil 65 Veri Seti İlk 5 Değerleri

Numeric değerler için sonuç bu şekildedir;

Cinsiyet; 1=Erkek, 2=Kadın

Yas; 1=0-18, 2=18-30, 3=30-50, 4=50-60, 5=60+

Eğitim; 1=İlk Okul, 2=Lisans, 3=Lisans Üstü, 4=Lise, 5=Orta Okul, 6=Ön Lisans

Bölge; 1=Akdeniz, 2=Doğu Anadolu, 3=Ege, 4=Güneydoğu, 5=İç Anadolu, 6=Karadeniz, 7=Marmara

Sorular; 1=Evet, 2=Hayır

Partiler; 1=Parti1 2=Parti2 3=Parti3 4=Parti4 5=Parti5

4.4.2 K-Means İçin Veriyi Anlama

```
desc_stats<-data.frame(  
  Min=apply(data_k, 2, min), # minimum  
  Med=apply(data_k, 2, median), # median  
  Mean=apply(data_k, 2, mean), # mean  
  SD=apply(data_k, 2, sd), # Standard deviation  
  Max=apply(data_k, 2, max)# Maximum  
)  
desc_stats<-round(desc_stats, 1)# Yuvarlama  
desc_stats
```

Şekil 66 Veri Anlama Kodları

	Min	Med	Mean	SD	Max
Cinsiyet	1	1	1.1	0.3	2
Yas	1	2	1.8	0.9	5
Bolge	1	6	5.2	2.1	7
Egitim	1	4	3.6	1.2	6
soru1	1	2	1.6	0.5	2
soru2	1	1	1.1	0.2	2
soru3	1	1	1.5	0.5	2
soru4	1	1	1.4	0.5	2
soru5	1	2	1.9	0.2	2
soru6	1	2	1.7	0.5	2
soru7	1	1	1.2	0.4	2
soru8	1	2	1.7	0.5	2
soru9	1	1	1.3	0.5	2
soru10	1	1	1.2	0.4	2
parti	1	3	3.1	1.5	6

Şekil 67 Min - Max Değerlerinin Çıktısı

K-Means için hazırladığımız veri setinin Minimum, Medyan(Ortanca Değer), Mean(Ortalama Değer), Standard Deviation(Standart Sapma),Maximum değerlerini Şekil 67’de göstermektedir.

```
data_k_scale<-scale(data_k)  
head(data_k_scale)
```

Şekil 68 Anlamli Aralik Kodları

```
> head(data_k_scale)  
  Cinsiyet    Yas    Bolge    Egitim    soru1    soru2    soru3    soru4    soru5    soru6    soru7  
[1,] -0.310717 0.2795273 0.8268137 -1.4203421 0.7613812 -0.2393384 -0.907640 1.1791179 0.2419632 0.6305985 -0.4314123  
[2,] -0.310717 0.2795273 0.8268137 -1.4203421 -1.3119184 -0.2393384 1.100513 -0.8471333 0.2419632 -1.5840033 2.3153493  
[3,] 3.214726 0.2795273 0.8268137 -1.4203421 0.7613812 -0.2393384 1.100513 1.1791179 0.2419632 0.6305985 -0.4314123  
[4,] -0.310717 -0.8873672 0.8268137 0.3134548 -1.3119184 -0.2393384 1.100513 -0.8471333 0.2419632 -1.5840033 -0.4314123  
[5,] -0.310717 0.2795273 0.8268137 2.0472517 0.7613812 -0.2393384 -0.907640 -0.8471333 0.2419632 0.6305985 -0.4314123  
[6,] -0.310717 3.7802107 0.8268137 -0.5534436 0.7613812 4.1734634 1.100513 1.1791179 0.2419632 0.6305985 2.3153493  
  soru8    soru9    soru10    parti  
[1,] 0.6888129 -0.6586873 -0.5365482 1.21950029  
[2,] -1.4501325 1.5164555 1.8616596 -1.40005737  
[3,] 0.6888129 -0.6586873 -0.5365482 -0.09027854  
[4,] -1.4501325 -0.6586873 -0.5365482 1.21950029  
[5,] -1.4501325 1.5164555 1.8616596 1.21950029  
[6,] 0.6888129 1.5164555 1.8616596 0.56461087
```

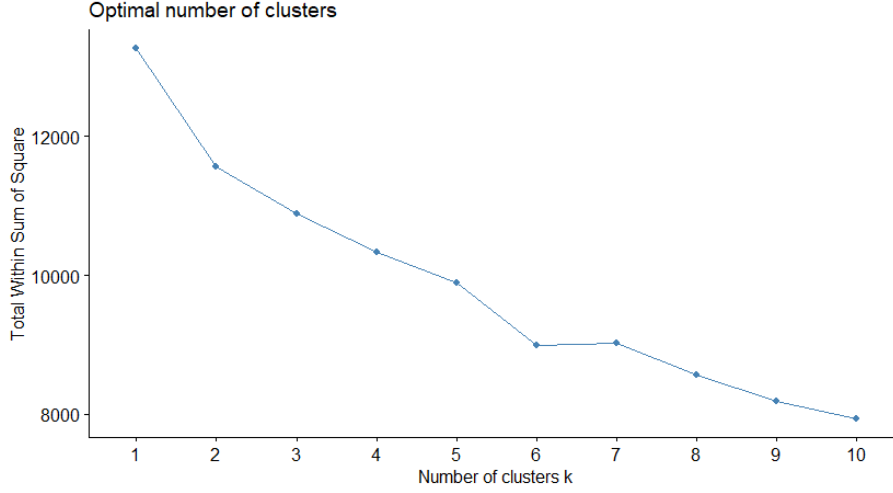
Şekil 69 Anlamli Aralik Özeti

Scale() komutu sayesinde Şekil 69’da değişkenlerin anlamlı aralık farklarını görebiliyoruz.

4.4.3 K-Means İçin Optimal Küme Sayısı Belirlenmesi

```
library(factoextra)
fviz_nbclust(data_k_scale, kmeans, method="wss")
```

Şekil 70 Optimum Küme Sayısı Bulma Kodları

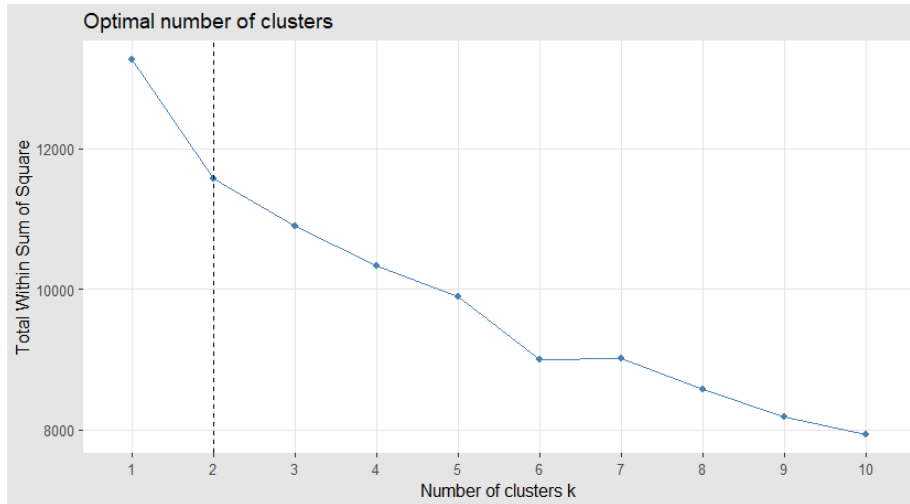


Şekil 71 Optimum Küme Sayısı Bulma Grafiği

Şekil 71’de veri setimiz için dirsek değeri inceliyoruz. Oluşturduğumuz grafik sayesinde optimal değerin 2. değer olduğunu görebiliyoruz.

```
library(ggthemes)
tma<-theme_igray()
fviz_nbclust(data_k_scale, kmeans, method="wss")+
  geom_vline(xintercept=2, linetype=2)+tma
```

Şekil 72 Optimum Küme Sayısı İşaretleme Kodları



Şekil 73 Optimum Küme Sayısı İşaretleme Çıktısı

Şekil 73’de bulduğumuz optimal değeri belirtiyoruz.

4.4.4 K-Means Modeli Oluşturma

```
set.seed(27)
km.res<-kmeans(data_k_scale, 2, nstart=1)
print(km.res)
```

Şekil 74 K-Means Modeli Oluşturma Kodları

Oluşturduğumuz data_k_scale veri setini bulduğumuz optimal değer ile modelini oluşturuyoruz.

```
> print(km.res)
K-means clustering with 2 clusters of sizes 196, 689

Cluster means:
  Cinsiyet      Yas      Bolge      Egitim      soru1      soru2      soru3      soru4      soru5      soru6
1  0.06700903  0.14854931  0.05072048  0.05692363 -0.3916273  0.5486619  0.3628245 -0.6403730 -0.5607178 -0.9060640
2 -0.01906207 -0.04225786 -0.01442847 -0.01619308  0.1114063 -0.1560780 -0.1032128  0.1821671  0.1595075  0.2577482
  soru7      soru8      soru9      soru10      parti
1  1.3343630 -0.7298754  1.0503535  0.8950350 -0.5847869
2 -0.3795866  0.2076278 -0.2987943 -0.2546108  0.1663545
```

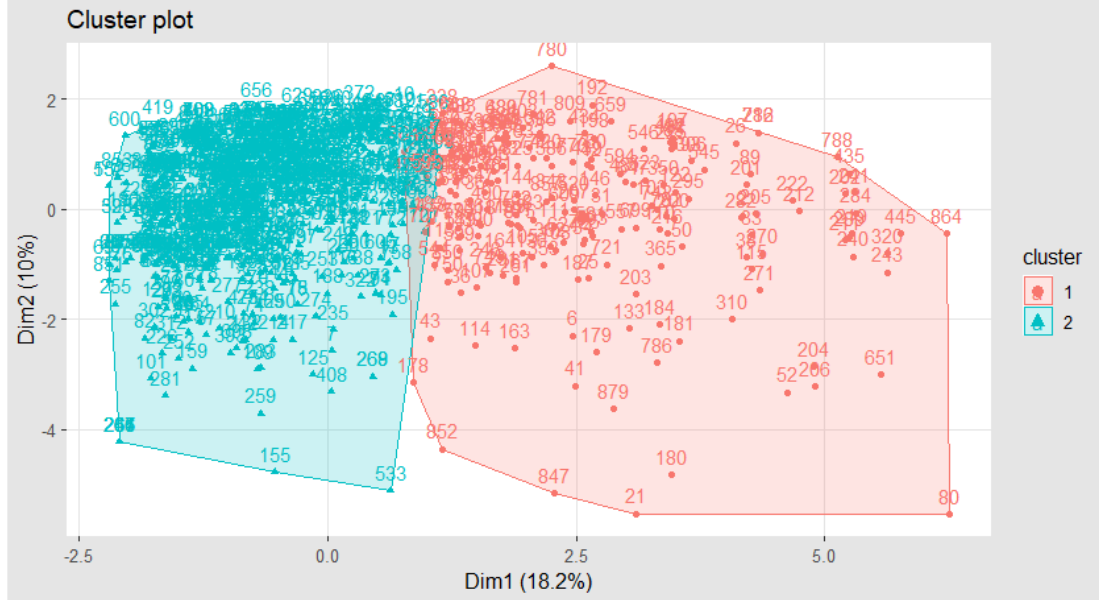
Şekil 75 K-Means Modeli

Şekil 75’de K-Means modelinin 2 kümesinin oluşturduğu değerleri görebiliyoruz. 1.Kümemiz 196 değer alırken 2.Kümemiz ise 689 değer almıştır.

4.4.5 K-Means Modelinin Grafiği

```
fviz_cluster(km.res, data=data_k_scale)+ tma
```

Şekil 76 K-Means Modeli Grafiği Kodları



Şekil 77 K-Means Modeli Grafiği

Şekil 77’de K-Means modelimizin grafiğini görmekteyiz. Grafiği incelediğimiz zaman 1.Kümemizin değerlerinin birbirlerinden uzak değerler olduğu görülmektedir.

2.Kümemizin değerleri ise birbirlerine daha yakındır. Kümelerin 1 ile 1,25 değerler arasında kesiştiği gözlemlenmiştir.

4.4.5.1 K-Means Modelinin Grafiğinin İncelenmesi

```
km.res.data <- km.res[["cluster"]]  
km.res.data <- as.factor(km.res.data)  
str(km.res.data)  
summary(km.res.data)
```

Şekil 78 km.res.data veri seti oluşturma

Şekil 78’de km.res verisinden cluster sütununu alıp km.res.data isimli veri setini oluşturuyoruz. Oluşturduğumuz veri setini İnteger’den Factore çeviriyoruz.

```
> str(km.res.data)  
Factor w/ 2 levels "1","2": 2 1 2 2 2 1 2 2 2 2 ...  
> summary(km.res.data)  
 1    2  
196 689
```

Şekil 79 km.res.data veri seti incelemesi

Şekil 79’da Str fonksiyonunu kullanarak verimizin cinsini kontrol ettik, Summary fonksiyonu ile 1.kümemizin 196 2.kümemizin ise 689 olduğu görülmüştür.

```
aa<- data.frame(kumeler=c(km.res.data),  
                parti=c(data$parti),  
                stringsAsFactors = FALSE)  
  
view(aa)  
summary(aa)
```

Şekil 80 küme ve partilerle yeni veri seti oluşturma

Şekil 80’de oluşturduğumuz km.res.data veri seti ile data veri setindeki parti sütununu birleştirip yeni bir veri seti oluşturduk.

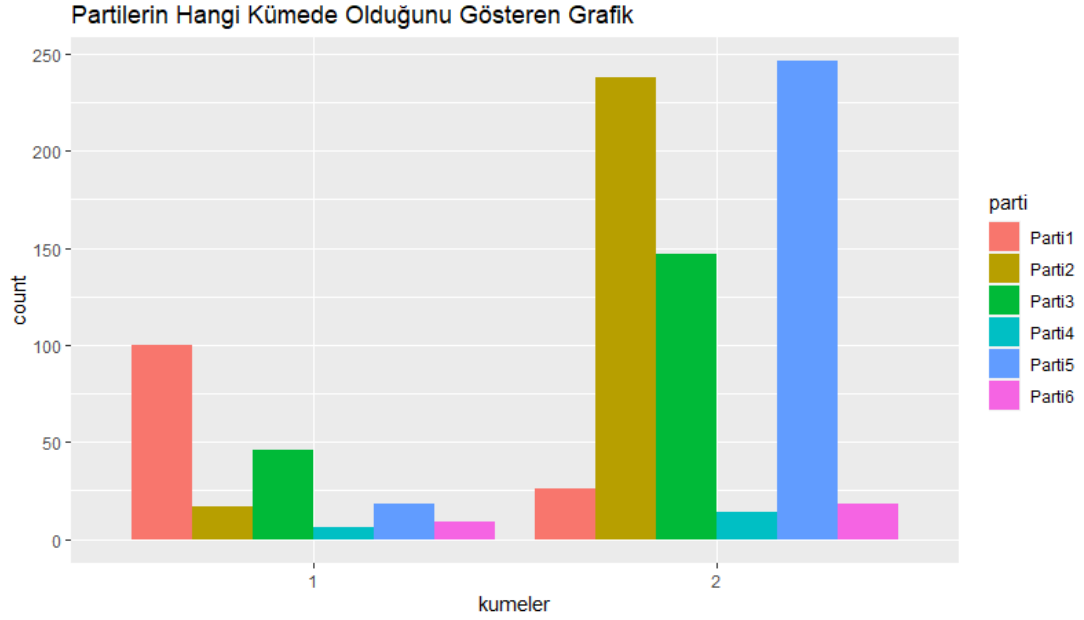
```
> summary(aa)  
kumeler      parti  
1:196      Parti1:126  
2:689      Parti2:255  
           Parti3:193  
           Parti4: 20  
           Parti5:264  
           Parti6: 27
```

Şekil 81 aa veri setinin özeti

Oluşturduğumuz yeni veri setinin özet tablosu Şekil 81’de görülmektedir.

```
karsilastirma <- ggplot(aa, aes(kumeler, fill = parti))
karsilastirma + geom_bar(position = "dodge") +
  labs(title = "Partilerin Hangi Kümede Olduğunu Gösteren Grafik")
```

Şekil 82 Partilerin Hangi Kümede Olduğunu Gösteren Grafik kodu



Şekil 83 Partilerin Hangi Kümede Olduğunu Gösteren Grafik

Şekil 83’de Partilerin hangi kümede olduğunu görmekteyiz. Bu grafiğe baktığımız zaman 1.kümede 196 kişinin çoğunu parti 1’in oluşturduğunu görebiliyoruz. 2.Kümede ise 689 kişinin diğer partilerin çoğunluğunun oluşturduğunu görebilmekteyiz.

Bu durumda kümeler karşılaştırıldığı zaman Parti 1’in diğer partilerden ayrıldığını gözlemleyebiliriz.

Sonuç

Kullandığımız “Türkiye Politik Görüşü” veri setine göre sorular, yaşanan bölge, yaş ve eğitim durumuna bağlı olarak kişilerin desteklediği partileri incelemiş olduk. İncelediğimiz veriyi grafiklerle yorumladık ve karşılaştırmalarını yaptık. Karşılaştırmalar sonucunda Şekil 42’de en fazla eğitim durumunun Marmara bölgesinde, Şekil 44’de en fazla partilere verilen desteğin Marmara bölgesinde, Şekil 46’da en çok partilere destek veren yaş aralığı 0-18, Şekil 48’de en fazla partilere destek veren eğitim durumu ise Lise olarak çıkmıştır. Bu sonuçlara bakıldığı zaman katılımcıların çoğunun Marmara bölgesinde, 0-18 yaş aralığında, eğitim durumunun ise Lise olduğu gözlemlenmiştir. Şekil 50’de partiler arasında en çok desteklenen Parti 5 olduğu görülmektedir.

Naive Bayes ve Karar Ağacı modellerini kullanarak veri setinin performansını değerlendirdik çıkan sonuçlara göre Naive Bayes modelinde 0,45’lik bir doğruluk oranı alırken Karar Ağacı modelinde 0,43’lük bir oran aldık. Bu oranlara bakıldığı zaman veri setimizin ortalama 0,44’lük bir doğruluk oranı aldığını görmekteyiz, doğruluk oranımız 1 e uzak olduğundan dolayı güvenilir bir veri olmadığı gözlemlenmiştir. Doğruluk oranları düşük olduğundan dolayı K-Means Algoritması kullanarak veriyi tekrardan inceledik. K-Means incelemesi sonucunda 1.kümede 196 kişinin çoğunu parti 1’in oluşturduğunu görebiliyoruz. 2.Kümede ise 689 kişinin diğer partilerin çoğunluğunun oluşturduğunu görebilmekteyiz. Kümeler karşılaştırıldığı zaman Parti 1’in diğer partilerden ayrıldığını gözlemleyebiliriz.

Seçtiğimiz veri seti için iki adet sınıflandırma modeli, bir adet ise kümeleme modeli oluşturarak inceledik. Veri setimiz için model olarak kümeleme modelinin kullanılmasının daha doğru olduğunu tespit ettik.

Değerlendirme

Büyük veri günlük hayatımızın içerisinde var olmaktadır. Büyük veriyi videolar, görseller, log dosyaları ve ses dosyaları oluşturabilir. Büyük Veri Analizi dersi sayesinde büyük veriyi incelenebilir bir veri seti olarak hazırlamayı öğrendik. Öğrendiğimiz bilgiler sayesinde bu projeyi gerçekleştirdik. Kaggle'dan bulduğumuz Türkiye Politik Görüşü veri setini inceledik ve modelledik.

Bu yaptığımız proje sayesinde derste öğrendiğimiz bilgileri hem bir projede kullanıp hem de pekiştirmiş olduk ayrıca çalıştığımız dönem boyunca grup çalışması yeteneğimizi geliştirmiş olduk.

KAYNAKÇA

10 Tips for Choosing the Optimal Number of Clusters (2019, 27 Ocak) Eriřim Adresi: <https://towardsdatascience.com/10-tips-for-choosing-the-optimal-number-of-clusters-277e93d72d92>

Balaban E. ve Kartal E. (bs.2,2018). *Veri madencilięi ve makine öğrenmesi temel algoritmaları ve r dili ile uygulamaları*. İstanbul: Çaęlayan Yayın

CELEBBİ K. (2018, 08 Ağustos) R ile Karar Ağaçları. [Blog Yazısı] Eriřim Adresi: <https://www.kutaycelebi.com/2018/08/08/r-ile-karar-agaclari/>

DEMİREL S. (2018, 08 Kasım) R ekosisteminde ggplot2 paketi ile veri nasıl görselleřtirilir?. [Blog Yazısı] Eriřim Adresi: <https://www.newslabturkey.org/2018/11/08/r-ekosisteminde-ggplot2-paketi-ile-veri-nasil-gorsellestirilir/>

Ertuęrul AKTAN (2018). *Büyük Veri: Uygulama Alanları, Analitięi ve Güvenlik Boyutu* (Makale) Eriřim Adresi: <https://dergipark.org.tr/tr/>

MANAV N.(2021). BÜYÜK VERİ ANALİZİ. Yayımlanmamış ders notu, Yönettim Biliřim Sistemleri Bölümü, İstanbul Geliřim Üniversitesi, İstanbul. Eriřim Adresi: https://alms-cdn.gelisim.edu.tr/BlmsFileRepo/1/course/368490/activity/422976/ders_notu2.txt

R İLE KÜMELEME ANALİZİ (CLA) ALGORİTMALARI (2016, 02 Temmuz) Eriřim Adresi: <https://www.linkedin.com/pulse/r-ile-k%C3%BCmeleme-analizi-cla-algoritmaları-datalab-tr/?originalSubdomain=tr>

R'da Veri Görselleřtirme. (2020, 23 Mayıs) Eriřim Adresi:

<https://rpydaneogrendim.github.io/rViz/>

Rstudio (Sürüm 2021.09.1+372) [Yazılım]. Tedarik edilebileceęi adres:

<https://www.rstudio.com/>

Sosyal Bilimler Arařtırmaları için R. (2018, 11 Aralık) Eriřim Adresi:

<https://bookdown.org/content/2096/>

Turkey Political Opinions.(2018) Eriřim Adresi: <https://www.kaggle.com/>

Ünal SEZER (2088). *KARAR AęAęLARININ BİRLİKTELİK KURALLARI İLE İYİLEřTİRİLMESİ*

(YÜKSEK LİSANS TEZİ) Eriřim Adresi:

<http://dspace.kocaeli.edu.tr:8080/xmlui/bitstream/handle/11493/1236/232705.pdf?sequence=1&isAllowed=y>