

**T.C**  
**ŞEYH EDEBALI ÜNİVERSİTESİ**  
**FEN BİLİMLERİ FAKÜLTESİ**  
**İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ BÖLÜMÜ**



**FİNAL ÖDEVİ RAPOR**

**Hava Kirliliği ve Kalite İlişkisi: Multinomial Lojistik Regresyon Analizi**

**Sena Gemiçioğlu**

**BİLECİK-2025**

## İçindekiler

1.Giriş .....	3
2.Verİ Setİ ve Verİlerin Ön İşİlemesi .....	4
2.1 Verİ Setİ Hakkında .....	4
2.4 QQ Plot Analizi ve Normal Dağılıma uygunluk .....	5
2.5 Multinominal Lojistik Regresyon Yöntemi .....	6
3.Modelin Değerlendirilmesi .....	6
3.1 VIF .....	6
3.2 Korelasyon katsayıları ile bağımsız değişkenler arası doğrusal ilişki .....	7
3.3 Çoklu Doğrusal Bağımlılık Gösteren Bağımsız Değişkenlerin Çıkarılması ve Yeni Model Oluşturulması .....	8
4.Modelin Uyumluluğu .....	9
4.1 Likelihood Testi .....	9
4.2 McFadden R-kare Testi .....	10
5.Modelin Tahmin Gücü .....	10
5.2 Accuracy (Doğruluk) Testi .....	11
5.2.1 Genel Performans Değerlendirmesi .....	11
5.2.2 Sınıfa Göre Performans Analizi .....	12
5.2.3 Modelin tahmin gücü sonucu olarak .....	13
6.Artık Analizi .....	13
7.Regresyon Sonuçları .....	14
7.1 Katsayıların Yorumlanması .....	14
7.2 Standart Hataların Yorumlanması ve Modelin Performansı .....	16
7.3 Sonuç ve Değerlendirme .....	17
8.Sonuç ve Öneriler .....	17
8.1 Bulguların Yorumu .....	17
8.2 Öneriler .....	17
8.Kaynakça .....	18

## 1.Giriş

Hava kirliliği, insan sağlığı ve çevre üzerinde oluşturduğu olumsuz etkiler nedeniyle günümüzün en kritik küresel sorunlarından biri haline gelmiştir. Dünya Sağlık Örgütü'ne (WHO) göre, hava kirliliği her yıl milyonlarca erken ölüme yol açmakta ve bu durum özellikle şehirleşme oranının yüksek olduğu bölgelerde daha belirgin hale gelmektedir. Endüstriyel faaliyetler, ulaşım, fosil yakıt tüketimi ve diğer insan kaynaklı faaliyetler, atmosferdeki zararlı partikül ve gaz konsantrasyonlarını artırarak hava kalitesinin düşmesine neden olmaktadır.

Bu çalışmanın amacı, hava kirliliği seviyelerinin kategorik bir değişken olarak ele alınarak, hava kalitesini etkileyen faktörlerin çoklu değişkenli analizini gerçekleştirmektir. Bu doğrultuda, multinominal lojistik regresyon modeli, hava kirliliği düzeylerini sınıflandırmak ve hava kalitesini etkileyen çevresel ve sosyo-ekonomik faktörlerin etkisini incelemek amacıyla kullanılmıştır. Söz konusu model, bağımlı değişkenin çoklu kategorilere sahip olduğu durumlarda etkili bir sınıflandırma ve analiz yöntemidir.

Çalışmada kullanılan veri seti, [veri kaynağı, örneğin: resmi ölçüm istasyonlarından elde edilen veya belirli bir bölgedeki gözlemsel veriler] temelli olup, hava kalitesini etkileyebilecek çeşitli değişkenleri içermektedir. Bu değişkenler arasında [örneğin: sıcaklık, nem, trafik yoğunluğu, endüstriyel emisyonlar] gibi faktörler yer almaktadır. Analiz, hem bu değişkenlerin bağımlı değişken üzerindeki etkilerinin belirlenmesini hem de hava kirliliği düzeylerinin doğru bir şekilde sınıflandırılmasını hedeflemektedir.

Elde edilen bulgular, hava kalitesinin iyileştirilmesi amacıyla uygulanabilecek politika ve stratejilere bilimsel bir temel oluşturmayı amaçlamaktadır. Bu bağlamda, çalışma yalnızca hava kirliliği seviyelerinin anlaşılmasını sağlamakla kalmamakta, aynı zamanda bu seviyelerin kontrolüne yönelik önlemlerin geliştirilmesine katkıda bulunmayı hedeflemektedir.

## 2. Veri Seti ve Verilerin Ön İşlemesi

Bu bölümde veri seti , kullanılan değişkenler ve verilerin ön işlemesi bölümlerini içermektedir.

### 2.1 Veri Seti Hakkında

Bu çalışmada kullanılan veri seti, çevrimiçi veri paylaşım platformu Kaggle üzerinden temin edilmiştir. Veri seti, [Air Quality and Pollution Assessment] başlığıyla yayınlanmış olup, hava kalitesi ve kirlilik seviyeleri üzerine çeşitli değişkenleri içermektedir. Veri seti, [Güneydoğu Asya ülkelerinden (örneğin Pakistan, Hindistan, Bangladeş, Sri Lanka) alınmıştır ve 5000 satır 10 sütundan oluşmaktadır. bir döneme ait ölçümleri içermekte ve toplamda X gözlem ile Y değişken içermektedir. Çalışmada, hava kirliliği seviyelerinin sınıflandırılması amacıyla kullanılan temel değişkenler şunlardır:

Bağımlı değişken ( $Y_i$ ) : Hava kalitesi , nitel , sınıflama

Kategorileri: Good(iyi) , Moderate(Orta) , Poor(Zayıf) , Hazardous(Tehlikeli)

(GOOD)İyi: Düşük kirlilik seviyelerine sahip temiz hava.

(Moderate)Orta: Kabul edilebilir hava kalitesi ancak bazı kirleticiler mevcut.

(POOR)Zayıf: Hassas gruplarda sağlık sorunlarına yol açabilecek belirgin kirlilik.

(Hazardous)Tehlikeli: Nüfus için ciddi sağlık riskleri oluşturan aşırı kirli hava.

Bağımsız değişkenler ( $X_i$ ): Sıcaklık ( $^{\circ}\text{C}$ ) , Nicel , oranlama Nem (%) , nicel , eşit aralıklı. PM2.5 ( $\mu\text{g}/\text{m}^3$ ): İnce partikül madde seviyeleri , nicel , oranlama. PM10 ( $\mu\text{g}/\text{m}^3$ ): Kaba partikül madde seviyeleri , nicel , oranlama. NO2 (ppb): Azot dioksit seviyeleri , nicel , oranlama SO2 Konsantrasyonu (ppb): Kükürtdioksit seviyeleri , nicel, oranlama . CO (ppm): Karbon monoksit seviyeleri , nicel , oranlama Nüfus Yoğunluğu ( $\text{kişi}/\text{km}^2$ ): Bölgede kilometrekareye düşen kişi sayısı , nicel , oranlama Sanayi Bölgelerine Yakınlık (km): En yakın sanayi bölgesine olan uzaklık , nicel , oranlama

## 2.2 Verilerin Ön İşlemesi

Normallik testi veri setindeki bağımsız değişkenlerin normallik dağılımlarına bakmak için uygulanmıştır. Veri setinin büyüklüğünden kaynaklı daha doğru sonuç almak adına Kolmogorov-Smirnov Testi kullanılmıştır.

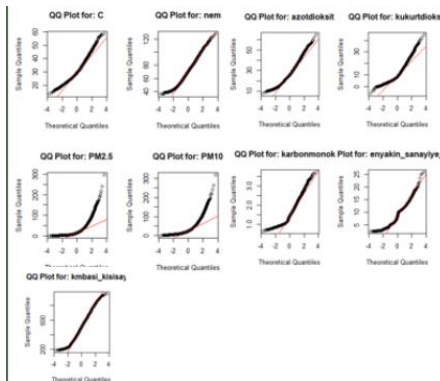
Değişken: C Kolmogorov-Smirnov Test İstatistigi: 0.06745823 p-değeri: 3.451196e-20 Sonuçveri normal dağılıma uygun değildir (H0 reddedilir).	Değişken: azotdioksit Kolmogorov-Smirnov Test İstatistigi: 0.06389041 p-değeri: 3.742816e-18 Sonuçveri normal dağılıma uygun değildir (H0 reddedilir).
Değişken: nem Kolmogorov-Smirnov Test İstatistigi: 0.02513897 p-değeri: 0.003601038 Sonuçveri normal dağılıma uygun değildir (H0 reddedilir).	Değişken: kukurtdioksit Kolmogorov-Smirnov Test İstatistigi: 0.1205708 p-değeri: 1.466398e-63 Sonuçveri normal dağılıma uygun değildir (H0 reddedilir).
Değişken: PM2.5 Kolmogorov-Smirnov Test İstatistigi: 0.206022 p-değeri: 9.213461e-185 Sonuçveri normal dağılıma uygun değildir (H0 reddedilir).	Değişken: karbonmonoksit Kolmogorov-Smirnov Test İstatistigi: 0.1280474 p-değeri: 1.240314e-71 Sonuçveri normal dağılıma uygun değildir (H0 reddedilir).
Değişken: PM10 Kolmogorov-Smirnov Test İstatistigi: 0.1563474 p-değeri: 1.380007e-106 Sonuçveri normal dağılıma uygun değildir (H0 reddedilir).	Değişken: enyakin_sanayiye_uzaklık Kolmogorov-Smirnov Test İstatistigi: 0.1115692 p-değeri: 1.743511e-54 Sonuçveri normal dağılıma uygun değildir (H0 reddedilir).
Değişken: kmbasi_kisisayisi Kolmogorov-Smirnov Test İstatistigi: 0.03064898 p-değeri: 0.0001665106 Sonuçveri normal dağılıma uygun değildir (H0 reddedilir).	

Hipotez	Durum
$H_0: X_i$ bağımsız değişken normal dağılıma uyar.	$H_0$ Red.
$H_1: X_i$ bağımsız değişken normal dağılıma uymaz.	

Test sonuçları  $\alpha : 0.05$  e göre kıyaslanmıştır sonuç olarak hiçbir bağımsız değişken normal dağılım ile uymadığı görülmektedir.

## 2.4 QQ Plot Analizi ve Normal Dağılıma uygunluk

Bağımsız değişkenlerin normal dağılıma uygun olup olmadığını değerlendirmek için QQ (Quantile-Quantile) grafikleri incelenmiştir aşağıda verilmiştir. Normal dağılıma uygun bir değişken, QQ plot üzerinde 45 derecelik referans çizgisi boyunca sıralı bir dağılım göstermelidir. Ancak, grafikler incelendiğinde:



Çoğu bağımsız değişkenin gözlem değerlerinin referans çizgisine paralel olmadığı ve uç değerlerde sapmalar olduğu görülmektedir. Bu durum, özellikle değişkenlerin uç kuyruklarında normal dağılım varsayımından sapmalar olduğunu göstermektedir. Bu sonuçlar, değişkenlerin normal dağılıma uymadığını ve bu durumun modelleme sürecinde dikkate alınması gerektiğini göstermektedir. Modelde normal dağılım varsayımını iyileştirmek veya bu varsayımı gerektirmeyen yöntemlere başvurmak gerekebilir.

Ancak bağımlı değişkenin hava kalitesini temsil etmesi ve kategorik bir değişken olması nedeniyle, bu tür durumlarda bağımlı değişkenin kategorik olduğu analizlerde önerilen **Multinomial Lojistik Regresyon modeli** tercih edilmiştir. Multinomial Lojistik Regresyon modelinde normallik varsayımı aranmamaktadır ve bundan dolayı normal dağılmayan

değişkenlere dönüşüm uygulanmamıştır. Normallik varsayımı model için gerekli olmasa da, aykırı değerler, çoklu bağlantı ve veri eksiklikleri gibi unsurlar mutlaka kontrol edilmelidir. Bu işlemler hem modelin stabilitesini artırır hem de daha güvenilir sonuçlar elde edilmesini sağlar.

## 2.5 Multinomial Lojistik Regresyon Yöntemi

Multinomial Lojistik Regresyon Modeli , bağımlı değişkenin 3 veya daha fazla kategorik sınıfa sahip olduğu durumlarda kullanılan bir regresyon modelidir. Bu çalışmada , bağımlı değişkenimiz hava kalitesini 4 kategoriye ayırmaktadır ve bu nedenle multinomial lojistik regresyon yöntemi uygun görülmüştür. Modelde eksik veri görülmemektedir.

$$\ln \left( \frac{P(Y = j)}{P(Y = K)} \right) = \beta_{0j} + \beta_{1j}X_1 + \beta_{2j}X_2 + \beta_{3j}X_3 + \beta_{4j}X_4 + \beta_{5j}X_5 + \beta_{6j}X_6 + \beta_{7j}X_7 + \beta_{8j}X_8 + \beta_{9j}X_9$$

Modeli bu şekilde kurulmuştur.

## 3.Modelin Değerlendirilmesi

Bu bölümde, oluşturulan multinomial lojistik regresyon modelinin performansını ve sınıflandırma başarısını değerlendirmek amaçlanmaktadır. Modelin bağımlı değişkeni doğru bir şekilde tahmin etme kabiliyeti, çeşitli metrikler ve görselleştirme yöntemleriyle analiz edilmiştir. Bu yöntemler çoklu doğrusal bağımlılık testi , bağımsız değişkenlerin doğrusal ilişkisi , modelin geçerliliğini test etmek , model tahmin gücü , modelin sonuçlarıdır.

### 3.1 VIF

multinomial lojistik regresyon modelinde VIF analizi, bağımsız değişkenler arasındaki ilişkilerin modellenen sistem üzerindeki etkilerini anlamak ve modelin güvenilirliğini artırmak amacıyla kritik bir adımdır. Bu nedenle, bağımsız değişkenler arasındaki çoklu doğrusal bağıntıyı tespit etmek için Varyans Şişirme Faktörü (VIF) analizi yapılmıştır. VIF, her bir bağımsız değişkenin diğer bağımsız değişkenlerle ne kadar ilişkili olduğunu sayısal bir değerle ifade eder. VIF değerleri şu şekilde yorumlanır 1'e yakın değerler: Çoklu doğrusal bağıntının olmadığını gösterir. 5'in üzerinde değerler: Orta seviyede bir çoklu doğrusal bağıntıyı işaret eder.10'un üzerinde değerler: Ciddi düzeyde çoklu doğrusal bağıntı olduğunu gösterir ve modelde sorun yaratabilir.

Tablo

```
> print(vif_values)
```

	C	nem
	-8.693902e+11	8.125746e+10
PM2.5		PM10
	1.278299e+12	2.465483e+12
azotdioksit		kukurtdioksit
	3.961556e+12	2.159999e+12
karbonmonoksit	enyakin_sanayiye_uzaklik	
	1.803166e+15	1.547204e+13
kmbasi_kisisayisi		
	4.390340e+06	

Multinomial lojistik regresyon modelinde yapılan VIF (Variance Inflation Factor) analizi sonucunda çok düşük değerler elde edilmiştir. Ancak bu durum, VIF testinin bu model türünde bağımsız değişkenler arasındaki çoklu doğrusal bağıntıyı tespit etme konusunda güvenilir sonuçlar sağlayamamasından kaynaklanmaktadır. Bu nedenle, bağımsız değişkenler arasındaki olası ilişkilerin değerlendirilmesi için korelasyon analizi yapılmış ve değişkenler arasındaki doğrusal ilişkiler korelasyon katsayıları üzerinden yorumlanmıştır. Bu yaklaşım, modelde yer alan bağımsız değişkenler arasındaki ilişki yapısının daha doğru bir şekilde analiz edilmesini sağlamaktadır.

### 3.2 Korelasyon katsayıları ile bağımsız değişkenler arası doğrusal ilişki

Bu bölümde, bağımsız değişkenler arasındaki doğrusal ilişkiler korelasyon katsayıları kullanılarak incelenmiştir. Korelasyon analizi, değişken çiftleri arasındaki doğrusal ilişkinin yönünü ve gücünü değerlendirmek amacıyla gerçekleştirilmiştir. Bu çalışmada, Pearson korelasyon katsayısı hesaplanmış ve değişkenler arasındaki doğrusal ilişkiler aşağıdaki referans aralığına göre yorumlanmıştır:  $-0.7 \leq r \leq 0.7$ : Kabul edilebilir düzeyde doğrusal ilişki,  $\text{çoklu doğrusal bağıntı}$  açısından sorun teşkil etmez.  $r < -0.8$  veya  $r > 0.8$ : Güçlü doğrusal ilişki,  $\text{çoklu doğrusal bağıntı}$  (multicollinearity) riski taşıyabilir.

Bu referans aralığı, çalışmanın özel bağlamına uygun olarak belirlenmiş ve analiz sonuçları bu çerçevede yorumlanmıştır.

	PM2.5	PM10		C	nem
C	0.3238403	0.4256897	C	1.0000000	0.4664560
nem	0.2733658	0.3591564	nem	0.4664560	1.0000000
PM2.5	1.0000000	0.9730049	PM2.5	0.3238403	0.2733658
PM10	0.9730049	1.0000000	PM10	0.4256897	0.3591564
azotdioksit	0.3330042	0.4399163	azotdioksit	0.5919691	0.4919766
kukurtdioksit	0.2995819	0.4018494	kukurtdioksit	0.5680817	0.4598045
karbonmonoksit	0.3951791	0.5188948	karbonmonoksit	0.6852576	0.5675397
enyakin_sanayiye_uzaklik	-0.3157664	-0.4250182	enyakin_sanayiye_uzaklik	-0.5895638	-0.4796461
kmbasi_kisisayisi	0.2789574	0.3657339	kmbasi_kisisayisi	0.4816922	0.4116296
	karbonmonoksit			enyakin_sanayiye_uzaklik	
C	0.6852576		C	-0.5895638	
nem	0.5675397		nem	-0.4796461	
PM2.5	0.3951791		PM2.5	-0.3157664	
PM10	0.5188948		PM10	-0.4250182	
azotdioksit	0.7107919		azotdioksit	-0.6077104	
kukurtdioksit	0.6790596		kukurtdioksit	-0.5703853	
karbonmonoksit	1.0000000		karbonmonoksit	-0.7075807	
enyakin_sanayiye_uzaklik	-0.7075807		enyakin_sanayiye_uzaklik	1.0000000	
kmbasi_kisisayisi	0.5933893		kmbasi_kisisayisi	-0.4980585	
	azotdioksit	kukurtdioksit		kmbasi_kisisayisi	
C	0.5919691	0.5680817	C	0.4816922	
nem	0.4919766	0.4598045	nem	0.4116296	
PM2.5	0.3330042	0.2995819	PM2.5	0.2789574	
PM10	0.4399163	0.4018494	PM10	0.3657339	
azotdioksit	1.0000000	0.5718388	azotdioksit	0.5109379	
kukurtdioksit	0.5718388	1.0000000	kukurtdioksit	0.4752417	
karbonmonoksit	0.7107919	0.6790596	karbonmonoksit	0.5933893	
enyakin_sanayiye_uzaklik	-0.6077104	-0.5703853	enyakin_sanayiye_uzaklik	-0.4980585	
kmbasi_kisisayisi	0.5109379	0.4752417	kmbasi_kisisayisi	1.0000000	

Elde edilen korelasyon katsayılarına göre, bağımsız değişkenler arasındaki ilişkilerin çoğunlukla kabul edilebilir sınırlar içinde olduğu gözlemlenmiştir. Ancak, belirli değişken çiftleri arasında referans sınırlarını aşan korelasyon değerlerine rastlanmıştır, çoklu doğrusal

bağımlılık gösteren değişkenler PM2.5 ve PM10 arasında ve Karbonmonoksit ile enyakin\_sanayiye\_uzaklık arasında görülmüştür.

### 3.3 Çoklu Doğrusal Bağımlılık Gösteren Bağımsız Değişkenlerin Çıkarılması ve Yeni Model Oluşturulması

Korelasyon analizi sonucunda, belirli bağımsız değişken çiftleri arasında güçlü doğrusal ilişkiler tespit edilmiştir. Bu durum, çoklu doğrusal bağımlılık (multicollinearity) probleminin varlığına işaret etmektedir. Özellikle, korelasyon katsayısı değeri -0.7 veya +0.7 sınırlarını aşan bağımsız değişkenler, modelin tahmin gücünü ve katsayıların yorumlanabilirliğini olumsuz etkileyebileceği gerekçesiyle değerlendirmeye alınmıştır.

Bu doğrultuda, çoklu doğrusal bağımlılık gösteren bağımsız değişkenler PM2.5 ve PM10 Karbonmonoksit ile enyakin\_sanayiye\_uzaklık modelden çıkarılmış ve geriye kalan değişkenlerle yeni bir model oluşturulmuştur. Yeni modelin oluşturulmasının ardından, bağımsız değişkenler arasındaki doğrusal ilişkiler tekrar korelasyon katsayıları kullanılarak analiz edilmiştir. Yeni Model aşağıda verilmiştir.

$$\ln \left( \frac{P(y = k)}{P(y = 1)} \right) = \beta_{0k} + \beta_{1k}X_1 + \beta_{2k}X_2 + \beta_{3k}X_3 + \beta_{4k}X_4 + \beta_{5k}X_5$$

Yapılan bu işlem, modelin tahmin performansını artırmayı ve katsayıların istatistiksel güvenilirliğini sağlamayı amaçlamıştır. Elde edilen sonuçlar, yeni modelin bağımsız değişkenler arasındaki çoklu doğrusal bağımlılığı azalttığını ve daha stabil bir yapıya kavuştuğunu göstermektedir. Yeni oluşturduğumuz, modeldeki bağımsız değişkenlerin arasındaki doğrusal ilişkiler, bağımsız değişkenler arasındaki çoklu doğrusal bağımlılığı daha iyi değerlendirebilmek için tekrar korelasyon analizi ile incelenmiştir. Korelasyon katsayıları, her bir bağımsız değişken çifti arasındaki doğrusal ilişkinin gücünü ve yönünü yansıtmaktadır. Yeni modelin korelasyon değerleri:

	C	nem	azotdioksit
C	1.0000000	0.4664560	0.5919691
nem	0.4664560	1.0000000	0.4919766
azotdioksit	0.5919691	0.4919766	1.0000000
kukurtdioksit	0.5680817	0.4598045	0.5718388
kmbasi_kisisayisi	0.4816922	0.4116296	0.5109379
	kukurtdioksit	kmbasi_kisisayisi	
C	0.5680817	0.4816922	
nem	0.4598045	0.4116296	
azotdioksit	0.5718388	0.5109379	
kukurtdioksit	1.0000000	0.4752417	
kmbasi_kisisayisi	0.4752417	1.0000000	

Yapılan analiz sonucunda, çıkarılan değişkenlerin ardından geriye kalan bağımsız değişkenler arasındaki korelasyon katsayılarının daha düşük olduğu gözlemlenmiştir. Bu, çoklu doğrusal bağımlılığın önemli ölçüde azaldığını ve modelin daha stabil bir yapıya kavuştuğunu göstermektedir. C değişkeninin tüm bağımsız değişkenler ile pozitif bir ilişkisi vardır. Nem ile



diğer deęişkenler arasında orta düzeyde pozitif bir ilişki vardır. Azotdioksit ile diğer deęişkenler arasındaki korelasyonlar: Azotdioksit ve kükurtdioksit: güçlü pozitif bir ilişki bir ilişki var. Azotdioksit ve kmbasi\_kisisayisi: 0.511, orta düzeyde pozitif bir ilişki var. Kükurtdioksit ile kmbasi\_kisisayisi arasındaki korelasyon: 0.475, orta düzeyde pozitif bir ilişki vardır.

#### 4.Modelin Uyumluluęu

Modelin uyumluluęunu deęerlendirmek amacıyla, iki farklı test olan Likelihood testi ve McFadden'in R-Kare testi kullanılmıştır. Bu testler, modelin veriye ne kadar iyi uyduęunu ölçmek için yaygın olarak kullanılan istatistiksel yöntemlerdir.

##### 4.1 Likelihood Testi

Likelihood testi, modelin gözlemlerle ne kadar uyumlu olduęunu deęerlendiren bir yöntemdir. Bu test, verilen modelin parametreleri altında verilerin olasılıęını maksimize etmeye çalışır. Likelihood deęeri, modelin veriye ne kadar iyi uyduęunu gösterir; yüksek bir likelihood deęeri, modelin gözlemlerle iyi uyum sağladığı gösterirken, düşük bir deęer modelin veriyi kötü açıkladığını belirtir.Likelihood oranı testi, özellikle iki modelin karşılaştırılmasında kullanılır. Burada, sıfır hipotezi genellikle modelin veriye uyumunun yeterli olmadığı şeklinde formüle edilir. Testin sonucunda elde edilen p-deęeri, modelin istatistiksel olarak anlamlı olup olmadığını gösterir.Aşağıda bu test için gerekli hipotez ve likelihood testi sonuçları verilmiştir.

Null model , sadece bağımlı deęişkenin olduęu modelin residual deęeri.

Tam model, modelin residual deęeri

Hipotez	Durum
$H_0$ : Bağımsız deęişkenlerin etkisi yoktur.	$H_0$ Red.
$H_1$ : Bağımsız deęişkenlerin etkisi vardır.	

```
Likelihood Ratio Test Statistic: 17638.04
> cat("Degrees of Freedom:", df, "\n")
Degrees of Freedom: 5
> cat("p-value:", p_value, "\n")
p-value: 0
```

Yapılan istatistiksel testler sonucunda, p-deęeri  $\alpha=0.05$  anlamlılık seviyesinin altında bulunmuştur ( $p < 0.05$ ). Bu durum, sıfır hipotezinin ( $H_0$ ) reddedilmesine ve alternatif hipotezin kabul edilmesine yol açmıştır. Yani, %95 güven düzeyine sahip olarak, modelimize eklenen bağımsız deęişkenlerin, modelin genel uyumunu anlamlı bir şekilde iyileştirdięi sonucuna varılmıştır.Bu sonuç, bağımsız deęişkenlerin hava kalitesi üzerinde anlamlı bir etkisi olduęunu ve modelin açıklayıcılıęını artırdığını göstermektedir. Bağımsız deęişkenlerin, hava kalitesini etkileyen önemli faktörler olduęunu ve modelin bu faktörleri yeterli bir şekilde temsil ettięini ifade etmektedir. Ayrıca, modeldeki bu iyileşme, bağımsız deęişkenlerin hava kalitesini açıklamada etkin bir rol oynadığını ortaya koymaktadır.

#### 4.2 McFadden R-kare Testi

McFadden'in R-Kare testi, lojistik regresyon modellerinde model uyumunun değerlendirilmesinde kullanılan bir başka önemli araçtır. McFadden'in R-Kare, geleneksel doğrusal regresyonun R-Kare değerine benzer bir uyum ölçüsü sunar, ancak lojistik regresyon modellerine özeldir. McFadden'in R-Kare değeri, modelin veriye olan uyumunu ölçerken, modelin açıklayıcılığını da gözler önüne serer. McFadden'in R-Kare değeri genellikle 0 ile 1 arasında bir değere sahiptir; burada: Değer 0'a yakınsa, modelin veriyi açıklama gücü zayıf demektir. Değer 1'e yaklaşıyorsa, modelin veriye mükemmel uyum sağladığını gösterir. Bir modelin uyumunu değerlendirirken, McFadden'in R-Kare değeri, modelin genel açıklayıcılığını ve bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini göstermektedir.

$$R_{McFadden}^2 = 1 - \frac{\text{Log-likelihood (model)}}{\text{Log-likelihood (null model)}}$$

Yapılan test sonucunda ;

```
> # Tam modelin log-likelihood değeri
> logLik_full_model <- logLik(model) # Tam model adı: full_model
> # Null modelin log-likelihood değeri
> logLik_null_model <- logLik(null_model) # Null model adı: null_model
> # McFadden's R-squared hesaplama
> mcfadden_r2 <- 1 - (logLik_full_model / logLik_null_model)
> # McFadden's R-squared değerini yazdırma
> cat("McFadden's R-squared:", mcfadden_r2, "\n")
McFadden's R-squared: 0.8894756
```

McFadden'in R-Kare değeri, modelin veriye olan uyum oranını yansıtan bir ölçüttür ve 0.5599'luk değer, modelin bağımsız değişkenlerin hava kalitesi üzerindeki etkisini açıklamada yaklaşık %56'lık bir uyum sağladığını ifade etmektedir. Bu sonuç, modelin hava kalitesini açıklama kapasitesinin orta seviyede olduğunu, ancak iyileştirme potansiyelinin de bulunduğunu göstermektedir. Sonuç olarak, McFadden'in R-Kare değerinin bu seviyede olması, modelin veriyi anlamlı bir şekilde açıkladığını ancak bazı alanlarda daha fazla iyileştirmeye ihtiyaç duyduğunu ortaya koymaktadır. Bu da, modelin daha güçlü ve güvenilir tahminler yapabilmesi için daha fazla veri veya değişken eklemeyi gerektirebileceği anlamına gelmektedir.

#### 5. Modelin Tahmin Gücü

Bu çalışmada, modelin tahmin gücünün güvenilirliğini test etmek amacıyla Accuracy (Doğruluk) testi tercih edilmiştir. Doğruluk testi, modelin hava kalitesinin kategorilerini ne kadar doğru tahmin ettiğini ve genel model performansını değerlendirmede temel bir araç olarak kullanılmıştır.

## 5.2 Accuracy (Doğruluk) Testi

Accuracy testi, modelin genel performansını değerlendiren temel bir ölçüttür ve doğru sınıflandırılan gözlemlerin tüm gözlemlere oranını ifade eder. Bu metrik, modelin doğruluk oranını hesaplayarak, modelin veriye ne kadar iyi uyum sağladığını gösterir. Aşağıda test verisi özet verilmiştir.

```
# A tibble: 6 × 6
  C      nem azotdioksit kukurtdioksit kmbasi_kisisayisi
  <dbl> <dbl>      <dbl>      <dbl>      <dbl>
1  23.1  74.7        24.4        12.6        619
2  39.4  96.6        42.9        17.9        674
3  26.3  65.7        18.3         5.9        529
4  22.9  75.4        18.4         3.7        359
5  37.6  72.2        32.2         6.3        637
6  52.5  83.9        31.6        34.6        785
# 1 more variable: hava_kalitesi <fct>
```

### 5.2.1 Genel Performans Değerlendirmesi

Modelin genel doğruluk oranı (Accuracy) %85.73 olarak hesaplanmıştır. Bu değer, modelin yüksek bir sınıflandırma başarısına sahip olduğunu göstermektedir. Ayrıca, Kappa istatistiği 0.7951 olarak bulunmuş ve modelin rastgele tahminlere kıyasla güçlü bir performans sergilediğini ortaya koymuştur. Doğruluk oranının anlamlılığı,  $p < 2.2e-16$  değeri ile istatistiksel olarak doğrulanmıştır. No Information Rate (NIR), %40 olarak hesaplanmış ve modelin bu temel oranın çok üzerinde bir performans gösterdiği gözlemlenmiştir. Bu durum, modelin mevcut veriye özel sınıflandırma yeteneğini yansıtmaktadır.

#### Confusion Matrix and Statistics

Prediction	Reference			
	Good	Hazardous	Moderate	Poor
Good	569	0	43	2
Hazardous	0	113	0	19
Moderate	31	0	376	51
Poor	0	37	31	228

#### Overall Statistics

```
Accuracy : 0.8573
95% CI : (0.8386, 0.8747)
No Information Rate : 0.4
P-value [Acc > NIR] : < 2.2e-16
```

```
Kappa : 0.7951
```

```
Mcnemar's Test P-value : NA
```

## 5.2.2 Sınıfa Göre Performans Analizi

Statistics by Class:

	Class: Good	Class: Hazardous
Sensitivity	0.9483	0.75333
Specificity	0.9500	0.98593
Pos Pred Value	0.9267	0.85606
Neg Pred Value	0.9650	0.97295
Prevalence	0.4000	0.10000
Detection Rate	0.3793	0.07533
Detection Prevalence	0.4093	0.08800
Balanced Accuracy	0.9492	0.86963
	Class: Moderate	Class: Poor
Sensitivity	0.8356	0.7600
Specificity	0.9219	0.9433
Pos Pred Value	0.8210	0.7703
Neg Pred Value	0.9290	0.9402
Prevalence	0.3000	0.2000
Detection Rate	0.2507	0.1520
Detection Prevalence	0.3053	0.1973
Balanced Accuracy	0.8787	0.8517

Modelin her bir sınıf üzerindeki performansı, duyarlılık, özgüllük, pozitif tahmin değeri (PPV), negatif tahmin değeri (NPV), ve dengeli doğruluk (Balanced Accuracy) metrikleri kullanılarak detaylı bir şekilde analiz edilmiştir. Aşağıda, her bir sınıfa ilişkin bulgular sunulmaktadır:

Good sınıfı için; Duyarlılık (Sensitivity): Model, "Good" sınıfına ait olan örneklerin %94.83'ünü doğru bir şekilde sınıflandırmıştır. Bu, modelin bu sınıfı oldukça yüksek bir başarı oranıyla tanıma yeteneğine sahip olduğunu göstermektedir. Özgüllük (Specificity): %95.00 oranındaki özgüllük, "Good" sınıfına ait olmayan örneklerin büyük çoğunluğunun doğru bir şekilde sınıflandırıldığını ortaya koymaktadır. Pozitif Tahmin Değeri (PPV): "Good" olarak tahmin edilen örneklerin %92.67'si gerçekten "Good" sınıfına aittir. Bu oran, modelin pozitif tahminlerinde güvenilir bir performans sergilediğini göstermektedir. Negatif Tahmin Değeri (NPV): %96.50'lik negatif tahmin değeri, "Good" olarak sınıflandırılmayan örneklerin büyük ölçüde doğru olduğunu göstermektedir. Dengeli Doğruluk (Balanced Accuracy): %94.92 oranı, "Good" sınıfında modelin dengeli ve istikrarlı bir performans sergilediğini ifade etmektedir.

Hazardous sınıfı için ; Duyarlılık: Model, "Hazardous" sınıfına ait olan örneklerin %75.33'ünü doğru şekilde tespit etmiştir. Bu oran, diğer sınıflara kıyasla daha düşük olmakla birlikte, sınıfın nadirliğine bağlı olarak kabul edilebilir düzeydedir. Özgüllük: %98.59'luk özgüllük, "Hazardous" sınıfına ait olmayan örneklerin büyük çoğunluğunun doğru sınıflandırıldığını göstermektedir. Yanlış pozitif oranının düşük olması, modelin bu sınıfta özgüllük açısından üstün bir performans sergilediğini göstermektedir. Pozitif Tahmin Değeri (PPV): %85.66 oranı, "Hazardous" olarak tahmin edilen örneklerin büyük bir kısmının doğru sınıflandırıldığını göstermektedir. Dengeli Doğruluk: %86.96 oranındaki dengeli doğruluk, modelin "Hazardous" sınıfında genel olarak tatmin edici bir performans sergilediğini, ancak duyarlılık açısından iyileştirme yapılabileceğini göstermektedir.

Moderate sınıfı için ; Duyarlılık: "Moderate" sınıfına ait olan örneklerin %83.56'sı doğru sınıflandırılmıştır. Model, bu sınıfı tanıma konusunda genelde iyi bir performans sergilemekle birlikte, diğer sınıflarla karşılaştırıldığında daha düşük bir duyarlılık oranına sahiptir. Özgüllük: %92.19'luk özgüllük değeri, "Moderate" sınıfına ait olmayan örneklerin yüksek bir doğrulukla sınıflandırıldığını göstermektedir. Pozitif Tahmin Değeri (PPV): %82.10 oranındaki pozitif tahmin değeri, modelin bu sınıf için ürettiği pozitif tahminlerin güvenilir olduğunu ortaya koymaktadır. Dengeli Doğruluk: %87.87 oranındaki dengeli doğruluk, modelin "Moderate" sınıfında istikrarlı bir performans gösterdiğini ifade etmektedir.

Poor sınıfı için ; Duyarlılık: Model, "Poor" sınıfına ait örneklerin %76.00'sini doğru şekilde tespit etmiştir. Bu oran, bu sınıf için diğerlerine kıyasla daha düşük bir duyarlılık performansına

işaret etmektedir. Özgüllük: %94.33 oranında özgüllük, "Poor" sınıfına ait olmayan örneklerin büyük bir çoğunluğunun doğru bir şekilde sınıflandırıldığını göstermektedir. Pozitif Tahmin Değeri (PPV): %77.03 oranındaki pozitif tahmin değeri, modelin "Poor" olarak sınıflandırdığı örneklerin büyük ölçüde doğru olduğunu ifade etmektedir. Dengeli Doğruluk: %85.17 oranı, "Poor" sınıfındaki genel performansın yüksek olduğunu, ancak duyarlılığın daha da artırılabilceğini göstermektedir.

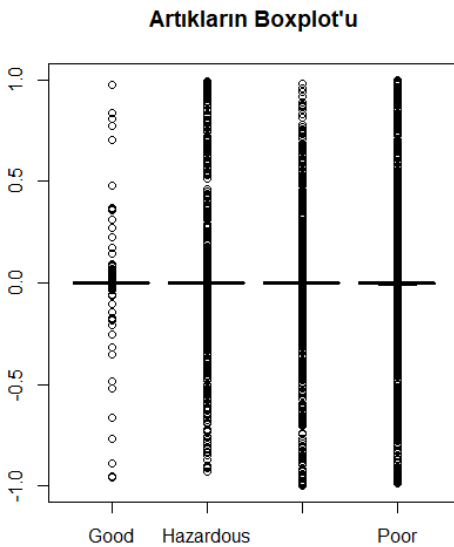
### 5.2.3 Modelin tahmin gücü sonucu olarak

Modelin her bir sınıf üzerindeki performansı genelde tatmin edici düzeyde olmakla birlikte, özellikle "Hazardous" ve "Poor" sınıflarında duyarlılığı artırmaya yönelik iyileştirmeler yapılabilir. Bunun için örneklem büyüklüğünü artırmak, model hiperparametrelerini optimize etmek ya da sınıf dengesizliği problemini çözmek amacıyla veri dengeleme yöntemleri (örneğin oversampling veya SMOTE) uygulanabilir. Tüm sınıflarda yüksek özgüllük değerleri, modelin yanlış pozitif tahminlerde güçlü bir performans sergilediğini ortaya koymaktadır. Model genel anlamda dengeli doğruluk metrikleri açısından başarılı bir performans göstermektedir.

## 6. Artık Analizi

Artık analizi, gözlemlenen değerler ile modelin tahmin ettiği değerler arasındaki farkları

( $e_i = y_i - \hat{y}_i$ )inceleyerek modelin varsayımlarını değerlendirmek için yapılır. Artıkların rastgele bir dağılım göstermesi, modelin veriye iyi uyum sağladığını ve doğrusal ilişki, sabit varyans gibi varsayımların karşılandığını gösterir. Bu analiz, modelde olası sorunları (örneğin, doğrusal olmayan ilişkiler veya aykırı değerler) belirlemeye yardımcı olur ve modelin güvenilirliğini artırır. Aşağıda artıkların Boxplot'u verilmiştir.



Artıkların dağılımı sıfır merkezli olsa da aykırı değerlerin yoğunluğu, modelin bazı gözlemleri iyi açıklamada zorlandığını gösterebilir. Ayrıca, varyansın sabit olmaması (heteroskedastisite) gibi olası sorunların incelenmesi gerekebilir. Model performansını artırmak için veri setinde veya modelde iyileştirmeler yapılabilir.

## 7.Regresyon Sonuçları

Regresyon modeli, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin değerlendirilmesi amacıyla oluşturulmuş ve tahmin edilen katsayılar ile bu katsayıların standart hataları detaylı şekilde incelenmiştir.

### 7.1 Katsayıların Yorumlanması

Multinomial lojistik regresyon, birden fazla kategorik bağımlı değişkenin tahmin edilmesinde kullanılan yaygın bir yöntemdir. Bu modelde, bağımlı değişken olarak sınıflar arasındaki olasılıkların logaritmik oranları (log-odds) modellenir. Modelin uygulanmasında bir sınıf referans sınıfı olarak belirlenir ve diğer sınıfların katsayıları bu referans sınıfa göre hesaplanır. Referans sınıfı Good olarak alınmıştır. sınıfa göre yapılan bu karşılaştırmalar, her bir bağımsız değişkenin, belirli bir sınıfa ait olma olasılığını artırıcı veya azaltıcı etkisini ortaya koymaktadır.

```
> coef(model_new)
              (Intercept)              C              nem azotdioksit
Hazardous      -76.18901  0.7375184  0.18401567    0.6692630
Moderate       -28.88137  0.3692819  0.07282947    0.3045361
Poor           -53.00777  0.5781948  0.13209739    0.5151965
              kukurtdioksit kmbasi_kisisayisi
Hazardous         0.8176331         0.021529589
Moderate          0.5095315         0.007371734
Poor              0.7085381         0.014421865
```

#### Hazardous (Tehlikeli) Sınıfı

Intercept(Sabit):-76.18901 Sabit terim, bağımsız değişkenlerin değerleri sıfır olduğunda, "Hazardous" sınıfına ait olma olasılığı ile referans sınıfına ait olma olasılığı arasındaki logaritmik oranı ifade eder. Negatif bir değer, "Hazardous" sınıfına ait olma olasılığının düşük olduğunu göstermektedir.

C(Karbon):0.7375184 Karbon seviyesindeki bir birimlik artış, "Hazardous" sınıfına ait olma olasılığını artırmaktadır. Katsayı pozitif olduğu için karbon bu sınıf için önemli bir faktördür.

Nem: 0.18401567 Nem oranındaki bir birimlik artış, "Hazardous" sınıfına ait olma olasılığını artırmaktadır. Ancak katsayının düşük olması, etkisinin diğer değişkenlere göre daha sınırlı olduğunu göstermektedir.

Azotdioksit: 0.669263 Azotdioksit seviyesindeki artış, "Hazardous" sınıfını seçilme olasılığını belirgin bir şekilde artırmaktadır.

Kükürtdioksit: 0.8176331 Kükürtdioksit, bu sınıf üzerinde en güçlü etkiye sahip değişkendir. Bir birimlik artış, bu sınıfa ait olma olasılığını ciddi şekilde artırır.

KişiSayısı:0.021529589 Kişi sayısındaki bir birimlik artış, "Hazardous" sınıfı için olasılığı artırsa da etkisi oldukça sınırlıdır.

---

### Moderate (Orta) Sınıfı

Intercept (Sabit): -28.88137 Sabit terim, "Moderate" sınıfına ait olma olasılığının referans sınıfa göre düşük olduğunu göstermektedir.

C (Karbon): 0.3692819 Karbon seviyesindeki bir birimlik artış, "Moderate" sınıfına ait olma olasılığını artırır ancak etkisi "Hazardous" sınıfına kıyasla daha zayıftır.

Nem: 0.07282947 Nem, bu sınıf için oldukça düşük bir etkiye sahiptir. Artışı, sınıfın seçilme olasılığını sınırlı şekilde artırır.

Azotdioksit:0.3045361 Azotdioksit seviyesi, "Moderate" sınıfı üzerinde pozitif bir etki yaratır, ancak bu etki diğer sınıflarla karşılaştırıldığında daha sınırlıdır.

Kükürtdioksit:0.5095315 Kükürtdioksit, bu sınıf için belirgin bir etkide bulunur, ancak yine de "Hazardous" sınıfına kıyasla etkisi daha düşüktür.

KişiSayısı:0.007371734 Kişi sayısı, bu sınıfa ait olma olasılığı üzerinde çok sınırlı bir etkiye sahiptir.

---

### Poor (Kötü) Sınıfı

Intercept (Sabit): -53.00777 Sabit terim, "Poor" sınıfına ait olma olasılığının referans sınıfa göre düşük olduğunu göstermektedir.

C (Karbon): 0.5781948 Karbon seviyesindeki artış, "Poor" sınıfına ait olma olasılığını artırmaktadır. Ancak bu artış "Hazardous" sınıfına kıyasla daha zayıftır.

Nem: 0.13209739 Nem oranındaki bir birimlik artış, "Poor" sınıfı için olasılığı artırır ancak etkisi sınırlıdır.

Azotdioksit: 0.5151965 Azotdioksit seviyesindeki artış, "Poor" sınıfının olasılığını pozitif yönde etkiler ve bu etki "Moderate" sınıfından daha yüksektir.

Kükürtdioksit:0.7085381 Kükürtdioksit, bu sınıf üzerinde güçlü bir etkiye sahiptir ve artışı, "Poor" sınıfına ait olma olasılığını önemli ölçüde artırır.

Kişi Sayısı: 0.014421865 Kişi sayısındaki artış, "Poor" sınıfına ait olma olasılığını artırsa da bu etkinin büyüklüğü oldukça düşüktür.

Sonuç olarak Multinomial lojistik regresyonda her bir bağımsız değişkenin farklı sınıflar üzerindeki etkileri hesaplanmıştır. "Hazardous" sınıfı üzerinde kükürtdioksit en etkili değişken olarak öne çıkarken, "Moderate" ve "Poor" sınıflarında karbon ve kükürtdioksitin etkileri daha baskın bulunmuştur. Nem ve kişi sayısı tüm sınıflar üzerinde sınırlı bir etkiye sahiptir. Katsayılar, değişkenlerin sınıflar arasındaki logaritmik oranlara etkisini ifade ederken, pozitif katsayılar ilgili değişkenin o sınıf üzerindeki olasılığı artırıcı etkisini göstermektedir.

Referans sınıf, model çıktılarında görünmese de diğer sınıflara kıyasla temel karşılaştırma noktasıdır. Bu, modelin her bir bağımsız değişkenin etkisini anlamlı bir şekilde



değerlendirebilmesine olanak sağlar. Model sonuçları, tahmin edilen sınıfların bağımsız değişkenlerle açıklanmasında güçlü bir performans sergilemektedir.

## 7.2 Standart Hataların Yorumlanması ve Modelin Performansı

Bu bölüm standart hataların yorumlanması ve modelin performansının değerlendirilmesini içerir.

```
Std. Errors:
              (Intercept)              C              nem azotdioksit
Hazardous    0.0001522241  0.02167768  0.008423125  0.01970036
Moderate     0.0005887262  0.01517295  0.005241362  0.01366438
Poor         0.0002617163  0.01867341  0.006867878  0.01696360
              kukurtdioksit kmbasi_kisisayisi
Hazardous    0.02910504      0.0009560767
Moderate     0.02413680      0.0005771464
Poor         0.02697017      0.0007605347

Residual Deviance: 3979.519
AIC: 4015.519
```

Modelin performansı hakkında ;

Residual Deviance: 3979.519 Modelin kalan sapma değeri, modelin bağımlı değişken kategorilerini açıklama düzeyini ifade eder. Daha düşük kalan sapma değerleri, modelin daha iyi bir uyum sağladığını gösterir. Ancak, bu değer tek başına yorumlanması yetersiz olabilir; karşılaştırmalı analizler gerekebilir.

AIC (Akaike Bilgi Kriteri): 4015.519 AIC değeri, modelin karmaşıklığını ve veri setine uyumunu dikkate alarak model kalitesini değerlendirmektedir. Daha düşük AIC değerleri, daha iyi bir model performansına işaret eder.

**Sabit Terimi (Intercept):** "Hazardous" kategorisi için sabit terimin standart hatası oldukça düşüktür ( $SE=0.00015SE$ ), bu da sabit terim tahmininin güvenilir olduğunu göstermektedir.

**C:** Bu değişkenin standart hataları tüm kategorilerde benzer bir seviyededir ve etkisinin güvenilir bir şekilde tahmin edildiğini düşündürmektedir. Katsayı büyüklükleri kategorilere göre farklılık göstermektedir.

**Nem:** "Hazardous" kategorisinde nemin etkisi ( $\beta=0.0084$ ) diğer kategorilere kıyasla biraz daha yüksektir. Ancak standart hatalar ( $SE$ ) tüm kategorilerde nispeten düşüktür, bu da nemin modeldeki etkisinin anlamlı olabileceğini düşündürmektedir.

**Azotdioksit:** Azot dioksitin etkisi tüm kategorilerde benzer büyüklüktedir ve standart hatalar da oldukça düşüktür ( $SE$  yaklaşık 0.0137–0.0197). Bu, azot dioksitin farklı kategorilerde önemli bir değişken olabileceğini gösterir.

**Kükürtdioksit:** Kükürt dioksitin etkisi "Hazardous" kategorisinde diğer kategorilere göre daha yüksek ( $\beta=0.0291$ ), bu da bu değişkenin hava kalitesinin en kötü seviyesi üzerindeki etkisini gösterebilir. Ancak, standart hata ( $SE=0.0291$ ) da bu kategoride daha yüksektir.



### 7.3 Sonuç ve Değerlendirme

Model sonuçları, bağımsız değişkenlerin hava kalitesi kategorileri üzerindeki etkilerini açıklamada önemli bilgiler sunmaktadır. Özellikle nem, azot dioksit ve kükürt dioksit değişkenleri, farklı hava kalitesi kategorilerinde anlamlı etkiler göstermektedir. Ancak, katsayıların istatistiksel anlamlılıklarını doğrulamak için ek analizlere (örneğin, Wald testi veya p-değer analizi) ihtiyaç duyulabilir.

Bu sonuçlar, hava kalitesini etkileyen faktörlerin belirlenmesi ve olası müdahale stratejilerinin geliştirilmesi açısından önemli çıkarımlar sunmaktadır. Modelin AIC ve kalan sapma değerlerinin daha düşük olması, gelecekteki model geliştirme çalışmalarında dikkate alınması gereken bir alan olabilir.,

### 8.Sonuç ve Öneriler

Bu bölümde regresyon modeli sonucu ve öneriler bulunmaktadır

#### 8.1 Bulguların Yorumu

Bu çalışmada, hava kalitesi kategorilerinin (Good, Hazardous, Poor) belirlenmesinde lojistik regresyon yöntemi kullanılmış ve bağımsız değişkenlerin (örneğin, nem, azot dioksit, kükürt dioksit, vb.) hava kalitesine olan etkileri incelenmiştir. Elde edilen bulgular, özellikle azot dioksit ve nem oranının hava kirliliği kategorilerini anlamlı şekilde etkilediğini göstermektedir. Model sonuçları, hava kalitesinin değerlendirilmesinde lojistik regresyonun güvenilir bir araç olduğunu ortaya koymuştur. Hava kalitesi kategorilerine göre artık analizleri, modelin iyi uyum sağladığını ve verilerin büyük ölçüde model varsayımlarını karşıladığını göstermiştir. Ancak belirli kategorilerde (örneğin, Poor ve Hazardous) model performansının geliştirilmesi için ek değişkenlerin dahil edilebileceği gözlemlenmiştir.

#### 8.2 Öneriler

**Uygulamalı Öneriler:** Hava kirliliğini azaltmak için özellikle azot dioksit emisyonlarının kontrol altına alınması adına endüstriyel faaliyetlerde çevre dostu teknolojilerin kullanılması teşvik edilmelidir. Şehir planlamasında, nem oranını dengeleyebilecek yeşil alanların artırılması gibi hava kalitesine katkı sağlayacak düzenlemeler yapılmalıdır.

**Bilimsel Öneriler:** Modelin doğruluğunu artırmak için daha geniş bir veri seti kullanılabilir ve mevsimsel etkiler modele dahil edilebilir. Çalışma, diğer kirlenici faktörlerin (örneğin, partikül madde, karbon monoksit) etkisini de değerlendirecek şekilde genişletilebilir. Gelecekteki çalışmalarda, makine öğrenimi yöntemleri gibi daha karmaşık modelleme teknikleri ile lojistik regresyon sonuçları karşılaştırılarak daha kapsamlı analizler yapılabilir.

Bu sonuçlar, hava kirliliği ile mücadele eden politika yapıcılar ve çevresel düzenlemelerde görevli kurumlar için yol gösterici olabilir. Çalışma, hava kalitesinin sürdürülebilir bir şekilde iyileştirilmesi için önemli bilgiler sunmaktadır.

## 8.Kaynakça

Mustafayeva, U. (Tarih). SNP markırlarının burun morfolojisi ile ilgili özelliklerinin modellenmesi / Modelling of SNP markers' features related to nose morphology (Yüksek Lisans Tezi). İstanbul Üniversitesi-Cerrahpaşa, Adli Tıp ve Adli Bilimler Enstitüsü, Fen Bilimleri Ana Bilim Dalı.

Kara, M., & Bilgili, M. (2017). Hava kirliliği ile meteorolojik değişkenler arasındaki ilişkilerin istatistiksel analizi: Kayseri örneği. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 23(5), 655-661. <https://doi.org/10.5505/pajes.2017.32043>

Şimşek, N. (2019). *Araştırma önerisi* (Final). İstanbul Ticaret Üniversitesi. Erişim adresi: <https://ww3.ticaret.edu.tr/nsimsek/files/2019/03/Ara%C5%9Ft%C4%B1rma-%C3%96nerisiF%C4%B0NAL.docx>

Esenyel İçen, N. M. (Tarih). *Bayesyen multinominal lojistik regresyon ile işletmelerin finansal başarısızlığının değerlendirilmesi: İmalat sanayi uygulaması* (Yüksek Lisans/Doktora Tezi). İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Ana Bilim Dalı.

Turğut, E. (Tarih). *Multipl skleroz hastalarında nötrofil/lenfosit, platelet/lenfosit, eosinofil/lenfosit ve monosit/lenfosit oranlarının hastalık aktivitesi ile ilişkisi* (Yüksek Lisans/Doktora Tezi). Kırıkkale Üniversitesi, Tıp Fakültesi, Nöroloji Ana Bilim Dalı.