

Lojistik Elastik Net Yönteminin Alternatif Yöntemlerle Karşılaştırılması

Sevim Simge Uysal

**YÜKSEK LİSANS TEZİ**

İstatistik Anabilim Dalı

Kasım 2020

Comparison of The Logistic Elastic Net Method with Alternative Methods

Sevim Simge Uysal

**MASTER OF SCIENCE THESIS**

Department of Statistics

Kasım 2020

Lojistik Elastik Ağ Yönteminin Alternatif Yöntemlerle Karşılaştırılması

Sevim Simge Uysal

Eskişehir Osmangazi Üniversitesi  
Fen Bilimleri Enstitüsü  
Lisansüstü Yönetmeliği Uyarınca  
İstatistik Anabilim Dalı  
İstatistik Teorisi Bilim Dalında  
YÜKSEK LİSANS TEZİ  
Olarak Hazırlanmıştır

Danışman: Prof. Dr. Arzu Altın Yavuz

Kasım 2020

## ETİK BEYAN

Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna göre, Prof. Dr. Arzu ALTIN YAVUZ danışmanlığında hazırlamış olduğum “Lojistik Elastik Net Yönteminin Alternatif Yöntemlerle Karşılaştırılması” başlıklı YÜKSEK LİSANS tezimin özgün bir çalışma olduğunu; tez çalışmamın tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; tezimde verdiğim bilgileri, verileri akademik ve bilimsel etik ilke ve kurallara uygun olarak elde ettiğimi; tez çalışmamda yararlandığım eserlerin tümüne atıf yaptığımı ve kaynak gösterdiğimi ve bilgi, belge ve sonuçları bilimsel etik ilke ve kurallara göre sunduğumu beyan ederim. 04/11/2020

Sevim Simge Uysal

İmza

## ÖZET

İstatistiksel çalışmalarda bağımlı değişken ile bağımsız değişkenler arasındaki nedensellik ilişkisini açıklamak için sıklıkla regresyon analizi kullanılmaktadır. Klasik çoklu doğrusal regresyon analizinde bağımlı değişkenin sürekli ve normal dağılıma sahip olma koşulu vardır. Ancak bazı durumlarda bağımlı değişken kesikli değerler alan bir değişken olabilir. Bu durumda regresyon analizi lojistik regresyon analizi adını alır. Çoklu doğrusal regresyon modelinde olduğu gibi lojistik regresyon modelinde de bağımsız değişkenler arasında yüksek derecede korelasyonların olması çoklu iç ilişki problemini ortaya çıkaracaktır. Bu problemin üstesinden gelmek için yanlı regresyon yöntemleri kullanılabilir. Yanlı regresyon yöntemlerinden olan ridge regresyon, lojistik regresyon modeline uyarlanmasıyla elde edilen ve  $L_2$  kısıtı altında parametreleri tahmin etmeye çalışan lojistik ridge regresyon yöntemi çoklu iç ilişkinin üstesinden gelmek için önerilmiştir. Çoklu iç ilişki probleminin üstesinden gelmenin bir diğer yolu modeldeki bağımsız değişken sayısının azaltılmasıdır. Hem çoklu iç ilişki problemini çözen hem de cezalı regresyon yöntemlerinde biri olan LASSO (En küçük büzülme ve seçim yöntemi)  $L_1$  kısıtı altında parametre tahmini yapmaktadır. Lojistik regresyon modeline uyarlanmasıyla elde edilen lojistik-LASSO boyut indirgeyerek değişken seçimi yapan bir yöntemdir. Bu yöntemde hem değişken seçimi yapılırken aynı zamanda model parametreleri de tahmin edilmektedir. LASSO' nun dezavantajlarını azaltmak için  $L_1$  ve  $L_2$  normu arasında bir denge sağlamaya çalışan Elastik ağ yöntemi lojistik regresyon modeline uyarlanarak elde edilen Lojistik Elastik ağ yöntemi, bağımsız değişkenler arasında ikili korelasyonlar yüksek olsa bile etkin sonuçlar sağlamaktadır. Regresyon analizindeki bir diğer sorun ise aykırı değer problemidir. Lojistik regresyonda veri setinde aykırı değerlerin olması durumunda EÇO tahmin edicisinin etkinliği azalmaktadır. Bu durumda veri setindeki aykırı değerlerin etkilerini azaltmak için sağlam yöntemler kullanılmaktadır.

Bu tez çalışmasında sağlam lojistik regresyon yöntemi kullanılmıştır. Bu tez çalışmasında Sakarya Nehri üzerinden seçilmiş sekiz istasyona ait su verileri lojistik regresyon, lojistik ridge, lojistik-LASSO, lojistik Elastik ağ ve sağlam lojistik regresyon yöntemleri kullanılarak iki istasyon arasındaki değişime neden olan su kalite karakteristikleri belirlenmeye çalışılmıştır.

**Anahtar Kelimeler:** Çoklu İç İlişki, Lojistik Regresyon, Lojistik Ridge Regresyon, Lojistik-LASSO, Lojistik-Elastik Ağ, Sağlam Lojistik Regresyon.

## SUMMARY

Regression analysis is used to explain the causality relationship between the dependent variable and the independent variables in statistical studies. In classical multiple linear regression analysis, the dependent variable must have a continuous and normal distribution. However, in some cases, the dependent variable may take discrete values. In this case, logistic regression analysis is widely used. In logistic regression analysis, the problem of multicollinearity can also be encountered. Penalized regression methods can be used to overcome this problem. One of the penalized regression method ridge regression is a method that tries to estimate parameters under the  $L_2$  constraint in the logistic regression model and is called the logistic ridge regression model. LASSO (Least absolute shrinkage and selection operator), which is another penalized regression method, makes parameter estimation under the  $L_1$  constraint. Logistics-LASSO, obtained by adapting it to the logistic regression model, is a method that performs variable selection and parameter estimation simultaneously. In order to reduce the disadvantages of Ridge and LASSO, the Elastic net method, which tries to provide a balance between  $L_1$  and  $L_2$  norms, has also been adapted to the logistic regression model. Logistics Elastic net method provides effective results in case of multicollinearity. In addition, the presence of outliers in the data set in the logistic regression analysis reduces the effectiveness of the MLE estimator. In this case, robust methods should be used to reduce the effects of outliers in the data set.

In this thesis, robust logistic regression method is discussed. These methods discussed in this thesis study were compared using the water quality data of eight stations selected over Sakarya River. The water quality characteristics causing the change between the two stations were determined and it was determined that the Logistic Elastic Net method gave the most efficient classification.

**Keywords:** Multicollinearity, Logistic Regression, Logistic-Ridge Regression, Logistic-LASSO, Logistic-Elastic Net, Robust Logistic Regression.

## İÇİNDEKİLER

	<u>Sayfa</u>
<b>ÖZET .....</b>	<b>vi</b>
<b>SUMMARY .....</b>	<b>vii</b>
<b>TEŞEKKÜR .....</b>	<b>viii</b>
<b>İÇİNDEKİLER .....</b>	<b>ix</b>
<b>ŞEKİLLER DİZİNİ .....</b>	<b>xi</b>
<b>ÇİZELGELER DİZİNİ .....</b>	<b>xii</b>
<b>SİMGELER ve KISALTMALAR DİZİNİ .....</b>	<b>xiii</b>
<b>1. GİRİŞ .....</b>	<b>1</b>
<b>2. LİTERATÜR ARAŞTIRMASI .....</b>	<b>5</b>
<b>3. YÖNTEM .....</b>	<b>11</b>
<b>3.1. Çoklu Doğrusal Regresyon Modeli .....</b>	<b>11</b>
<b>3.2. İkili Lojistik Regresyon Modeli .....</b>	<b>12</b>
<b>3.3. Cezalı Regresyon.....</b>	<b>17</b>
<b>3.4. Çoklu İç İlişki .....</b>	<b>18</b>
<b>3.4.1. Çoklu İç İlişkinin Nedenleri .....</b>	<b>18</b>
<b>3.4.2. Çoklu İç İlişkinin Belirlenmesi .....</b>	<b>18</b>
<b>3.5. Ridge Regresyon .....</b>	<b>19</b>
<b>3.6. Lojistik Ridge Regresyon .....</b>	<b>20</b>
<b>3.7. En Küçük Açık Regresyonu (LARS) .....</b>	<b>22</b>
<b>3.7.1. LASSO .....</b>	<b>25</b>
<b>3.7.1.1. LASSO'nun Grafiksel Yorumu .....</b>	<b>27</b>
<b>3.7.2. Lojistik LASSO.....</b>	<b>28</b>
<b>3.8. Elastik Ağ.....</b>	<b>29</b>
<b>3.9. Lojistik-Elastik Ağ .....</b>	<b>32</b>
<b>3.10. Sağlam Lojistik Regresyon .....</b>	<b>34</b>

**İÇİNDEKİLER (devam)**

	<b><u>Sayfa</u></b>
3.11. Model Seçimi.....	35
3.11.1. Ayar Parametrelerinin Seçimi.....	35
3.11.2. Çapraz Geçerlilik Testi (Cross Validation).....	35
3.12. İstatistiksel Çıkarılma.....	41
<b>4. UYGULAMA.....</b>	<b>41</b>
4.1. Su Parametreleri Hakkında Genel Bilgiler.....	42
4.1.1. Sıcaklık .....	42
4.1.2. pH.....	43
4.1.3. Çözünmüş Oksijen .....	43
4.1.4. Elektriksel İletkenlik .....	43
4.1.5. Sülfat İyonu .....	44
4.1.6. Amonyum Azotu .....	44
4.1.7. Biyolojik Oksijen İhtiyacı .....	44
4.1.8. Magnezyum .....	44
4.1.9. Klor .....	45
4.1.10. Sodyum .....	45
4.1.11. Toplam Sertlik ve Toplam Alkalinite .....	45
4.1.12. İletkenlik ve Tuzluluk .....	46
4.1.13 Toplam Çözünmüş Madde .....	46
4.2. Çalışma Alanı.....	46
<b>5. BULGULAR VE TARTIŞMA.....</b>	<b>66</b>
<b>6. SONUÇ VE ÖNERİLER.....</b>	<b>100</b>
<b>KAYNAKLAR DİZİNİ.....</b>	<b>102</b>



## ŞEKİLLER DİZİNİ

<u>Sekil</u>	<u>Sayfa</u>
3.1. Ridge ( $L_2$ cezalı) Lojistik Regresyon .....	22
3.2. $l$ -q normu için farklı $q$ değerlerinin Grafiksel Gösterimi.....	25
3.3. LARS ve LASSO Profili .....	26
3.4. EKK, Ridge ve LASSO'nun Grafiksel Yorumu .....	27
3.5. LASSO (sağda) ve Ridge (solda) İki Boyutlu Tahmin Durumu .....	27
3.6. LASSO ( $L_1$ ) cezalı Lojistik Regresyon .....	28
3.7. Artıkların Konturleri ve Elastik Ağ Tahmin Edicisinin Sınırlama Bölgesi.....	30
3.8. LASSO ve Elastik ağ Yöntemlerinin $\alpha$ Parametresine Göre Davranışları .....	31
3.9. Ridge, LASSO ve Elastik Ağ yöntemlerinin Grafiksel Gösterimi.....	32
3.10. c-Kat Çapraz Geçerlilik Örneği .....	37
3.11. Çapraz Geçerlilik Hata Eğrisi.....	38
4.1. Akmeşe Deresi ve İnegöl Yenicedere Mansap İstasyonlarının Konumları .....	48
4.2. Dere Mahallesi Köprü, Sarıyar Baraj Çıkışı ve Kayaboğazı Baraj Çıkışı İst. Kon .....	49
4.3. Çubuk 1 Baraj Çıkışı, Kuruçay Baraj Aksı ve Bayındır Barajı Çıkışı İst. Kon. ....	50
4.4. Birinci Karşılaştırma için Korelasyon Grafiği.....	52
4.5.İkinci Karşılaştırma için Korelasyon Grafiği.....	54
4.6. Üçüncü Karşılaştırma için Korelasyon Grafiği.....	56
4.7. Dördüncü Karşılaştırma için Korelasyon Grafiği.....	59
4.8. Beşinci Karşılaştırma için Korelasyon Grafiği.....	62
4.9. Altıncı Karşılaştırma için Korelasyon Grafiği.....	63
4.10. Yedinci Karşılaştırma için Korelasyon Grafiği.....	65

## ÇİZELGELER DİZİNİ

<u>Cizelge</u>	<u>Sayfa</u>
3.1. Bernoulli Olasılık Dağılımı.....	13
4.1. Birinci Karşılaştırma için Korelasyon Matrisi.....	51
4.2. İkinci Karşılaştırma için Korelasyon Matrisi.....	54
4.3. Üçüncü Karşılaştırma için Korelasyon Matrisi.....	55
4.4. Dördüncü Karşılaştırma için Korelasyon Matrisi.....	57
4.5. Beşinci Karşılaştırma için Korelasyon Matrisi.....	60
4.6. Altıncı Karşılaştırma için Korelasyon Matrisi.....	61
4.7. Yedinci Karşılaştırma için Korelasyon Matrisi.....	64
5.1. Akmeşe Deresi ve İnegöl Yenicedere Mansap İstasyonlarına ait Analiz Sonuçları.....	67
5.2. Sarıyar Baraj Çıkışı ve Dere Mahallesi Köprü İstasyonuna ait Analiz Sonuçları.....	71
5.3. Kayaboğazı Baraj Çıkışı ile Dere Mahallesi Köprü İstasyonları için An. Sonuçları....	77
5.4. Sarıyar Baraj Çıkışı ve Kayaboğazı Baraj Çıkışı Analiz Sonuçları.....	81
5.5. Çubuk-1 Baraj Çıkışı ve Kuruçay Baraj Aksı'na ait Analiz Sonuçları.....	86
5.6. Çubuk-1 Baraj Çıkışı ve Bayındır Barajı Çıkışına ait Analiz Sonuçları.....	91
5.7. Kuruçay Baraj Aksı ile Bayındır Barajı Çıkışı İst. Analiz Sonuçları.....	96
6.1. İstatistiksel Olarak Anlamlı Olan Değişkenler.....	101

## SİMGELER VE KISALTMALAR DİZİNİ

### Kısaltmalar

EKK

EÇO

LARS

LASSO

MSE

CV

DSİ

BOD<sub>5</sub>Ca<sup>++</sup>Cl<sup>-</sup>

DO

EC

K<sup>+</sup>Mg<sup>+2</sup>Na<sup>+</sup>NH<sub>4</sub>NO-PO<sub>4</sub>

Ph

Pv

SO<sub>4</sub>

T

TDS

TH

### Açıklamalar

En Küçük Kareler

En Çok Olabilirlik

En Küçük Açık Regresyonu

En Az Mutlak Büzülme ve Seçim Operatörü

Artık Kareler Ortalaması

Çapraz Geçerlilik Testi

Devlet Su İşleri

Biyolojik Oksijen İhtiyacı

Kalsiyum

Klor

Çözünmüş Oksijen

Elektriksel İletkenlik

Potasyum

Magnezyum

Sodyum

Amonyum Azotu

Orto Fosfat

Alkalinite

Permanganat Değeri

Sülfat

Sıcaklık

Toplam Çözünmüş Katılar

Toplam Sertlik

## 1.GİRİŞ

Regresyon analizi uygulamalı çalışmalarda sıklıkla kullanılan istatistiksel bir yöntemdir. Regresyon analizi bağımlı değişken ile bağımsız değişkenler arasındaki nedensellik ilişkisini açıklamak ve regresyon modelindeki bilinmeyen parametreleri tahmin etmek amacıyla kullanılmaktadır (Erar, 2013). Regresyon analizinde ele alınan bağımlı ve bağımsız değişkenlerin sayısı ve ölçüm düzeyleri çeşitli regresyon modellerinin ortaya çıkmasına neden olmaktadır. Ölçüm düzeyi sürekli olan tek bağımlı ve tek bağımsız değişken durumunda regresyon modeli basit doğrusal regresyon modeli adını almaktadır. Eğer ele alınan sürekli bir bağımlı değişkene karşı birden fazla bağımsız değişken söz konusu ise çoklu doğrusal regresyon modeli olarak adlandırılmaktadır (Akdi, 2014).

Çoklu doğrusal regresyon modelinde parametre tahmini için yaygın olarak En Küçük Kareler (EKK) yöntemi kullanılmaktadır. EKK yöntemi artık kareler toplamını minimize etmeye dayalı olarak tahmin sonuçları veren bir yöntemdir. Ancak EKK tahmin edicilerinin etkin sonuçlar vermesi için, çoklu doğrusal regresyon modelinin dayandığı varsayımların sağlanması gerekmektedir. Bu varsayımlardan ilki bağımlı değişkenin sürekli değerler alan ve normal dağılıma sahip olan bir değişken olmasıdır. Hata terimlerinin birbirleriyle ilişkisiz, sabit varyanslı normal dağılıma sahip olması gerekmektedir. Ayrıca, bağımsız değişkenler arasında yüksek derecede ilişki olmaması gerekmektedir. Bağımsız değişkenler ile hata terimi arasında ilişki olmaması ve hata teriminin otokorelasyonlu olmaması diğer bir varsayımdır. Çoklu doğrusal regresyon modelinde sıklıkla karşılaşılan bir problem, bağımsız değişkenler arasında yüksek derecede korelasyon olmasıdır. Bağımsız değişkenler arasında yüksek derecede korelasyon olması durumu çoklu iç ilişki problemi olarak adlandırılmaktadır. (Montgomery ve Peck,2001). Bu durumda EKK tahmin edicilerinin etkinliği azalmaktadır (Chang vd., 2018). Ayrıca regresyon analizinde bazı gözlemler veri setindeki diğer gözlemlerden önemli derecede farklı olabilir. Bu gözlemler aykırı değer olarak adlandırılmaktadır. Aykırı değerlerin varlığı da EKK tahmin edicilerinin etkinliğini azaltan diğer bir problemdir.

Bazı durumlarda bağımlı değişkenin sürekli olması mümkün olmayabilir. Bağımlı değişkenin ölçüm düzeyi sınıflayıcı ya da sıralayıcı olabilmektedir. Bir başka ifadeyle

bağımlı değişkenin ikili veya daha fazla kategorik değer aldığı durumlarda lojistik regresyon kullanılabilir. Doğrusal regresyon analizinde olduğu gibi lojistik regresyon analizinde de bağımlı değişken ile bağımsız değişkenler arasındaki nedensellik ilişkisi tespit edilmeye çalışılır. Bir başka ifade ile bağımlı değişken üzerinde etkili olan değişkenlerin tespiti amaçlanır. Doğrusal regresyon modelinde, bağımlı değişkenin değeri tahmin edilirken, lojistik regresyon analizinde bağımlı değişkenin alacağı değerlerden birinin gerçekleşme olasılığı tahmin edilmektedir (Bircan, 2004).

Çoklu iç ilişki problemi varlığında, geleneksel olarak kullanılan tahmin edicilerin varyansı yüksek çıkmakta ve buna bağlı olarak parametre tahminleri gerçeği yansıtmamaktadır. Bu durumda klasik tahmin edicilerinin etkinliğini azalmaktadır. Çoklu iç ilişki probleminin üstesinden gelmek için örneklem hacminin arttırılması ya da yüksek korelasyonlu değişkenlerden bazılarının çalışmadan çıkarılması gerekmektedir. Değişken seçimi uygulanması durumunda hangi değişkenin seçilmesi gerektiği konusu önem kazanmaktadır. Örneklem hacminin arttırılması durumunda ise bazı çalışmalar için etkin sonuçlara ulaşılamamaktadır. Çünkü incelenen çalışmada ele alınan değişkenlerin yapısı gereği yüksek korelasyon giderilememektedir. Çoklu iç ilişki probleminin üstesinden gelmek için literatürde çeşitli yöntemler önerilmiştir. Hoerl ve Kennard (1970) tarafından önerilen ridge regresyon çoklu iç ilişki problemi ile karşılaşıldığında sıklıkla kullanılan yanlı regresyon yöntemlerinden birisidir. Ridge regresyon hem çoklu doğrusal regresyon analizinde hemde lojistik regresyon analizinde kullanılmaktadır. Lojistik regresyon analizinde bu yöntem, lojistik-Ridge yöntemi olarak adlandırılmaktadır. Lojistik ridge regresyon modelinde,  $L_2$  kısıtı altında bağımsız değişkenlerin tamamı için bir katsayı tahmini söz konusudur. Dolayısıyla lojistik ridge yönteminde değişken seçimi yapılmaksızın model kurulur. Ancak bu durum elde edilen sonuçların kolay yorumlanamaması ve verilerdeki küçük değişikliklerin tahmin doğruluğunu azaltması gibi sonuçlara neden olur (Hastie vd., 2015).

Çoklu iç ilişki probleminin üstesinden gelmek için kullanılan diğer bir yöntem olan değişken seçimi yöntemi, son dönemlerde makine öğrenme algoritmalarının geliştirilmesinde yaygın olarak kullanılmaktadır. Tibshirani (1996), cezalı bir regresyon yöntemi olan ve değişken seçimi ile parametre tahminini aynı anda gerçekleştiren En Küçük Mutlak Büzülme ve Seçim Operatörü (Least Absolute Shrinkage and Selection Operatör,

LASSO) yöntemini önermiştir. LASSO, çoklu iç ilişki varlığında kullanılan cezalı regresyon yöntemlerinden biridir. LASSO regresyonu katsayıların mutlak değerinin toplamının belirli bir sabitten daha az olması kısıtı yani  $L_1$  normu altında artık kareler toplamının minimizasyonuna dayanmaktadır (Tibshirani, 1996). Ancak, LASSO yönteminin de bazı dezavantajları bulunmaktadır. LASSO'nun dezavantajlarından biri bağımsız değişkenler arasındaki çok yüksek korelasyona karşı sağlam olmamasıdır. Bu durumda yüksek derecede korelasyonlu değişken grubundan bir değişkeni rastgele seçip geri kalanını gözardı etmektedir. LASSO'nun bir başka dezavantajı ise,  $p > n$  olduğunda yüksek boyutlu verilerde, en fazla  $n$  bağımsız değişken seçmesidir (Zeny, 2012; Zou ve Hastie, 2005; Zhou, 2013). Bu iki dezavantajının üstesinden gelmek için LASSO regresyona ikinci bir kısıt eklenerek elastik ağ yöntemi önerilmiştir (Zou ve Hastie, 2005).

Elastik ağ (Elastic-Net) hem  $L_1$  hem de  $L_2$  ceza terimlerinin birleştirilmesi sonucu ortaya çıkmıştır. Lojistik regresyon modelinde Elastik ağ yöntemi LASSO yönteminde olduğu gibi negatif-log olabilirlik fonksiyonuna iki kısıt eklenerek elde edilir. Çok boyutlu ikili sınıflandırma verilerinde lojistik-elastik ağ regresyonu kullanılarak değişken azaltılması yoluna başvurulmaktadır. Hem katsayıları tahmin etmek hem de değişken seçimini aynı anda gerçekleştirmek için, elastik ağ cezası yüksek boyutlu ikili sınıflandırmada başarılı sonuçlar vermektedir (Algarnal ve Lee 2015).

Regresyon analizinde karşılaşılan bir diğer problem aykırı değer sorunudur. Aykırı değerlerin varlığında kurulan lojistik regresyonda her bir gözlemin ayrıntılı olarak incelenmesi ve aykırı değerlerin belirlenmesi gerekir. Lojistik regresyon analizinde aykırı değerler açıklayıcı değişkenlerde görülmektedir. Bu değerler kurulan model yardımıyla yapılacak sınıflandırmada yanlış sınıflandırma problemine neden olmaktadır. Veri setinde aykırı değerler bulunması durumunda bu gözlemlerin silinmesi en ilkel yöntemdir. Gözlem kaybı elde edilecek tahminlerin ve modelin geçerliliğini etkilemektedir. Bunun yerine, aykırı değerlerin etkisini azaltarak veri kaybını engelleyen sağlam yöntemler kullanılabilir. Lojistik regresyon modelinde aykırı değer sorununun üstesinden gelmek için geleneksel olarak kullanılan En Çok Olabilirlik tahmin edicisi yerine, Bianco ve Yohai (1996) tarafından geliştirilen ve bir  $M$  sınıfı tahmin edici olan Bianco-Yohai tahmin edicisi kullanılmaktadır.

Bu tez çalışmasında bağımsız değişkenler arasında yüksek derecede korelasyon olması durumunda lojistik regresyona alternatif olarak lojistik-Ridge, lojistik-LASSO ve lojistik-elastik ağ yöntemleri tanıtılmıştır. Ayrıca, aykırı değer probleminin etkisini azaltmak için sağlam lojistik regresyon yöntemi tanıtılmıştır. Çalışmanın üçüncü bölümünde bu yöntemler hakkında teorik bilgiler verilmiştir. Uygulama bölümü olan dördüncü bölümde ise öncelikle su kalitesi verilerine ilişkin açıklamalara yer verilmiştir. Çalışma alanı olarak Sakarya Nehri'nden seçilmiş 8 adet gözlem istasyonu için 2014-2018 yılları arasındaki su kalitesi ölçümleri elde edilmiştir. Daha sonra ele alınan yöntemler kullanılarak su kalitesinin istasyonlar arasında farklılığına neden olan su kalite karakteristikleri belirlenmeye çalışılmıştır.

## 2.LİTERATÜR ARAŞTIRMASI

Lojistik regresyon analizi, son yıllarda tıp, ekonomi, ziraat, spor, biyoloji alanındaki uygulamalarda sıklıkla kullanılmaktadır. Lojistik regresyon analizinde amaç, diğer regresyon modellerinde olduğu gibi az sayıda değişken kullanarak bağımlı ve bağımsız değişkenler arasındaki ilişkiyi tanımlayabilen, kabul edilebilir bir model kurmaktır. Ayrıca, lojistik regresyon analizi normallik varsayımının sağlanamaması durumunda diskriminant analizinin alternatifi olarak sınıflandırma problemlerine çözüm olarak kullanılmaktadır.

Berkson (1944), lojistik regresyon modelini ilk olarak biyolojik deneylerin analizi için kullanmıştır. Cox (1970), bu modeli gözden geçirerek çeşitli uygulamalar yapmış, özet gelişmeler ise ilk olarak Andersson (1979, 1983) tarafından verilmiştir. Ayrıca verilerin lojistik modele uyumu ile ilgili birçok çalışma yapılmıştır. Bunlar arasında Aranda ve Ordaz (1981) ve Johnson (1985) tarafından yapılan çalışmalar en önemlileridir. Pregibon (1981) iki grup lojistik regresyon modelinde etkili (influential), aykırı (outlier) gözlemleri ve belirleme ölçütlerini (diagnostic) incelemiştir. Lee (1984), basit dönüşümlü deneme planları için lojistik modeller üzerinde çalışmıştır. Schaeffer (1986), çoklu iç ilişki varlığında lojistik regresyon modeline alternatif olarak Stein, Temel Bileşenler ve Ridge tahmincilerini önermiştir. Lesaffre (1986), Lesaffre ve Albert (1989) ise çoklu grup lojistik modellerde etkin ve aykırı gözlemleri belirleme ölçütlerini incelemiştir. Bonney (1987) ve Duffy (1990), lojistik regresyonda hata terimlerinin dağılımını ve parametre tahminlerinin gerçek değerlere yakınsamasını incelemiştir. Arı ve Önder (2013), yaptıkları çalışmada lojistik regresyon ve diğer farklı regresyon modellerini karşılaştırarak bu regresyon modellerinin hangi durumlarda kullanılabileceğini incelemiştir. Chang (2015), lojistik regresyon modelinde çoklu iç ilişki probleminin üstesinden gelmek için lojistik Ridge ve lojistik Lui tahmin edicilerini kullanmıştır. Arıcan (2016), lojistik regresyon modelinde çoklu iç ilişki probleminin etkilerini azaltmak için en çok olabilirlik tahmin edicisine (EÇO) alternatif olarak yeni parametre tahmin yöntemleri önermiştir.

Lojistik regresyon analizinin teorisi kadar uygulama çalışmaları da ilgi çekicidir. Gardside ve Glueck (1995) insanlarda beslenme şekli, sigara ve alkol kullanımı, fiziksel aktivite gibi risk faktörlerinin kalp hastalığı üzerindeki etkilerini incelemiştir. Kloiber vd.



(1996), Peoples vd. (1991), Buescher vd. (1993) kadınlarda düşük doğum ağırlığını etkileyen risk faktörlerini; Santos vd. (1998) kafein tüketimi ve düşük doğum ağırlığı arasındaki ilişkiyi; Sable ve Herman (1997) erken doğum ve düşük doğum ağırlığı arasındaki ilişkiyi incelemişlerdir. O'Neil ve Barry (1995), trafik kazalarına ait verileri ele alarak yaş ve cinsiyet gibi bağımsız değişkenlerin kazalarda ölüm sebebi olabileceğini araştırmışlardır. Bayram (2004), ikili ve çoklu lojistik regresyon analizi yardımıyla ücretsiz gelir karşılığı çalışanların iş gücünü açıklamaya çalışmıştır. Bircan (2004), çocuklarda doğum ağırlıklarını etkileyen önemli risk faktörlerini belirlemek için tıp verileri kullanarak bu verilere lojistik regresyon analizi uygulamıştır. Coşkun vd., (2004), diş hekimliği verileri kullanarak protez kullanımını etkileyen önemli risk faktörlerini belirlemek için lojistik regresyon analizi kullanmışlardır. Akın ve Şentürk (2012), Avrupa yaşam kalitesi anketi aracılığıyla ordinal lojistik regresyon analizi kullanılarak sosyo-demografik özelliklere göre mutluluk düzeylerini incelenmiş ve elde edilen sonuçları daha önce yapılan çalışmalar ile karşılaştırmıştır. Akyol vd. (2020), Parkinson hastalığının teşhisi için en iyi modeli bulmak için lojistik regresyon ve rassal orman algoritmaları kullanılmışlardır.

Lojistik regresyon analizi uygulamalarında veri setinde aykırı değer bulunması durumu tahmin edicilerin etkinliği üzerinde olumsuz etkiye sahiptir. Lojistik regresyonda, açıklayıcı değişkenlerde bulunan ve verinin kalan kısmından oldukça farklı yapıya sahip olan gözlemler aykırı değer olarak adlandırılmaktadır. Aykırı değerlerin varlığı durumunda lojistik regresyonda birimlerin yanlış gruplara atanma sorunu yaşanmaktadır. Bir başka ifade ile ayırıcı fonksiyonun etkinliği azalmaktadır. Bu problemin üstesinden gelmek için çeşitli sağlam tahmin ediciler kullanılarak sağlam lojistik regresyon analizi önerilmiştir. Bu yöntem ilk kez Bianco ve Yohai (1996) tarafından öne sürülmüştür. Croux ve Haesbroeck (2003) ise Bianco ve Yohai'nin hazırladığı algoritmayı modifiye ederek diğer sağlam lojistik regresyon algoritmalarına göre daha hızlı bir algoritma geliştirmişlerdir. Bilgin (2012), ikili yapıda bağımlı değişken içeren heterojen veri setlerinin analizlerinde genel lojistik regresyon, sağlam lojistik regresyon yöntemlerinin performanslarını karşılaştırmıştır. Ergül ve Yavuz (2016), Eskişehir ili otomobil tercihlerinden duyulan memnuniyetin performanslarını değerlendirmek amacıyla lojistik regresyon, diskriminant analizi ve sınıflandırma ağaçları analizi yöntemleri kullanılmışlardır. Ergül vd. (2016), Avrupa bayanlar şampiyonlar ligi takımlarının performanslarını değerlendirmek amacıyla lojistik regresyon, sağlam lojistik regresyon ve lojistik ridge yöntemleri kullanarak analiz edilmiştir. Karacasu vd. (2016),

Eskişehir ilindeki trafik kazalarını etkileyen faktörleri belirlemek amacıyla lojistik regresyon ve diskriminant analizi yöntemlerini kullanmışlardır. Aşık ve Yavuz (2018), yaşam kalitesini etkilediği düşünülen ve yaşam kalitesini oluşturan diğer alt endeks değerleri yardımıyla bir sınıflandırma mekanizması oluşturmuş, bu sınıflandırma problemi için yaşam memnuniyetini etkileyen değişkenlerin istatistiksel olarak anlamlı olup olmadığı belirlenmeye çalışmıştır. Bunun için diskriminant analizi, sağlam diskriminant analizi, lojistik regresyon, sağlam lojistik regresyon, lojistik ridge ve sağlam lojistik ridge regresyon yöntemlerini kullanmışlardır. Yavuz ve Çilengiroğlu (2020), TÜİK 2017 dönemine ait yaşam memnuniyeti verileri kullanılarak CART algoritması ve lojistik regresyon yöntemlerinin performanslarını değerlendirmişlerdir.

Lojistik regresyon analizi uygulanırken karşılaşılan bir diğer problem çoklu ilişki sorunudur. Bu sorunun üstesinden gelmek için ise genellikle yanlı regresyon yöntemleri kullanılmaktadır. Yanlı regresyon yöntemleri içerisinde en çok kullanılan yöntem ise Ridge regresyon yöntemidir. İlk kez Hoerl ve Kennard (1970) tarafından önerilen Ridge tahmin edicisi EKK yöntemine alternatif bir çözüm olarak sunulmuştur. Ridge regresyon, EKK tahmin edicilerinin yansızlık özelliğinden ödün vererek varyanslarını minimum yapmaya çalışmaktadır. Bağımlı değişkenin iki düzeyli olduğu durumlarda ise EÇO tahmin edicine alternatif olarak, lojistik-Ridge tahmin edicisi kullanılmaktadır. Lojistik Ridge tahmin edicisi ilk olarak Schaefer vd. (1984) tarafından verilmiştir. Dufy ve Santer (1989) lojistik regresyon modeli için lojistik ridge regresyon tahmin edicisini önermişler ve bu tahmin edicinin Schaefer vd. (1984) tarafından sunulan lojistik ridge tahmin edicisi ile asimtotik olarak denk olduğunu göstermişlerdir.

Le Cessie ve Van Houwelingen (1992), lojistik ridge tahmin edicisini kanserli DNA hücreleri ile ilgili veri kümesinin analizinde kullanmışlardır. Barker ve Brown (2001) lojistik regresyonda çoklu iç ilişkinin varlığı durumunda, EÇO tahmin edicisini, lojistik ridge ve temel bileşenler tahmin edicilerini bir simülasyon çalışması yardımıyla karşılaştırmışlardır. Aguilera vd. (2006), yüksek derecede çoklu iç ilişkinin olması durumunda lojistik regresyon analizinde temel bileşenler yönteminin kullanılması incelemişlerdir. Simülasyon çalışması ile farklı optimum temel bileşen seçimi metotları karşılaştırmalı olarak değerlendirilmiştir. Vágó ve Kemény (2006), sürekli ve kesikli bağımsız değişkenlerin bulunduğu bir biyomedikal veri kümesine lojistik ridge regresyon analizi uygulamıştır. Muniz ve Kibria

(2009), ridge parametresini belirlemek için kullanılan yöntemleri derleyerek yeni bir ridge tahmin edicisi önermişlerdir. Önerilen bu yeni ridge tahmin edicisini bir simülasyon çalışması literatürde varolanlar ile karşılaştırmışlardır. Kibria vd., (2011), ridge parametresini tahmin etmek için önerilen yöntemleri genelleştirerek Monte Carlo simülasyonu yardımıyla karşılaştırmışlardır. Mansson ve Shukur (2013), lojistik ridge regresyonda ridge parametresini belirlemek için yeni yaklaşımlar önermişlerdir. Bir simülasyon çalışmasıyla yöntemler artık kareler ortalamasına (MSE) göre değerlendirilmiştir. Ertuğrul (2018), farklı tuzluluğa sahip göllerde popülasyonun büyüklüğünün alansal ve zamansal olarak modellenmesi için lojistik regresyon ve lojistik ridge regresyon yöntemlerini kullanmışlardır.

Çoklu iç ilişki probleminin çözümünde kullanılan ve cezalı regresyon yöntemlerinden biri olan LASSO, son zamanlarda sıklıkla karşımıza çıkmaktadır. İlk kez Tibshirani (1996) değişken sayısını en aza indirerek eş anlolu olarak parametre tahmini ve değişken seçimi yapan LASSO yöntemini tanıtmıştır. LASSO katsayıların mutlak değerinin toplamının belirli bir sabitten daha az olması kısıtı altında artık kareler toplamının minimizasyonuna dayanmaktadır (Montgomery ve Peck, 2001). Melkumova ve Shatskikh (2017), kırmızı şarap verisini kullanarak EKK, Ridge ve LASSO'nun performansları değerlendirilmiştir. LASSO ilk zamanlarda sadece çoklu doğrusal regresyonda EKK tahmin yöntemine alternatif olarak ortaya çıkmasına rağmen, zaman içerisinde yapılan çalışmalar sonucunda uygulama alanları genişletilmiştir (Zou ve Hastie, 2005). Bağımlı değişkenin kategorik olduğu lojistik regresyon için LASSO tahmin edicileri negatif log-olabilirlik fonksiyonuna ceza terimi eklenerek elde edilmektedir ve lojistik LASSO olarak adlandırılmaktadır (Tibshirani, 1996). Bu yöntem ile ilgili literatürde yapılan çalışmalar incelendiğinde Wu vd. (2009), gen seçimi için lojistik-LASSO yöntemini kullanmışlardır. Bielza vd. (2011), kanseri sınıflandırmak amacıyla lojistik LASSO ve lojistik Ridge yöntemleri kullanmışlardır. Sabbe vd. (2012), eksik veri varlığında lojistik LASSO yöntemini 273 akciğer kanseri hastasının verilerine uygulamışlardır. Pereira vd. (2015), bir şirketin iflasının tahmini için lojistik-LASSO ve lojistik ridge regresyonu kullanmışlardır. Brooks vd. (2016), 2010-2013 yaz plaj sezonunda Wisconsin Great Lakes plajındaki dışkı göstergesi (FIB) konsantrasyonunu 14 regresyon yöntemi ile değerlendirmişlerdir. Bu çalışmada kısmi en küçük kareler, seyrek kısmi en küçük kareler, çoklu doğrusal regresyon ve ikili lojistik regresyon modeli kullanılmıştır. Liu vd. (2016), Çin'in Shenzhen şehrindeki

2011-2014 yılları dönemine ait su kalitesi verisini kullanılarak Grup-LASSO uygulamışlardır. Yaman (2017), çok değişkenli Bernoulli lojistik modellerde LASSO tahmincilerini incelemiştir. Kim vd. (2018), meme kanserini tahmin etmek amacıyla görüntü analizinin performansını değerlendirmek için aşamalı lojistik (stepwise logistic) ve lojistik LASSO regresyonu kullanmışlardır. Robles-Guerrero vd. (2019), arı sesini izleyerek kraliçesiz durumları belirlemek için çok sınıflı bir sınıflandırma probleminde, özellik seçimi ve düzenlemesi için lojistik LASSO modelini kullanmışlardır. Garcia-Carretero vd. (2020), D vitamini eksikliğini tahmin etmek için lojistik regresyon, lojistik LASSO ve Elastik Ağ yöntemlerini kullanarak performansları karşılaştırılmıştır.

Yaygın kullanımının yanında LASSO'nun iki dezavantajı vardır. Bunlar, bağımsız değişkenler arasındaki yüksek korelasyona karşı sağlam olmaması ve bu değişkenlerden birini rastgele seçip geri kalanını gözardı etmesidir. Bu dezavantajların üstesinden gelmek için Zou ve Hastie (2005) tarafından,  $L_1$  ve  $L_2$  normu birleştirerek değişken seçimi ve parametre tahmini yapan Elastik ağ yöntemi önerilmiştir. Elastik ağ, değişken seçiminde LASSO cezasından yararlanırken yüksek korelasyon problemi ile başa çıkmak için de ridge regresyon cezasını kullanmaktadır.

Baba vd. (2014), diyabet hastalarının hipertansif olma riskini artıran faktörleri belirlemek için lojistik regresyonda cezayı (Ridge, LASSO ve Elastik Ağ) değiştirerek hastalığın gelişmesine katkıda bulunan faktörleri ve bunun için en uygun yöntemin hangisi olduğunu belirlemeye çalışmışlardır. Zhang vd. (2015), lösemi hastalarına ilişkin veriler ile lojistik LASSO ve lojistik Elastik ağ yöntemlerinin performansları karşılaştırılmıştır. Demir (2017), 146 bebeğe ait verileri kullanarak EKK, Ridge, LASSO ve Elastik Ağ yöntemleri karşılaştırmıştır. Comber ve Harris (2018), 2016 ABD il başkanlık seçimlerinde il düzeyinde oylama modellerini, Trump' a oy vermeyle ilgili sosyo-ekonomik faktörlerin yapısı gibi değişkenlerden oluşan veri seti ile açıklamak için lojistik Elastik Ağ yöntemini uygulamıştır. Göv (2019), Türkiye'de kamu iç borçları ile bazı temel makroekonomik göstergeler arasındaki ilişkiyi ve gelişmekte olan 30 ülkede ekonomik büyüme, enerji tüketimi, dış ticaret dengesi ve finansal gelişme arasındaki ilişkiyi LASSO ve Elastik ağ yöntemleri yardımıyla incelemiştir. Rençber ve Bağcı (2019), sermaye yeterliliğini etkileyen faktörleri belirlemek için elastik ağ yöntemini kullanmışlardır. Tütmez (2020), hava kalitesinin değerlendirilmesi için Ridge, LASSO ve Elastik Ağ yöntemlerini kullanmıştır. Kalkan

(2020), çoklu iç ilişki durumunda Ridge Regresyon, LASSO, Elastik Ağ ve Liu yöntemlerini karşılaştırmıştır.

### 3. YÖNTEM

#### 3.1. Çoklu Doğrusal Regresyon Modeli

Regresyon analizi, bağımlı değişken ile bağımsız değişkenler arasındaki nedensellik ilişkisini açıklamak için kullanılan istatistiksel bir yöntemdir. Bir olguyu açıklayabilmek için tek değişkenin yeterli olması ihtimali oldukça düşüktür. Bu nedenle regresyon analizi çalışmalarında birden çok bağımsız değişkenin olduğu durumlar ile sıklıkla karşılaşmaktadır. Bağımsız değişken sayısının 2 veya daha fazla olduğu ve bağımlı değişkenin sürekli olduğu durumda çoklu doğrusal regresyon modeli kullanılmaktadır (Akdi, 2014). Çoklu doğrusal regresyon modelinin matris notasyonu ile gösterimi Eşitlik (3.1)' de verilmiştir.

$$y = X\beta + \varepsilon \quad (3.1)$$

$$\varepsilon = y - X\beta$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{2n} & & x_{nk} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Eşitlik (3.1) de,

$y$ :  $n \times 1$  boyutlu bağımlı değişken vektörünü,

$X$ :  $n \times p$  boyutlu bağımsız değişkenler matrisini,

$\beta$ :  $k + 1 \times 1$  boyutlu regresyon katsayıları vektörünü,

$\varepsilon$ :  $n \times 1$  boyutlu rasgele hatalar vektörünü göstermektedir.

Doğrusal regresyon modelinde bağımlı değişkenin sabit varyanslı ve normal dağılımlı olması koşulu her zaman sağlanmayabilir. Böyle bir durumda sıklıkla bağımlı değişken için değişken dönüşümü önerilmektedir. Ancak bağımlı değişkenin kategorik olduğu durumlarda değişken dönüşümü doğrusal regresyon modelinin varsayımlarını sağlamada yetersiz kalabilmektedir. Bağımlı değişkenin kategorik olduğu durumlarda çeşitli regresyon modelleri kullanılmaktadır. Bunlar arasında lojistik regresyon modeli uygulamada sıklıkla tercih edilmektedir. Bağımlı değişkenin yapısına bağlı olarak ikili (binary), sıralı

(ordinal), sınıflayıcı (multinomial) lojistik regresyon modelleri bulunmaktadır. Bir regresyon modelinde yanıt değişkeni sadece iki düzeyli kategorik bir değişken ise genellikle ikili lojistik regresyon modeli kullanılmaktadır. İlerleyen kısımda ikili lojistik regresyon modeli açıklanmıştır.

### 3.2. İkili Lojistik Regresyon Modeli

Bağımlı değişkenin kategorik olduğu durumlarda lojistik regresyon modeli yaygın olarak kullanılmaktadır. Çoklu doğrusal regresyon modelinde olduğu gibi lojistik regresyon modelinde de amaç, bağımlı değişken ile bağımsız değişkenler arasındaki nedensellik ilişkisini açıklamaktır. Ayrıca lojistik regresyon bir sınıflandırma yöntemidir. Son zamanlarda makine öğrenme yöntemleri içerisinde yaygın kullanıma sahiptir. Çok değişkenli istatistiksel sınıflandırma problemlerinde veri setinin çok değişkenli normal dağılmaması durumunda da diskriminant analizine alternatif olarak kullanılmaktadır (Algamal ve Lee, 2015).

Lojistik regresyonda bağımlı değişken sadece iki olası sonuç alabiliyorsa ikili lojistik regresyon modeli olarak adlandırılır. Bağımlı değişken 0 ve 1 şeklinde kodlanabilen, başarı ve başarısızlık şeklinde adlandırılan iki düzeye sahiptir. Başarı ya da başarısızlık çalışmanın amacına bağlı olarak araştırmacı tarafından atanmaktadır. Doğrusal regresyon modelinde, bağımlı değişkenin değeri tahmin edilirken, lojistik regresyon analizinde bağımlı değişkenin alacağı değerlerden her birinin gerçekleşme olasılığı tahmin edilmektedir (Bircan 2004).

Lojistik regresyon modelinin sağlaması gereken bazı varsayımları vardır (Aldrich ve Nelson, 1984). Bu varsayımlarından ilki, bağımlı değişkenin alacağı değerlerin karşılıklı zıt olaylar olmasıdır. Bağımlı değişkenin iki düzeyli olması durumunda başarı ve başarısızlık durumları karşılıklı zıt olaylardır ve gerçekleşme olasılıkları toplamı daima 1'dir. Lojistik regresyon analizi, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin doğrusal olduğunu varsaymaz. Lojit dönüşümü sayesinde doğrusallık şartı sağlanmış olur. Ayrıca, çoklu doğrusal regresyonda olduğu gibi, lojistik regresyonda da bağımsız değişkenler arasında yüksek derecede çoklu iç ilişki bulunmamalıdır. Hata terimleri arasında

otokorelasyon olmaması diğerk bir varsayımdır. Lojistik regresyon analizinde gözlem sayısı tahmin edilecek parametre sayısından daima fazla olmalıdır.

İkili lojistik regresyonda, bağımlı değışken iki düzeyli olduğundan Bernoulli dağılımına sahip bir değışkendir. Bağımlı değışkenin alabileceğı değıerlere ilişkin olasılıklar Çizelge (3.1)' de verildiğı gibi ifade edilebilir.

Çizelge 3.1: Bernoulli olasılık dağılımı

$y_i$	Gerçekleşme Olasılığı
0	$P(y_i=0)=1 - \pi_i$
1	$P(y_i=1)=\pi_i$

Çizelge 3.1' de verilen  $y_i$  rassal değışkenin Bernoulli dağılımına sahip olduğundan, lojistik regresyon modeli  $E(\varepsilon_i) = 0$  olmak üzere Eşitlik (3.1)' de verilmiştir.

$$y_i = E(y_i) + \varepsilon_i \quad (3.1)$$

Eşitlik (3.1)' de yer alan  $y_i$ : bernoulli rastgele değışkenini göstermektedir.

Doğrusal modele benzer şekilde, lojistik regresyon modelinde ortalama bağımlı değışkenin bir fonksiyonu olarak ifade edilir. Bu nedenle, lojistik model için dağılımlarının bernoulli olduğu belirtilmektedir (Myers vd., 2012).

Lojistik regresyonda  $E(Y|X = x_i) = \pi(x_i)$  koşullu olasılığı olmak üzere, lojistik regresyon modelinin koşullu olasılığı Eşitlik (3.2)' de verilmiştir.

$$\pi(x_i) = P(Y = 1|x_i) = \frac{e^{(\beta_0 + \beta_1 x_1)}}{1 + e^{(\beta_0 + \beta_1 x_1)}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1)}} \quad (3.2)$$



Lojistik regresyon modelinin beklenen değeri Eşitlik (3.4)'de verilmiştir.

$$E(y_i) = E(x'_i\beta + e_i) = E(x'_i\beta) + E(e_i) = E(x'_i\beta) \quad (3.3)$$

$$E(y_i) = x'_i\beta = \frac{\exp(x'_i\beta)}{1 + \exp(x'_i\beta)} = \pi_i \quad (3.4)$$

Eşitlik (3.4) bağımlı değişken değerinin 1 olduğu durumda bu değeri alma olasılığını göstermektedir.

Denklem (3.1)'deki regresyon modeli ile ilgili bazı temel problemler vardır. İlk olarak eğer bağımlı değişken ikili ise o zaman  $\varepsilon_i$  hata teriminin sadece iki değer alabileceğine dikkat edilmelidir.

$$\begin{aligned} y_i = 1 \text{ iken } \varepsilon_i &= 1 - x'_i\beta \\ y_i = 0 \text{ iken } \varepsilon_i &= -x'_i\beta \end{aligned}$$

Yukarıda görüldüğü gibi hata terimlerinin dağılımı normal dağılmamaktadır. Eşitlik (3.1) için hata varyansı Eşitlik (3.5)'de verilmiştir.

$$\begin{aligned} \sigma_{y_i}^2 &= E[y_i - E(y_i)]^2 \\ &= (1 - \pi_i)^2\pi_i + (0 - \pi_i)^2(1 - \pi_i) \\ &= \pi_i(1 - \pi_i) \end{aligned} \quad (3.5)$$

Elde edilen bu varyans sabit değildir (Erar, 2013).

Lojistik regresyonda parametre tahmini için EÇÖ yöntemi kullanılmaktadır. Bağımlı değişken ikili düzeyli olduğu durumda  $y = 1$  olma olasılığı  $\pi(x_i)$  ve  $y = 0$  olma olasılığı  $1 - \pi(x_i)$  olduğundan, lojistik regresyon modelinin *i*'inci gözlemi için olasılık dağılımı Eşitlik (3.6)'de gösterildiği gibidir.

$$f(y_i) = \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (3.6)$$

Gözlemler bağımsız olduğundan Eşitlik (3.6) yardımıyla olabilirlik fonksiyonu Eşitlik (3.7) biçiminde elde edilir (Hosmer ve Lemeshow,2000).

$$l(\beta_0, \beta_1) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (3.7)$$

Eşitlik (3.7) ifadesini maksimum yapan  $\beta$  değeri en çok olabilirlik tahmin edicileri olarak adlandırılır. Eşitlik (3.7) yardımıyla log-olabilirlik fonksiyonu Eşitlik (3.8) şeklinde yazılabilir.

$$\begin{aligned} \ln l(\beta_0, \beta_1) &= \ln \left( \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \right) \\ &= \sum_{i=1}^n (\ln (\pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i})) \\ &= \sum_{i=1}^n (\ln (\pi(x_i)^{y_i}) + \ln ([1 - \pi(x_i)]^{1-y_i})) \\ &= \sum_{i=1}^n (y_i \ln \pi(x_i)) + (1 - y_i) \ln (1 - \pi(x_i)) \\ &= \sum_{i=1}^n (y_i \ln \pi(x_i)) - (y_i) \ln \pi(x_i) + \ln (1 - \pi(x_i)) \\ &= \sum_{i=1}^n \left( y_i \ln \left( \frac{e^{(\beta_0 + \beta_1 x_1)}}{1 + e^{(\beta_0 + \beta_1 x_1)}} \right) - y_i \ln \left( \frac{1}{e^{(\beta_0 + \beta_1 x_1)}} \right) + \ln \left( \frac{1}{e^{(\beta_0 + \beta_1 x_1)}} \right) \right) \end{aligned}$$

$$\ln l(\beta_0, \beta_1) = \sum_{i=1}^n (y_i (\beta_0 + \beta_1 x_1) - \ln (1 + e^{(\beta_0 + \beta_1 x_1)})) \quad (3.8)$$

Eşitlik (3.8)'in  $\beta_0$  ve  $\beta_1$  parametrelerine göre türevi alınıp sıfıra eşitlendiğinde,

$$\frac{d \ln l(\beta_0, \beta_1)}{d \beta_0} = \sum_{i=1}^n \left( y_i - \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_1}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_1}} \right)$$

$$\begin{aligned}
&= \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_1)) = 0 \\
\frac{d \ln l(\beta_0, \beta_1)}{d\beta_1} &= \sum_{i=1}^n \left( x_i y_i - x_i \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_1}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_1}} \right) \\
&= \sum_{i=1}^n x_i (y_i - \hat{\pi}(x_i)) \\
&= \sum_{i=1}^n x_i (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_1)) = 0 \quad (3.9)
\end{aligned}$$

elde edilir. Yukarıda görüldüğü üzere, parametre tahminleri Log-olabilirlik fonksiyonunun maksimize edilmesiyle bulunur. Ancak elde edilen denklemin türev işlemleriyle çözülmesi sonucunda kapalı bir çözüm elde edilememektedir. Bu amaçla nümerik çözüm için literatürde sıklıkla kullanılan Newton-Raphson algoritması kullanılabilir.

Bağımlı değişkeninin alacağı değerleri (0,1) aralığına çekebilmek için dönüşüm yapılması gerekmektedir.  $\frac{\pi(x)}{[1 - \pi(x)]}$  dönüşümü bağımlı değişkeninin sınırlarını (0,∞) aralığına taşımaktadır. Bağımlı değişkenin sınırlarının (-∞, +∞) aralığında olması için ise  $\frac{\pi(x)}{[1 - \pi(x)]}$  oranının logaritmasının alınması gerekmektedir. Bu dönüşüm yardımıyla elde edilen yeni bağımlı değişken bağımsız değişkenin doğrusal bir fonksiyonu olmaktadır ve bu dönüşüme logit dönüşüm adı verilmektedir (Atabey, 2010). Lojistik regresyon modelinin lojit dönüşümü Eşitlik (3.10)' de verilmiştir.

$$\ln \left[ \frac{\pi_i}{1 - \pi_i} \right] = \beta_0 + \sum_{j=1}^p x_j^T \beta_j \quad i = 1, 2, \dots, n \quad (3.10)$$

Bir lojistik regresyon modelinde parametreleri yorumlamak nispeten kolaydır. Lojistik regresyonda modelde anlamlı bulunan katsayıların yorumlanmasını kolaylaştırmak için Odd's oranı kullanılabilir. Odds oranı, bağımsız değişken değerindeki bir birim değişime karşılık gelen başarı olasılığındaki kestirilmiş artış olarak da tanımlanabilir (Erar, 2013). Bu oran bir bağlantı ölçümüdür. Lojistik regresyon analizi sonuçlarındaki tablolarda bu oran fazladan sütun olarak bulunur. Odds oranı Eşitlik (3.11)' de verilmiştir.

$$\hat{Q}_R = \frac{odds_{x_{i+1}}}{odds_{x_i}} = e^{\hat{\beta}_1} \quad (3.11)$$

İkili lojistik regresyon modeli için odd's oranı Eşitlik (3.12)' de verilmiştir.

$$\psi = e^{\beta} \quad (3.12)$$

Odds oranı yardımıyla lojistik regresyon modeli Eşitlik (3.13)' de ki gibi yazılabilir.

$$\frac{\pi(x_i)}{1-\pi(x_i)} = e^{\beta_0 + \beta_1 x_1} \quad (3.13)$$

### 3.3. Cezalı Regresyon

Veri setinde çoklu iç ilişki problemi mevcut olduğunda veya yüksek boyutlu verilerin ( $p > n$ ) olması durumunda klasik tahmin yöntemlerine alternatif olarak cezalandırılmış regresyon yöntemleri yaygın olarak kullanılmaktadır. Bu yöntemler aynı zamanda büzülme (shrinkage) ya da düzenleme (regularization) yöntemleri olarak da bilinmektedir (James vd. 2014; Bruce ve Bruce,2017). Geleneksel değişken seçim yaklaşımlarından farklı olarak, cezalandırılmış regresyon yöntemleri, model seçimi ile parametrelerin tahminini eş anlı olarak gerçekleştirmektedir. Cezalı regresyon yöntemlerinde, katsayılar üzerinde ceza terimleri kullanarak regresyon katsayıları tahmin edilir. Bu yöntemlerde genellikle regresyon katsayılarına bir miktar yanlılık eklenerek katsayıların varyansının azaltılması amaçlanmaktadır (Genç, 2020).

Cezalı regresyon yöntemlerine geçilmeden önce çoklu iç ilişki problemi açıklanacaktır.

### 3.4. Çoklu İç İlişki

Önceki bölümde ifade edilen lojistik regresyon modelinin varsayımlarından biri olan bağımsız değişkenler arasında ilişki yoktur varsayımının sağlanamaması durumunda çoklu iç ilişki problemi ortaya çıkmaktadır. Çoklu iç ilişkinin ortaya çıkmasındaki başlıca nedenler ilerleyen kısımda açıklanmıştır.

#### 3.4.1. Çoklu iç ilişkinin nedenleri

Bir veri setinde ele alınan değişkenler arasında çoklu iç ilişkinin bulunmasının birçok nedeni bulunmaktadır. Çoklu iç ilişkinin nedenleri, kullanılan veri toplama yöntemi, model ya da kitle üzerindeki kısıtlamalar, modelin özellikleri veya model tanımından kaynaklanan yanlışlıklar olabilir. Çoklu iç ilişkinin bu dört farklı kaynağı arasındaki farkları anlamak önemlidir. Verilerin analizi ve belirlenen modelin yorumlanması problemin nedenine bağlıdır. Kullanılan veri toplama yöntemi, bağımsız değişkenlerin tanım bölgesinin yalnızca bir alt uzayından örneklem aldığı zaman, çoklu iç ilişki sorununa yol açabilir. Model ya da kitle üzerindeki kısıtlamalar genellikle üretimsel ya da kimyasal süreçleri içeren problemlerde ortaya çıkmaktadır. Çoklu iç ilişki, modelin yanlış seçiminden de kaynaklanabilmektedir. Bir regresyon modeline polinomial terimler eklenmesi çoklu iç ilişki problemine yol açabilir. Bağımsız değişkenin tanım aralığı daha küçük iken, regresyon modeline bu rassal değişkenin karesel veya daha üst düzeyde dönüşümünü içeren terimler eklenmemelidir. Aşırı tanımlanmış bir model, gözlemlerden çok daha fazla sayıda bağımsız değişkenlerin olduğu modellerdir. Bu gibi durumlarda çoklu iç ilişki ile baş edebilmek için değişken seçimi yöntemlerinden yararlanılabilir.

#### 3.4.2. Çoklu iç ilişkinin belirlenmesi

Çoklu iç ilişkiyi belirlemek için korelasyon matrisinin incelenmesi, varyans şişirme faktörü (Varyans İnflasyon Faktörü-VIF), özdeğer ve özvektör analizi, koşul sayısı ve koşul endeksi sıklıkla kullanılan yöntemlerdir.

$X'X$  matrisinin köşegen dışı elemanlarının incelenmesi çoklu iç ilişki sorununun en basit ölçüsüdür.  $|r_{ij}| \rightarrow 1$  iken  $x_i$  ve  $x_j$  doğrusal bağımlıdır denir. VIF değerlerinin 10'dan büyük olması çoklu iç ilişki sorununun varlığını göstermektedir.  $X'X$  matrisinin öz değerlerinden bir veya daha fazlasının sıfıra eşit olması çoklu iç ilişki problemini göstermektedir. Çoklu iç ilişkinin varlığını gösteren bir başka yöntem ise  $|X'X| = 0$  olmasıdır. Bu durum tam çoklu iç ilişki olarak adlandırılmaktadır.

Çoklu iç ilişki durumunda elde edilen regresyon modeli tahmini gerçeği yansıtmaktan uzaktır. Böyle bir modelden elde edilecek sonuçlar ve çıkarımlar yanlış olacaktır. Çoklu iç ilişki problemini gidermek için ilk aşamada ek veri toplanabilir veya model yeniden belirlenebilir. Ancak ek veri toplanması, ekonomik kısıttan ya da çalışılan konunun artık örnekleme uygun olmamasından dolayı her zaman mümkün değildir. Ek veriler var olsa bile, bağımsız değişkenlerin aralığını araştırmacının inceleme alanı dışına taşıyabilir. Bu durumda ek verilerin toplanması ve kullanılması uygun olmayabilir. Modelin yeniden belirlenmesiyle ilgili bir yaklaşım, değişken seçimidir. Bu yaklaşım, bağımsız değişkenler doğrusal ilişkiliyse bu değişkenlerden birinin elenmesi şeklinde açıklanabilir. Ancak, modelden çıkartılan bağımsız değişkenlerin bağımlı değişkenin açıklanmasına katkısı varsa bu teknik modelin kestirim gücünü azaltmaktadır (Erar, 2013). Değişken seçimi yöntemi ya da ek veri toplanması bazı durumlarda çoklu iç ilişki probleminin üstesinden gelmekte yetersiz kalmaktadır. Özellikle son zamanlarda yüz tanıma, tıbbi verilerin sınıflandırılması gibi birbiriyle yapısal olarak ilişkili değişkenlerle çalışılması durumunda yanlı regresyon yöntemlerinin kullanılması kaçınılmazdır. Yanlı regresyon yöntemleri içerisinde yaygın kullanıma sahip olması nedeniyle bu çalışmada Ridge regresyon yöntemi incelenmiştir.

### 3.5. Ridge Regresyon

Çoklu iç ilişki probleminin üstesinden gelmek için literatürde sıklıkla kullanılan yanlı regresyon yöntemlerinden biri Ridge Regresyon'dur. Ridge regresyon ilk kez Hoerl ve Kennard (1970) tarafından önerilmiştir. Çoklu doğrusal regresyon modelinde  $X'X$  matrisinin köşegen elemanlarına belirli bir  $k$  sabiti eklenerek elde edilen ridge tahmin edicisi Eşitlik (3.14)'de verilmiştir.

$$\hat{\beta}_{Ridge} = (X'X + kI)^{-1}X'y, k \geq 0 \quad (3.14)$$

Ridge tahmin edicisinin, EKK'dan farklılığı  $k$  ridge parametresinin varlığıdır (Hoerl ve Kennard, 1970; Lawless ve Wang, 1976; Dempster vd. 1977; Gibbons, 1981; Kibria, 2003; Khalaf ve Shukur, 2005; Alkhamisi ve Shukur 2008; Mansson ve Shukur, 2013).  $k$  sabiti 'ridge' veya 'yanlılık' parametresi olarak adlandırılmaktadır.  $k$  değeri seçilirken varyanstaki azalış, yanlılığın karesindeki artıştan daha büyük olacak şekilde seçilmelidir. Bu durumda ridge tahmin edicisinin MSE değeri, EKK tahmincisinin MSE değerinden daha küçüktür (Gültay, 2009). Ridge tahmin edicisinin MSE değerini minimum yapmak  $k$  parametresinin seçimine bağlıdır. Bu yüzden  $k$ 'nın seçimi önemli bir ölçüttür.  $k$ 'nın değerini tahmin etmek için literatürde yapılmış çok sayıda çalışma vardır (Hoerl ve Kennard, 1970; 1975; Lawless ve Wang, 1976; Dempster vd. 1977; Gibbons, 1981; Kibria, 2003; Khalaf ve Shukur, 2005; Alkhamisi ve Shukur 2008). Çoklu doğrusal regresyon modelinde  $k$  ridge parametresinin belirlenmesi için literatürde birçok yöntem vardır. Bu tez çalışmasında lojistik ridge regresyon yöntemi için  $k$  ridge parametresi çapraz geçerlilik testi yardımıyla belirlenmiştir.

Bağımsız değişkenler arasında çoklu iç ilişki olduğunda, EÇO tahmin edicisi kullanıldığında bazı özdeğerler küçük olacağından varyans büyük çıkabilir. Lojistik regresyon modelinde çoklu iç ilişki probleminin üstesinden gelmek için sıklıkla lojistik ridge regresyon yöntemi kullanılmaktadır. Lojistik ridge regresyonu ilerleyen kısımda açıklanmıştır.

### 3.6. Lojistik-Ridge Regresyon

Doğrusal regresyonda çoklu iç ilişki probleminin giderilmesi için belirlenen yöntemler bazı araştırmacılar tarafından lojistik regresyon modeline uyarlanmıştır (Schafer vd., 1984; Weissfeld ve Sereika,1991). Bağımlı değişken iki düzeyli olduğu durumda lojistik ridge tahmin edicisi kullanılabilir. İlk kez Duffy ve Santer (1989) tarafından önerilen lojistik ridge regresyonu Eşitlik (3.15)' de verilmiştir.

$$\hat{\beta}_{\log\text{-ridge}} = \underset{\beta}{\operatorname{argmin}}[-\sum_{i=1}^n \{y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)\} + \lambda \sum_{j=1}^p \beta_j^2] \quad (3.15)$$

Eşitlik (3.15)' de verilen lojistik ridge tahmin edicisinin log-olabilirlik fonksiyonu Eşitlik (3.16)' da verilmiştir (Dufy ve Santer, 1989).

$$\ln l(\beta) = \sum_{i=1}^n y_i \ln\left(\frac{\pi_i}{1-\pi_i}\right) - \sum_{i=1}^n \ln(1 + e^{x_i' \beta}) + \lambda \sum_{j=1}^p \beta_j^2 \quad (3.16)$$

Lojistik regresyon modelinde matris gösterimiyle ridge regresyonun EÇO tahmin edicisi Eşitlik (3.17)' de verilmiştir.

$$\hat{\beta}_{E\text{ÇO}} = (X' \hat{V} X)^{-1} X' \hat{V} \hat{Z} \quad (3.17)$$

$\hat{\beta}_{E\text{ÇO}}$  tahmin edicisi  $\hat{V}$  ağırlık matrisinden dolayı çoklu iç ilişkiden etkilenmektedir.

$$\hat{V} = \operatorname{diag}[\hat{\pi}(1 - \hat{\pi})]$$

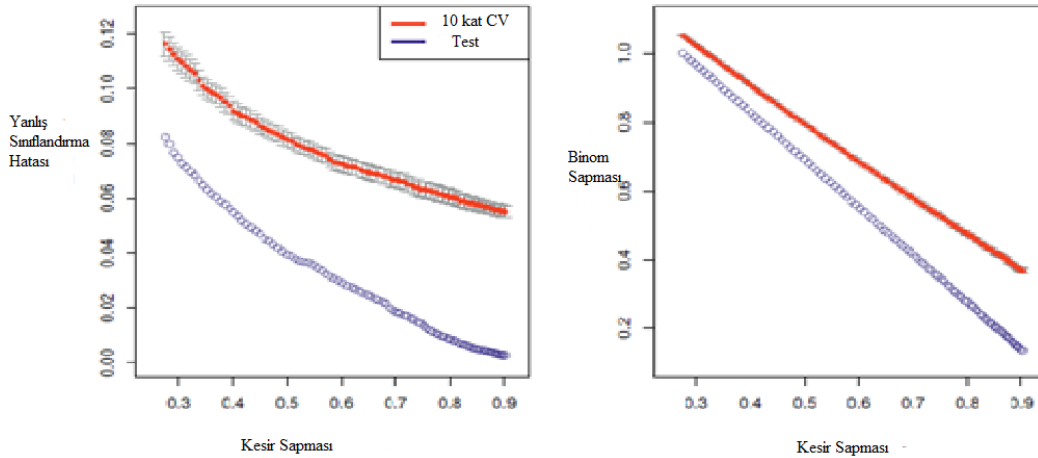
Eşitlik (3.27)' de verilen  $\hat{Z}$ ,  $i$ 'inci elemanı  $\hat{z}_i = \log(\hat{\pi}_i) + \frac{\hat{y}_i - \hat{\pi}_i}{1 - \hat{\pi}_i}$  olan bir vektördür. EÇO tahmin edicisinin,  $E(\hat{\beta}_{E\text{ÇO}}) = \beta$  ve  $\operatorname{Var}(\hat{\beta}_{E\text{ÇO}}) = (X' \hat{V} X)^{-1}$  dir.

Lojistik ridge tahmin edicisinin cezalı regresyon şeklinde ifadesi Eşitlik (3.18)' de verildiği gibidir.

$$\underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \{y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))\} + \lambda \sum_{j=1}^p \beta_j^2 \quad (3.18)$$

Lojistik ridge regresyonda, ayar parametresi  $\lambda$  büzülme miktarını kontrol eder, ancak hiçbir zaman bağımsız değişken katsayılarını tam olarak sıfır yapmaz. Burada  $\lambda$  ayar parametresi çapraz geçerlilik testiyle belirlenmiş olup ilerleyen kısımda bu konu hakkında bilgi verilmiştir.





Şekil: 3.1: Ridge ( $L_2$ ) cezalı lojistik regresyon (Hastie vd., 2015)

Şekil (3.1)' de Lang (1995) tarafından verilen Haber Grubu (Newsgroup Data) verilerine ait on kat çapraz geçerlilik grafikleri verilmiştir. Bu grafiklerde, on kat çapraz doğrulama eğrileri noktalı standart hata bantlarıyla birlikte kırmızı renkte gösterilir. Sol grafik yanlış sınıflandırma hatasını gösterirken sağdaki grafik çarpıklığı göstermektedir. Bu ölçümlerin her biri için eğitim hatası mavi renkte gösterilmektedir (Hastie vd., 2015).

Lojistik Ridge regresyon modelinde, bağımsız değişkenlerin tamamı için bir katsayı tahmini söz konusudur. Bu durum, model katsayılarının yorumlanmasını zorlaştırmakta ve her parametre için tahmini zorunlu kıldığından tahmin değerinin yanlılığını artırmaktadır. Değişken seçimi yaklaşımıyla çoklu iç ilişki probleminin üstesinden gelen farklı cezalı regresyon yöntemleri vardır. Cezalı regresyon yöntemleri içerisinde yer alan LASSO ilk kez Tibshirani (1996) tarafından önerilmiştir. Bağımlı değişkenin ikili olduğu durumlarda kullanılan lojistik regresyon modeli LASSO yöntemine uyarlanmıştır (Hastie vd., 2005). Lojistik LASSO modelinin açıklanmasına geçmeden önce En küçük açı regresyonuna değinilecektir.

### 3.7. En Küçük Açı Regresyonu (Least Angel Regression-LARS)

Kısıtlı optimizasyon problemlerinin çözümü için çeşitli nümerik yöntemler kullanılmaktadır. Cezalı regresyon analizi gibi kısıtlı optimizasyon problemleri için LARS en çok kullanılan algoritmalarından biridir (Efron vd., 2004).

Çoklu iç ilişki problemi varlığında cezalı regresyon yöntemleri de kullanılabilir. En küçük açı regresyonu (Least Angle Regression and Shrinkage- LARS) ilk adımda bağımlı değişkenle en ilişkili değişkeni tanımlar. LARS yöntemi, tanımlanan değişkenin katsayısını sürekli olarak en küçük kareler değerine doğru hesaplar. Başka bir değişken modele dahil edildiğinde, süreç duraklatılır. İkinci değişken daha sonra aktif kümeye katılır ve katsayıları korelasyonlara bağlı ve azaltacak şekilde birlikte hesaplar. Bu işlem, tüm değişkenler modelde olana kadar devam eder ve iterasyon durur.  $p > N - 1$  ise, LARS algoritması  $N - 1$  adımından sonra sıfır artık çözümüne ulaşır (Hastie vd., 2015). Başka bir ifadeyle bu yöntem, tüm mümkün aday değişkenlerden en fazla korelasyona sahip olan değişkeni seçer ve o değişkenle bağımlı değişken arasında regresyon gerçekleştirir. Diğer bağımsız değişkenleri geometrik bir düzene göre seçecek tekrarlı bir algoritma süreci başlatmaktadır. Belli bir basamak sonra algoritma sonlandırılır. Böylece aktif bir değişken seti oluşturulmaktadır. Seçilen değişkenler yardımıyla parametreler tahmin edilmektedir.

LASSO'nun çözümü için LARS ya da APG algoritması kullanılırken, Elastik Ağ'ın çözümü için APG algoritması kullanılabilir. LARS algoritması LASSO'nun artık karelerle çözülmesi için bir süreçtir. Bu süreçte tüm çözüm sonuçları LASSO'nun  $\lambda$  parametresinin bir fonksiyonu olarak elde edilir (Hastie vd, 2015). Elastik Ağ için Chang vd., (2018) çalışmasında kullanılmış olan hızlandırılmış proksimal gradyan (Accelerated Proximal Gradient - APG) algoritması Eşitlik (3.20)' de verilen minimizasyon problemini çözmek için kullanılmıştır.

APG algoritması aşağıda açıklanmıştır. Bu minimizasyon problemi  $f: R^n \rightarrow R$  diferansiyebilir konveks fonksiyon ve  $g: R^n \rightarrow (-\infty, \infty]$  düzgün-yarı sürekli konveks fonksiyon olmak üzere;

$$\underset{x}{\operatorname{argmin}}(F(x) = f(x) + g(x)) \quad (3.19)$$

şeklindedir. Eşitlik (3.19)' da verilen minimizasyon probleminin çoklu doğrusal regresyon modeli için LASSO' ya uyarlanmış hali Eşitlik (3.20)' de verilmiştir.

$$f(\beta) = \|y - X\beta\|^2 \text{ ve } g(\beta) = \lambda\|\beta\|_1 \quad (3.20)$$

Eşitlik (3.20)' de,  $\|\cdot\|$ ,  $L_1$  normudur. Gradient Eşitlik (3.21) ve proksimal Eşitlik (3.22)' de verilmiştir.

$$\nabla f(\beta) = X^T(X\beta - y) \quad (3.21)$$

$$\text{prox}_{t_g}(\beta) = \text{argmin} \left\{ \|z\|_1 + \frac{1}{2\lambda t} \|z - \beta\|^2 \right\} = S_{\lambda t}(\beta) \quad (3.22)$$

Eşitlik (3.22)' de  $S_{\lambda t}(\beta)$  ayarlanabilir başlangıç fonksiyonudur ve Johnstone (1994) tarafından ifade edilen bu fonksiyon Eşitlik (3.23)' de verilmiştir.

$$S_{\lambda t}(\beta) = \begin{cases} v - \lambda, & v > \lambda \\ 0, & |v| \leq \lambda \\ v + \lambda, & v < -\lambda \end{cases} \quad (3.23)$$

Duabechies vd., (2004) tarafından proksimal gradyan (PG) yöntemi Eşitlik (3.24)' de ifade edilmiştir.

$$\beta_{k+1} = S_{\lambda t_k}(\beta_k - t_k X^T(X\beta_k - y)) \quad (3.24)$$

Eşitlik (3.24)' de seçilen uygun adım boyutu  $0 < t_k < \frac{1}{L}$  şeklindedir.  $L$ ,  $\nabla f$  için Lipschitz sabitidir. Beck ve Teboulle (2009)' da PG yöntemindeki yakınsama oranının artırılması için APG yöntemini önermişlerdir. Bu yöntem PG yöntemi gibi hesaplama kolaylığına sahiptir ve bir global yakınsama oranı sunar. LASSO'nun çözümü için APG algoritmasını kullanmıştır.

$$\tilde{\beta}_k = \beta_k + \frac{k-1}{k+2}(\beta_k - \beta_{k-1}) \quad (3.25)$$

$$\beta_{k+1} = S_{\lambda t_k}(\tilde{\beta}_k - t_k X^T(X\tilde{\beta}_k - y)) \quad (3.26)$$

Burada, uygun seçilen adım boyutu  $0 < t_k < \frac{1}{L}$  şeklindedir (Chang vd., 2018).

Eşitlik (3.19)' da verilen minimizasyon probleminin çoklu doğrusal regresyon modeli için Elastik ağ yöntemine uyarlanmış hali Eşitlik (3.27)' de verilmiştir. Elastik Ağ yöntemi  $L_1$  ve  $L_2$  kısıtı altında parametreleri tahmin etmektedir.

$$f(\beta) = \|y - X\beta\|^2 \text{ ve } g(\beta) = \lambda\|\beta\|_1 + \lambda_2\|\beta\|_2^2 \quad (3.27)$$

### 3.7.1. En az mutlak büzülme ve seçim operatürü (Least Absolute Shrinkage and Selection - LASSO)

Çoklu iç ilişki probleminin giderilmesi için değişken seçimi yaklaşımının benimsenmesi durumunda hangi değişkenin modelden çıkarılacağı önemlidir. Model için önemli olan bir değişkenin modelden çıkarılması modelin doğruluğunu azaltacaktır. Değişken seçimi ve parametre tahminini eş zamanlı yapan daraltıcı (shrinkage) yöntemler bu amaçla geliştirilmiştir (Hoerl ve Kennard, 1970). Daraltıcı yöntemlerden biri olan LASSO Tibshirani (1996) tarafından önerilmiştir. LASSO, regresyon modelleri için bir büzülme ve değişken seçim yöntemidir (Melkumova ve Shatskikh, 2017). LASSO tahmin edicileri, çoklu regresyon modelinin herhangi bir biçimi için model katsayılarının vektörüne  $L_1$  norm kısıtlaması eklenmesiyle elde edilir. Genel olarak  $\beta$  üzerindeki  $L_d$  norm kısıtlaması Eşitlik (3.28) de verildiği gibidir.

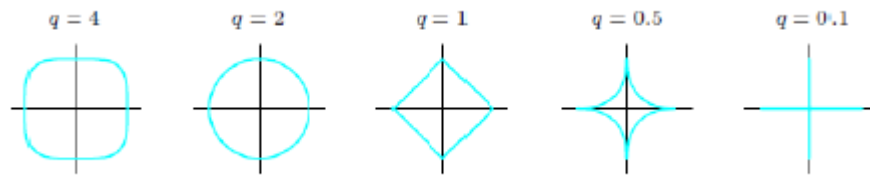
$$\|\beta\|_d = \left( |\beta_1|^d + |\beta_2|^d + \dots + |\beta_p|^d \right)^{1/d} \leq t \quad (3.28)$$

$$t > 0$$

Buradan hareketle  $L_1$  norm kısıtlaması aşağıdaki şekilde yazılabilir.

$$\sum_{k=1}^p |\beta_k| \leq t, t > 0$$

Burada  $t$ , kullanıcı tarafından belirlenebilen veya çapraz geçerlilik, genelleştirilmiş çapraz geçerlilik gibi çeşitli yöntemlerle hesaplanabilen bir ayar parametresidir (Tibshirani, 1996).



Şekil. 3.2: L-q normu için farklı  $q$  değerlerinin grafiksel gösterimi (Friedman vd., 2001)

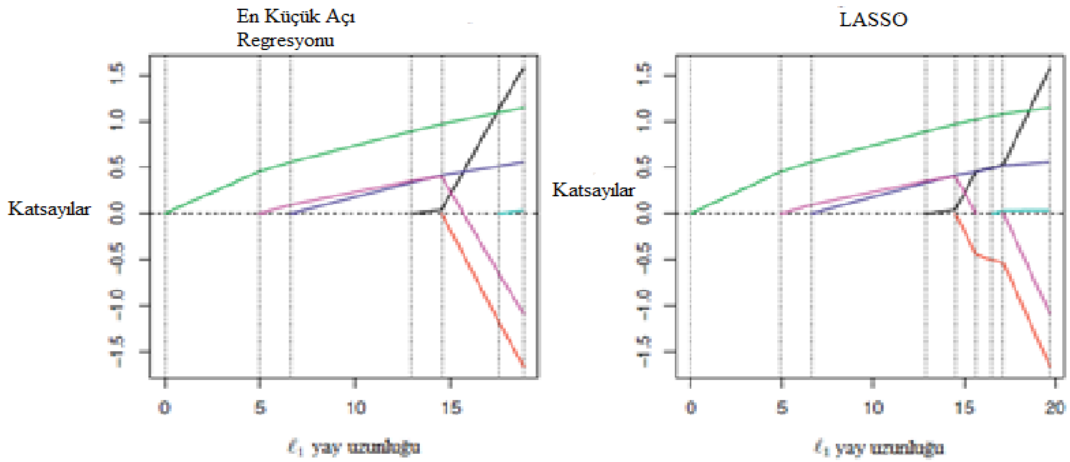
Şekil (3.2) incelendiğinde, düzlem üzerinde oluşturan alanlar model için belirlenen kısıt alanını göstermektedir.  $q = 1$  için LASSO tahmin edicisinde katsayılarından bazıları sıfır olmaktadır. EKK doğrusunun köşelere temas ettiği noktalarda katsayılar sıfır olmaktadır.

Tibshirani (1996), LASSO regresyonu  $\sum |\beta_j| < t$  koşulu ile  $\min_{\beta} (y - X\beta)'(y - X\beta)$  olarak tanımlamıştır.  $y_i$  yanıt değerleri ve  $x_{ij}$  bağımsız değişkenler olmak üzere çoklu doğrusal regresyon modeli için LASSO tahmincisi Eşitlik (3.29)' de verilmiştir.

$$\hat{\beta}_{LASSO} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - x_i' \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (3.29)$$

$$i = 1, 2, \dots, n \quad j = 0, 1, \dots, p$$

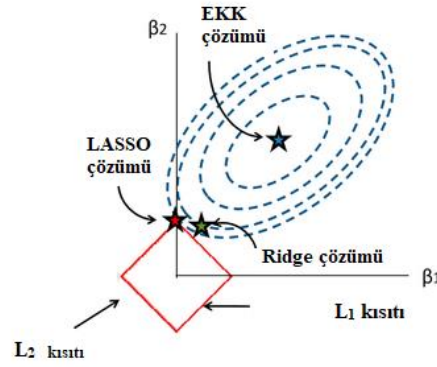
Eşitlik (3.38) ifadesini minimum yapacak  $\beta = \{\beta_j\}$  parametresini bulan cezalandırılmış  $L_1$  problemini çözer.



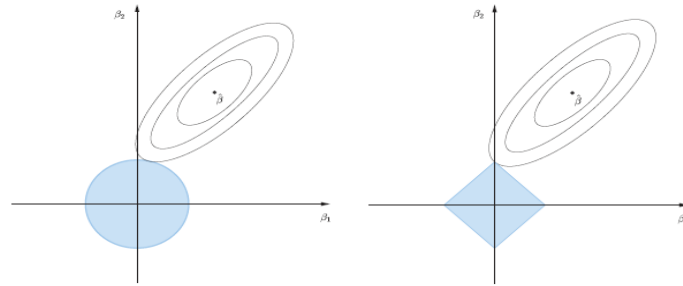
Şekil. 3.3: LARS ve LASSO profili (Hastie vd., 2015)

Şekil (3.3) incelendiğinde soldaki grafik  $L_1$  yay uzunluğunun bir fonksiyonu olarak gelişen LARS katsayı profilini göstermektedir. Sağdaki şekil ise soldaki panel ile neredeyse aynıdır ve LASSO katsayılarının profilini göstermektedir. Pembe çizgi ile gösterilen katsayı sıfırdan geri döndüğünde ilk kez farklılık göstermektedir. Panellere bakıldığında eğer sıfır olmayan bir katsayı, bir sonraki değişken algoritmaya girmeden önce sıfırı geçerse bu katsayı değişken aralığında bırakılır. (Hastie vd., 2015).

### 3.7.1.1. LASSO'nun grafiksel yorumu



Şekil 3.4: EKK, Ridge ve LASSO'nun Grafiksel Yorumu



Şekil 3.5: Ridge (solda) ve LASSO (sağda) İki Boyutlu Tahmin Durumu (Hastie vd., 2015)

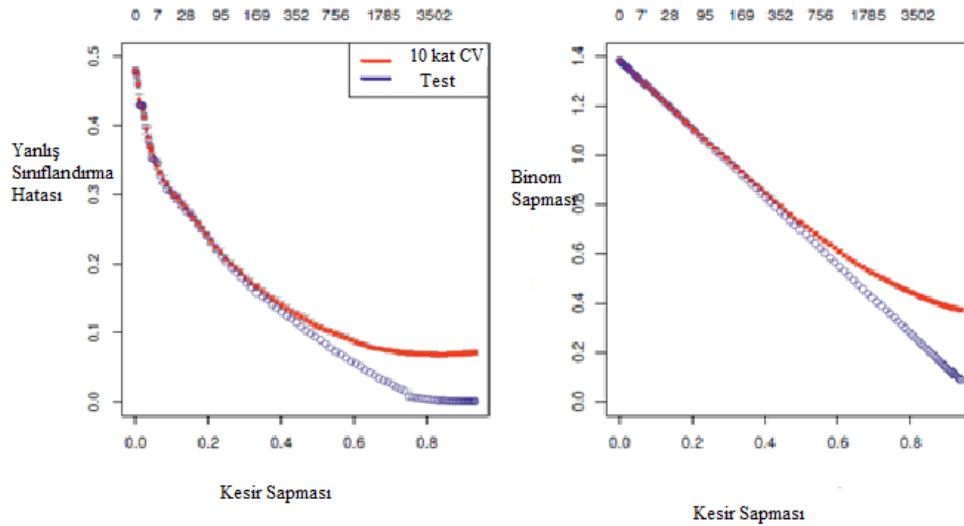
Şekil (3.4) için mavi bölgeler sırasıyla Ridge ve LASSO için kısıtlama alanlarıdır. Eliptik kontürler ise EKK tahmincisinin yayılımını göstermektedir. Bir vektör konveks fonksiyonunu belirli bir noktada kesiyor ise bu noktada fonksiyon minimum olmaktadır. Dolayısıyla Şekil (3.4) için LASSO tahmincisinin tam olarak sıfıra ayarlanabilen katsayılar elde edebildiği ve ridge regresyonunda EKK için belirli bir kısıta kadar daralabildiği fakat sıfır olabilen değerler elde edilemediği sonucuna ulaşılabilir (Varlickova, 2017; Yaman, 2017).

### 3.7.2. Lojistik-LASSO

Lojistik regresyon için LASSO tahmin edicileri, negatif log-olabilirlik fonksiyonuna ceza terimini ekleyerek elde edilmektedir. Ridge regresyonda olduğu gibi LASSO tahmin edicileri, log-olabilirlik fonksiyonunun maksimizasyonu yardımıyla elde edilir. LASSO, parametre uzayı üzerinde kısıt olarak  $L_2$  normu kullanmak yerine,  $L_1$ -normu kullanmaktadır. LASSO cezalı lojistik regresyon yöntemi negatif log-olabilirlik fonksiyonuna ceza teriminin eklenmesiyle elde edilir ve Eşitlik (3.30)'da verilmiştir.

$$\hat{\beta}_{\log-LASSO} = \underset{\beta}{\operatorname{argmin}} [-\sum_{i=1}^n \{y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)\} + \lambda \sum_{j=1}^p |\beta_j|] \quad (3.30)$$

LASSO birçok uygulamada yaygın olarak kullanılmasına rağmen, bazı dezavantajları vardır. Bu dezavantajlardan biri, bağımsız değişkenler arasındaki yüksek korelasyona karşı sağlam olmaması ve bu değişkenlerden birini rastgele seçip geri kalanını görmezden gelmesidir. LASSO'nun bir başka dezavantajı,  $p > n$  olduğunda yüksek boyutlu verilerde, en fazla  $n$  bağımsız değişken seçerken, son modelde sıfır olmayan değerlere sahip  $n$ 'den daha fazla bağımsız değişken katsayılarının olabilmesidir (Zhou, 2013).



Şekil 3.6: LASSO ( $L_1$ ) cezalı lojistik regresyon

Şekil (3.6) incelendiğinde, sol taraftaki grafik yanlış sınıflandırma hatasını gösterir. Sağdaki grafik ise çarpıklığı göstermektedir. Bu ölçümlerin her biri için eğitim hatası mavi renk ile gösterilmiştir. Her modeldeki sıfır olmayan katsayıların sayısı, her grafiğin üstünde gösterilmektedir.

LASSO'nun yaygın kullanımı nedeniyle dezavantajlarının giderilmesi amaçlanmıştır. Bu amaçla LASSO ya yeni bir kısıt eklenerek Elastik ağ regresyonu önerilmiştir (Zou ve Hastie, 2005). Bu tez çalışmasının bir sonraki adımında Elastik ağ yöntemi hakkında bilgiler verilmiştir.

### 3.8. Elastik Ağ

LASSO, model seçimi ve parametrelerin tahmininde yüksek boyutlu veriler için oldukça kullanışlıdır. Ancak LASSO'nun, verinin yapısına bağlı olarak bazı problemleri çözmeye dezavantajları vardır (Genç, 2020). Bu dezavantajlardan ilki, klasik bir veri setinde çoklu iç ilişki problemi varlığında yüksek derecede ilişkili değişkenlerden birini rasgele seçmesidir. Bu durumda modelde olması gereken bir değişkenin çoklu iç ilişkii nedeniyle modelden çıkarılmış olması sorunuyla karşılaşmaktadır. Bir diğer problem ise, bağımsız değişken sayısının ( $p$ ) gözlem sayısından ( $n$ ) fazla olduğu çok boyutlu veri durumunda ( $p \gg n$ ), LASSO tahmircisi en fazla  $n$  değişken için tahmin yapabilmektedir. Başka bir dezavantaj ise, LASSO tahmircisinin ceza fonksiyonunun tam olarak dışbükey olmamasıdır. Bu durumda modele uydurulduğunda farklı bağımsız değişkenlerin sırasına göre farklı tahminler elde edilmesine neden olur (Zou ve Hastie, 2005).

LASSO'nun belirtilen bu dezavantajlarının azaltılması amacıyla Zou ve Hastie (2005) tarafından Elastik ağ yöntemi önerilmiştir. Bu yöntem, Ridge ile LASSO cezaları arasında bir uzlaşma sağlamaktadır (Zou ve Hastie, 2005). Çoklu doğrusal regresyon modeli için  $\beta'$  yi minimum yapacak Elastik ağ tahmircisi Eşitlik (3.31)'de verilmiştir.

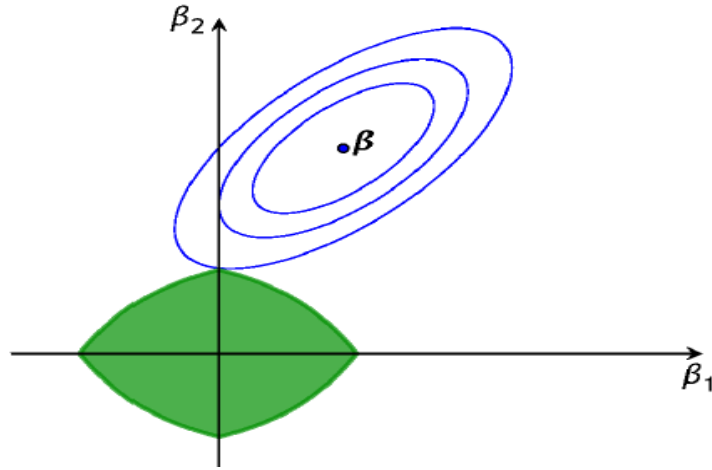
$$\hat{\beta}(\lambda_1, \lambda_2) = \underset{\beta}{\operatorname{argmin}} \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \frac{\lambda_2}{2} \|\beta\|_2^2 \quad (3.31)$$



Eşitlik (3.31)' de  $\lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2$  ifadesi elastik ağ tahmincisinin ceza fonksiyonudur. Bu terim  $L_1 + L_2$  norm ceza terimlerinin lineer birleşimidir. Burada  $\lambda_1$  ve  $\lambda_2$  lagrange katsayılarıdır. Eşitlik (3.40)' da  $0 \leq \alpha \leq 1$  olduğunda  $\lambda_1 = \lambda\alpha$  ve  $\lambda_2 = \lambda \left(\frac{1-\alpha}{2}\right)$  olur. Belirtilen  $\lambda_1$  ve  $\lambda_2$  Eşitlik (3.31)' de yerine yazıldığında Eşitlik (3.32) elde edilmektedir.

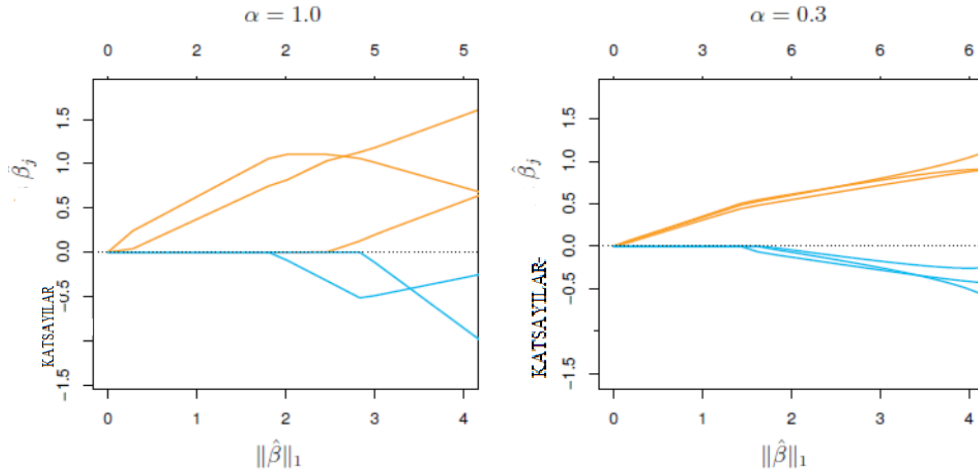
$$\hat{\beta}(\lambda_1, \lambda_2) = \underset{\beta}{\operatorname{argmin}} \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda\alpha \|\beta\|_1 + \frac{\lambda(1-\alpha)}{4} \|\beta\|_2^2 \quad (3.32)$$

Eşitlik (3.32)' de verilen elastik ağ tahmin edicisinin ceza fonksiyonunda  $\alpha = 1$  olduğunda  $L_1$  norma yani LASSO cezasına indirildiği görülmektedir. Benzer şekilde  $\alpha = 0$  olduğunda ise  $L_2$  norma yani Ridge cezasına indirgenir. Dolayısıyla Elastik ağ tahmin edicisi Ridge ve LASSO tahmin edicileri arasında denge bulmaya çalışan bir yöntemdir. Ridge regresyonda olduğu gibi tüm değişkenlere ait bir katsayı hesaplamak ile çok katı bir şekilde değişken seçimi yapan LASSO arasında denge kurmak amacıyla 2 ayar parametresi kullanılmaktadır.



Şekil 3.7: Artıkların Konturleri ve Elastik Ağ Tahmin Edicisinin Sınırlama Bölgesi (Zou ve Hastie, 2005)

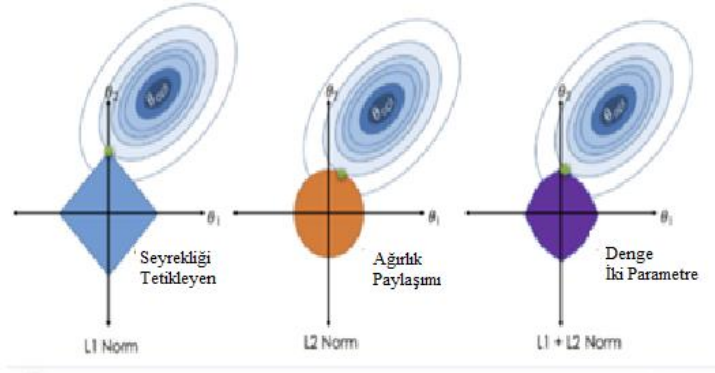
$p = 2$  için Elastik ağ tahmin edicisinin ceza fonksiyonunun grafiği Şekil (3.7)' de gösterilmektedir. Bu şekilde Elastik ağ tahmincisi, artık kareler toplamının konturlarının ilk kez temas ettiği noktaya karşılık gelmektedir. Aynı zamanda Ridge ve LASSO tahmin edicilerinin bir kombinasyonu olduğu görselden anlaşılmaktadır (Genç, 2020).



Şekil.3.8: LASSO ve Elastik Ağ Yöntemlerinin  $\alpha$  Parametresine Göre Davranışları (Hastie vd., 2015)

Şekil (3.8) incelendiğinde, üç değişkenli gruplar arasında yüksek korelasyonlu altı değişken için LASSO tahminleri ( $\alpha= 1$ ), sol panelde gösterildiği gibi, normalleştirme parametresi değiştikçe biraz düzensiz davranış gösterir. Sağ panelde,  $\alpha= 0.3$  için elastik ağ tüm değişkenleri içerir ve ilişkili gruplar birlikte çekilir. Elastik ağ, LASSO tahmin edicisinin aksine katsayıların yaklaşık olarak birlikte seçildiği ve yaklaşık olarak değerlerinin eşit olarak paylaştırıldığı bir yöntemdir. Elastik ağ, güçlü grup içi korelasyonları otomatik olarak kontrol eder (Hastie vd., 2015).

Elastik ağ yönteminde, LASSO'dan farklı olarak belirlenmesi gereken ek bir ayar parametresi bulunmaktadır. Bu ayar parametresi  $\alpha$  olarak gösterilmektedir. Elastik ağ için ayar parametre değerlerinin belirlenmesi, LASSO yönteminde olduğu gibi çapraz geçerlilik testi kullanılarak belirlenebilir.



Şekil 3.9: Ridge, LASSO ve Elastik Ağ yöntemlerinin Grafiksel Gösterimi (Anonim,2020)

Şekil (3.9)' da en soldaki grafik LASSO'yu, ortadaki Ridge regresyonu en sağdaki grafik ise Elastik ağ yöntemini temsil etmektedir. Bu şekilden Elastik ağın (sağdaki)  $L_1$  ve  $L_2$  normun özelliklerini paylaştığı görülmektedir. Keskin kenarlar ve köşeler değişken seçimi yani LASSO'yu, kavisli kontürler ise katsayıların paylaşımını yani ridge regresyonu göstermektedir (Sneiderman, 2020). Elastik Ağ yönteminin lojistik regresyona uyarlanmış hali olan lojistik Elastik ağ bir sonraki bölümde tanıtılmıştır.

### 3.9. Lojistik Elastik Ağ

Elastik ağ, Ridge ve LASSO cezaları arasında bir orta nokta bulmayı sağlar (Zou ve Hastie 2005). Elastik ağ, değişken seçiminin yanı sıra düzenleme için yeni bir yöntem olarak tanınır. Hem varsayımsal hem de gerçek veriler kullanılarak, Elastik ağın LASSO'ya benzediği gösterilmiştir. Ayrıca Elastik ağ yönteminin özellikle bağımsız değişken sayısının çok fazla olduğu durumlarda avantajlı olduğu literatürde yapılmış çalışmalarda gösterilmiştir (Yaseen vd., 2019).

Yüksek boyutlu veriler ( $p \gg n$ ) ölçülen değişken sayısının verilerdeki gözlem sayısından fazla olduğu durumu ifade eder. Bu, birkaç değişken üzerinde birçok gözlemimizin olduğu istatistiksel analiz için geleneksel veri kümelerinden farklıdır. Bu tür yüksek boyutlu veriler istatistiksel analize yeni zorluklar getirmiştir, çünkü birçok geleneksel istatistiksel yöntem bu veri kümelerine otomatik olarak uygulanmaz

(Pourahmadi, 2013). Yüksek boyutlu ikili sınıflandırma verilerinin cezalandırılmış lojistik regresyon kullanılarak azaltılması, bağımsız değişkenlerle ilişkilendirildiğinde karşılaşılan zorluklardan biridir. Hem katsayıları tahmin etmek hem de değişken seçimi aynı anda gerçekleştirmek için, elastik ağ cezası yüksek boyutlu ikili sınıflandırmada başarıyla uygulanmıştır (Algamal ve Lee, 2015).

Değişken seçiminde LASSO cezasından yararlanırken yüksek korelasyon problemi ile başa çıkmak için de ridge regresyon cezasını kullanarak  $L_2$  normu ve  $L_1$  norm cezaları arasında bir denge kurmaya çalışan lojistik Elastik ağ yöntemi Eşitlik (3.33)' de tanımlanmıştır.

$$\hat{\beta}_{log-elastic} = \sum_{i=1}^n \{y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))\} + \lambda \left[ \frac{1}{2} (1 - \alpha) \|\beta\|_2^2 + \alpha \|\beta\|_1 \right] \quad (3.33)$$

Zou ve Hastie (2005),  $\beta$  tahmin edicilerini bulmak için, verilen veri kümesi  $(\mathbf{y}, \mathbf{X})$  artırılmış bir veriye  $(\mathbf{y}^*, \mathbf{X}^*)$  genişletilir.  $\mathbf{X}^*$ , Eşitlik (3.34)' de verilmiştir.  $\mathbf{y}^*$  ise Eşitlik (3.35)' de verilmiştir.

$$X_{(n+p,p)}^* = (1 + \lambda_2)^{-\frac{1}{2}} \begin{pmatrix} X \\ \sqrt{\lambda_2} I \end{pmatrix} \quad (3.34)$$

$$y_{(n+p,1)}^* = \begin{pmatrix} y \\ 0 \end{pmatrix} \quad (3.35)$$

Veri setini bu şekilde artırmanın bir sonucu olarak, elastik ağ bir LASSO cezası olarak yazılabilir ve çözülebilir. Bu nedenle, elastik ağ, LASSO'daki gibi  $n$  tane bağımsız değişkeni değil, aynı zamanda  $p > n$  olduğunda da yüksek boyuttaki tüm  $p$  bağımsız değişkenleri seçebilir (Zou ve Hastie, 2005).

### 3.10. Sağlam Lojistik Regresyon

İstatistiksel analizlerde bazı gözlemlerin veri setindeki diğer gözlemlere göre çok büyük ya da çok küçük değerler aldığı durumlarla karşılaşabilmektedir. Bu tür gözlemler aykırı değer olarak tanımlanmaktadır. Bu durumda regresyon modelinde her bir gözlemin

ayrıntılı olarak incelenmesi gerekir. Bir tek aykırı değer durumunda dahi kurulan model ya da parametre tahminleri önemli ölçüde etkilenebilmektedir. Ayrıca, aykırı değerler kullanılacak istatistiksel modelin varsayımlarını sağlanamada sorunlar çıkarmaktadır. Model biçimi doğru oluşturulamamakta ve tahmin edilen parametre değerleri gerçeği yansıtmamaktadır. Ayrıca en çok olabilirlik yöntemi ile elde edilen parametre kestirimleri, ideal veri setleri için iyi sonuçlar verirken aykırı değerler bulunması durumunda oldukça hassastır (Hadi ve Imon, 2008). Veri setinde aykırı değerler bulunması durumunda istatistiksel analiz için sağlam yöntemler kullanılabilir. Oluşturulan regresyon modelinde aykırı değerlerin olması durumunda lojistik regresyonda bağımlı değişkenin 0 ve 1 değerlerini alma olasılıkları modellediğinden, modelde elde edilen başarı olasılığı küçük olduğu halde gözlenen durum başarılı, elde edilen başarı olasılığı yüksek olduğu halde gözlenen durumun başarısız olarak belirlenmesine neden olmaktadır (Yavuzkanat, 2011). Lojistik regresyonda aykırı değerlerle karşılaşılması durumunda sağlam lojistik regresyon kullanılabilir.

Bianco ve Yohai (1996) tarafından geliştirilen ve literatürde yaygın bir şekilde kullanılan sağlam lojistik regresyon tahmin edicisi Eşitlik (3.36)' da verilmiştir. Eşitlik (3.36)' da verilen  $\phi$  fonksiyonu Eşitlik (3.37)' da verilmiştir. Eşitlik (3.37)' da ki  $\tilde{n}$  fonksiyonu ise Eşitlik (3.38)' de verilmiştir.

$$\beta_{R-log} = \sum_{i=1}^n \phi(y_i, \pi_i(x'_i \beta)) \quad (3.36)$$

$$\phi(y_i, \pi_i(x'_i \beta)) = \tilde{n}(d_i(\beta)) + \tilde{n}_0(\pi_i(x'_i \beta)) \quad (3.37)$$

$d_i(\beta)$   $\beta$ 'ların varyansını göstermektedir ve  $\tilde{n}(t)$  sınırlandırılmış fonksiyondur.

$$\tilde{n}(0) = 0 \text{ ve } \tilde{n}(t) = (1 - t)I \quad (0 < t < 1) \quad (3.38)$$

Dengeleyici fonksiyon,  $\tilde{n}_0(\pi) = \tilde{n}_1(\pi) + \tilde{n}_1(1 - \pi)$  olmak üzere  $\tilde{n}_1$ ,  $\tilde{n}$ 'ye bağlı olarak;

$$\tilde{n}_1(\pi) = \int_0^\pi \tilde{n}(-\ln t) dt$$

$\pi \in (0,1)$  şeklinde elde edilmektedir (Hobza vd., 2012).

### 3.11. Model Seçimi

Regresyon modelleri arasında bir seçim yapılacağı zaman genellikle belirlilik katsayısının kullanıldığı görülmektedir. Ancak modelin uyumunu gösteren bu ölçünün değişken ve gözlem sayısından etkilendiği bilinmektedir. Bir başka ifade ile aynı sayıda gözlem ve parametreye sahip modeller için kullanışlı bir ölçü olmasına rağmen, aksi durumlarda güvenilir bir ölçü değildir. Yukarıda ele alınan lojistik regresyon modellerinde parametre sayılarının eşit olmaması nedeniyle genellikle model uyumu karşılaştırmaları için çapraz geçerlilik katsayısı kullanılmaktadır. Bu nedenle ilerleyen kısımda çapraz geçerlilik kavramı açıklanacaktır.

#### 3.11.1. $\lambda$ ayar parametresinin seçimi

Çapraz geçerlilik testi, lojistik-ridge regresyonun ayar parametresi  $k$ ' nin, lojistik-LASSO regresyonun ayar parametresi olan  $\lambda$ 'nin ve lojistik Elastik ağ yöntemindeki ayar parametrelerinin farklı değerlerine bağlı olan bir dizi katsayı tahmini üretilir. Bu ayar parametre tahminlerinden en uygun olanı seçmek, her üç yöntem için de kritik bir adımdır. Bu amaçla kullanılan farklı yaklaşımlar bulunmaktadır (Pereira vd., 2015). Bu yöntemler arasında sıklıkla kullanılan Cross-validation (çapraz geçerlilik testi) yöntemi alt kısımda açıklanmıştır.

#### 3.11.2. Çapraz geçerlilik testi (Cross-Validation) (CV)

Çapraz geçerlilik katsayısı, her defasında bir gözlem noktasını dışarıda bırakarak geriye kalan veri noktaları ile uygun bir model oluşturarak, dışarıda bırakılan gözlemler ile ortalama tahmin hatasını minimum yapmaya çalışmaktadır (Fidanoğlu, 2009). Çapraz geçerlilik katsayısı Eşitlik (3.39)' da verilmiştir.

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}^{-i})^2 \quad (3.39)$$

$$\hat{y}^{-i} = x_{i-1}^T \hat{\beta}^{-i}(\lambda)$$

$\hat{\beta}^{-i}(\lambda)$ ,  $(x_i, y_i)$  dışında kalan diğer gözlemlere dayalı modelin tahminidir.

Diğer bir ifadeyle çapraz geçerlilik, bir modelin bağımsız bir veri kümesine ne kadar iyi genelleştirilebileceğini değerlendiren bir yöntem olarak da ifade edilebilir.  $c$ -kat çapraz doğrulama yöntemiyle veriler yaklaşık olarak eşit boyutta  $c$  alt kümelerine bölünür ve alt kümelerden biri doğrulama kümesi haline gelir. Kalan  $c-1$  alt kümeleri eğitim verileri olarak kullanılır. Bu prosedür, her seferinde farklı bir doğrulama kümesiyle  $c$  kez tekrarlanır ve ayar parametrelerinin en iyi tahmin değeri, çapraz geçerlilik katsayısının en küçük olduğu nokta olarak belirlenir (Hastie vd., 2009; Goeman, 2010).

LASSO tahmin edicisi bir optimizasyon problemi olarak Eşitlik (3.40)'de verildiği gibi yazılabilir.

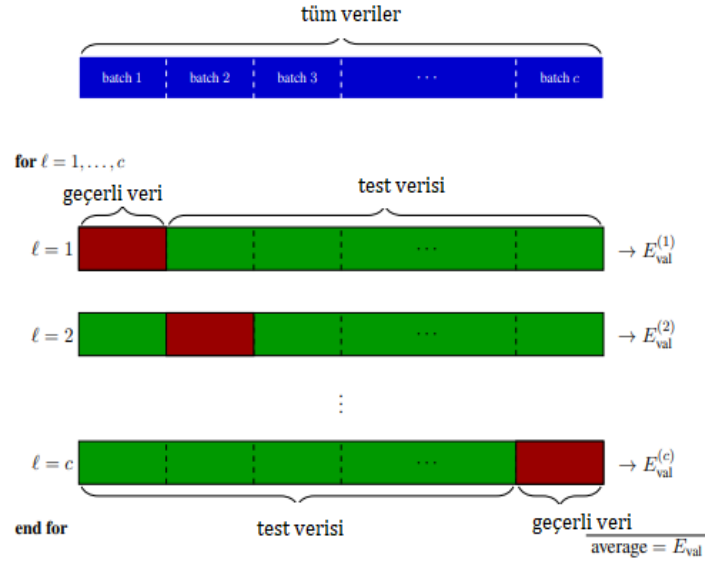
$$\min \left\{ \frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p X_{ij} \beta_j) \right\} \quad (3.40)$$

$$\text{Kısıt } \sum_{j=1}^p |\beta_j| \leq t$$

Burada  $t$  modelin karmaşıklığını kontrol eder.  $t$ 'nin daha büyük değerleri daha fazla parametreyi serbest bırakır ve modelin eğitim verilerine daha yakından uyum sağlamasına izin verir. Tersine daha küçük  $t$  değerleri parametreleri daha fazla kısıtlar ve verilere daha az uyan daha seyrek daha yorumlanabilir modeller oluşturur. Her iki durumda da bir test setindeki tahmin hatası şışecektir. Genellikle bu iki uç nokta arasında iyi bir denge sağlayan  $t$ 'nin bir ara değeri vardır ve bu süreçte bazı katsayıları sıfıra eşit bir model üretir (Tibshirani, 1996).

Çapraz geçerlilik yönteminde,  $t$ 'nin en iyi değerini tahmin etmek için verilen veri setini rasgele böler ve test verilerindeki performansı tahmin eden yapay eğitim ve test setleri oluşturulur. Verileri  $c > 1$  şeklinde grup sayısına bölünür. Tipik olarak  $c$ 'nin seçenekleri 5 veya 10 ve bazen  $c$  olabilir. Bir grubu test seti olarak belirlendikten sonra geri kalan  $c - 1$  grup eğitim seti olarak belirlenir. Daha sonra LASSO eğitim verisine farklı  $t$  değerleri için uygulanır ve test setindeki yanıtları tahmin etmek için her bir uydurulmuş (fitted) modelde

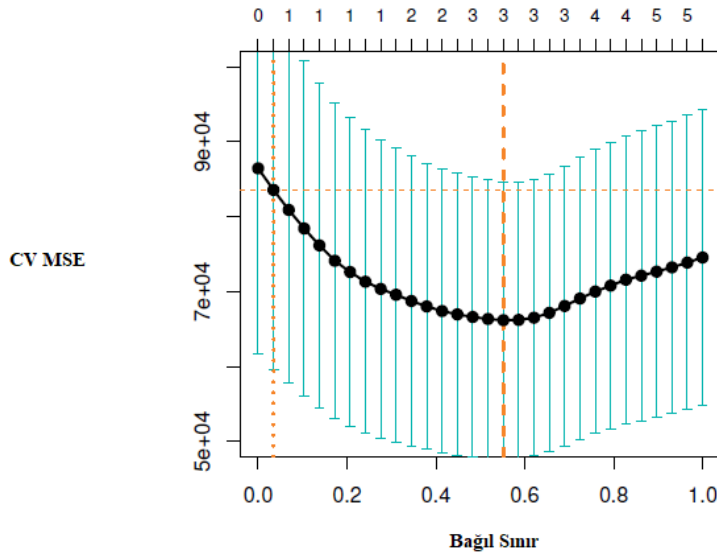
her  $t$  değeri için MSE değeri kaydedilir. Bu süreç toplam  $c$  kez tekrarlanır. Bu şekilde  $t$ 'nin bir değer aralığı üzerinde tahmin hatası için farklı tahminler elde edilir. Bu tahmin hatası tahminlerinin her bir  $t$  değeri için ortalaması alınarak bir çapraz doğrulama hatası eğrisi üretilir (Tibshirani vd., 2015)



Şekil 3.10: c-kat Çapraz Geçerlilik Örneği

Veriler benzer boyutlardaki  $c$  gruplarına bölünmüştür.  $\ell = 1, 2, \dots, c$  üzerinden döngü yaparken  $\ell$  geçerli veriler olarak tutulur ve model kalan  $c-1$  grup üzerinden eğitilir. Her seferinde eğitilmiş model doğrulama verilerinin ortalama  $E_{val}^{(\ell)}$  hatasını hesaplamak için kullanılır. Son model, mevcut tüm veriler kullanılarak eğitilmiştir ve yeni elde edilen bu modelin tahmini değeri  $E_{val}^{(\ell)}$ 'in ortalamasıdır.





Şekil 3.11: Çapraz Geçerlilik Hata Eğrisi (Hastie vd., 2015)

Şekil (3.11) örnek bir veri seti için çapraz geçerlilik hata eğrisini göstermektedir. Burada hata çubukları tahmin hatasının çapraz doğrulanmış tahminlerinde artı ve eksi bir standart hatayı göstermektedir. Dikey kesikli çizgi minimum konumuna çizilirken, kesikli çizgi 'bir standart hata kuralı' seçiminde çizilir (Tibshirani, vd. 2015).

### 3.12. İstatistiksel Çıkarsama

Lockhart vd. (2014), LASSO doğrusal regresyon modeli durumunda, LASSO çözümü için belirli bir aşamada mevcut LASSO modeline giren ortak değişkenin önemini değerlendirmek için kullanılacak bir kovaryans testi istatistiği önermiştir. Bağımsız değişkenler matrisi  $X$  üzerindeki varsayımlar sağlandığında ( $X$ 'in sütunları orthonormal),  $T_k$  olarak belirtilen kovaryans testi istatistiği, LASSO ile elde edilen mevcut tüm gerçek aktif değişkenleri içerdiğine dair sıfır hipotezi altında test edilir.

$k$ , LASSO çözümünün şu andaki çözümünü vermek üzere;  $A$ , mevcut ortak değişkenler kümesi ve  $X_A$ ,  $A$ 'daki  $X$ 'in sütunlarından oluşmak üzere kovaryans test istatistiği Eşitlik (3.41)' de tanımlanmıştır.

$$T_k = (\langle y, X\hat{\beta}(\lambda_{k+1}) \rangle - \langle y, X_A\tilde{\beta}_A(\lambda_{k+1}) \rangle) / \sigma^2 \quad (3.41)$$

İki düğüm arasında tüm  $\lambda$  değerleri için aynı kalan aktif bir  $A$  kümesi bulunmaktadır. Verilen herhangi bir düğüm olan  $\lambda_k$ , belirli bir değişkenin mevcut LASSO aktif değişkenler kümesine girişini veya çıkışını temsil eder. LASSO aktif değişkenler kümesi olan  $A$  kümesini, LASSO çözümünün destek kümesi olarak tanımlanır ve  $\hat{\beta}(\lambda)$  olarak ifade edilir. Buna göre (3.41) denkleminde,  $\langle a, b \rangle$ ; ilk koordinatı  $a$ , ikinci koordinatı  $b$  olan çarpım kümesi elemanını belirtmektedir.  $\hat{\beta}(\lambda_{k+1})$ ,  $A \cup \{j\}$  ortak değişkeni kullanılırken  $\lambda = \lambda_{k+1}$  için LASSO çözümüdür.  $\tilde{\beta}(\lambda_{k+1})$ , aktif değişkenler kullanılırken  $\lambda = \lambda_{k+1}$  için LASSO çözümüdür. Test istatistiği, daha sonraki aşamada iç çarpım kullanılarak elde edilir. (3.41) eşitliği,  $X\hat{\beta}$  ile  $X_A\tilde{\beta}_A$  arasındaki farkın fonksiyonu olarak da ifade edilir. Daha basit ifade ile Eşitlik (3.42) minimize edilir.

$$\tilde{\beta}_A(\lambda_{k+1}) = \underset{\beta_A \in R^{|A|}}{\operatorname{argmin}} \frac{1}{2} \|y - X_A\beta_A\|_2^2 + \lambda_{k+1} \|\beta_A\|_1 \quad (3.42)$$

(3.42) eşitliğinde,  $\sigma^2$  değerinin bilindiği varsayılmaktadır. Uygulamalarda genelde  $\sigma^2$  bilinmez.  $\sigma^2$  tahmini için genel olarak,  $n$  örneklem hacmi ve  $p$  bağımsız değişken sayısını göstermek üzere,  $n > p$  olmak üzere, ortalama hata karesi kullanılarak  $\sigma^2$  tahmin edilir. Eğer  $p \geq n$  ise, çapraz doğrulama ile elde edilen model uyumuna göre, en küçük kareler yöntemi ile elde edilen  $\sigma^2$  değeri kullanılmaktadır (Qui, 2017).

eNetXplorer, ridge ( $\alpha = 0$ ) den LASSO ( $\alpha = 1$ ) ya kadar birden fazla  $\alpha$  değeri için, elastik ağ modelleri ailesi oluşturur. Algoritma, kısaca üç ana modülden oluşur: (a) model oluşturma, (b) boş model oluşturma ve (c) her  $\alpha$  değeri için sıralı olarak model ve boş modelin karşılaştırmasının yapılması, sonuçların analizi ve görselleştirmesidir.

Model oluşturma algoritmasında, tüm veriler kullanılarak bir dizi  $n_\lambda$  değeri elde edilir.  $n_\lambda$  değerinden bağımsız olarak, log ölçeğini sabit tutarak,  $\lambda$  aralığını genişleterek bir değer belirleyebilir. Her bir  $\lambda$  için, elastik net çapraz doğrulama modelleri oluşturulur. Sonraki aşamada  $\lambda^*$ , bir kalite fonksiyonunun maksimize edilmesiyle seçilir. Doğrusal regresyon için kalite fonksiyonu olarak korelasyon seçilir. Kalite fonksiyonlarının seçimini analiz yapan kişi de belirleyebilir.

eNetXplorer'ın önemli bir özelliği, elastik ağ model ailesinin her bir ( $\alpha$ 'ya özgü) üyesiyle ilişkili bir boş modeller topluluğunun oluşturulmasıdır. Her bir  $n_r$  çalışması makine öğrenmesi yöntemi ile düğümlere atanır ve  $n_p$  boş modeli sayesinde bağımlı değişken değerleri ile karşılaştırılır. Bu işlem için yukarıda bahsedildiği gibi kalite fonksiyonlarından yararlanılır.

Modelin istatistiksel anlamlılığı ise, model ve boş model için kalite fonksiyonlarının arasındaki farka dayalı olarak elde edilir. Bu durum Eşitlik (3.43)' de verilmiştir.

$$p_{val} = \frac{1}{1+n_r n_p} \left\{ 1 + \sum_{r=1}^{n_r} \sum_{p=1}^{n_p} \Theta(KF_{boş}^{r,p} - KF_{model}^r) \right\} \quad (3.43)$$

Burada,  $\Theta$ ; sağdan sürekli Heaviside adım fonksiyonudur.  $p_{val}$  nümerik hesaplamalar ile bulunur. Yukarıdaki algortima lojistik regresyon modeli için de genişletilmiştir. R programındaki kütüphane kullanılarak Cov testi ve katsayıların anlamlılığı gerçekleştirilir.

#### 4. UYGULAMA

Su, canlıların bütün biyolojik yaşamı ve bütün insan faaliyetlerinde ihtiyaçtır (Bulut ve Kubilay, 2019). Su, yaşam ortamının oluşmasında temel öğelerden biri olduğu gibi aynı zamanda kendisi de bir yaşam ortamıdır. Yaşam için olmazsa olmaz ön koşullardan biri olduğu için, yaşam ortamında bulunması ve kalitesi de büyük önem taşımaktadır. Yeryüzünün  $\frac{3}{4}$ 'ünün sularla kaplı olmasına karşın içilebilir nitelikteki su oranı ancak % 0,74 civarındadır. (Bulut ve Kubilay, 2019; Yanık vd., 2001; Akın ve Akın, 2007).

Dere, çay ve nehir gibi akarsular, başlıca tatlı su kaynaklarıdır. Bu kaynaklardan elde edilen sular içme suyu, evlerde ve sanayide kullanma suyu ve tarımda sulama suyu olarak kullanılır. Ayrıca, akarsular enerji üretiminde de önemli bir yere sahiptir (Hacıoğlu ve Dulger 2009). Diğer yandan, akarsular çevre kirliliğinden en çok etkilenen ekosistemlerdir (Zeydan vd, 2019).

Günümüzde yerüstü su kaynaklarının sahip oldukları su kalitesinin belirlenmesi ve buna bağlı oluşturulan su kalite yönetimi tüm dünyada önemli araştırmalar arasında ciddi bir yer tutmaktadır. Su kalitesi ile ilgili yapılan çalışmalar, suyun kullanım amacının belirlenmesinde çok önemli olmasının yanı sıra, yüzey su kaynaklarının sürdürülebilir kullanımı açısından da önem arz etmektedir. Bu nedenle yüzey su kaynaklarının su kalitesinin sürekli ve düzenli izlenmesi gerekmektedir (Gölbaşı ve Şen, 2019).

Su kaynakları, endüstriyel ve tarımsal faaliyetler sonucu özellikle alıcı ortam olarak kullanılmaları nedeniyle son yıllarda kirlilik tehdidi ile karşı karşıya kalmaktadır. Sürdürülebilir bir su yönetimi ve bu kaynakların içme-sulama maksatlı kullanımı için su kütlelerinde meydana gelen fiziksel ve kimyasal değişimlerin sürekli izlenmesi ve değerlendirilmesi önem arz etmektedir (Özer ve Köklü, 2019).

Doğduğu noktadan itibaren birçok ilden geçen Sakarya Nehri, Aşağı Sakarya Alt Havzasına gelene kadar arıtma tesislerine sahip olmayan sanayi faaliyetlerinden ve yerleşim yerlerinden kaynaklanan atık suların, tarım faaliyetleri sonucunda oluşan kirlenmeden, kent içi küçük sanayilerden, yağmur sularının taşımış olduğu kirliliklerden ve geçtiği jeolojik formasyonlardan etkilenmektedir. Sakarya nehri tüm bu kirletici yüklerine rağmen

aynı zamanda geçtiği bölgelerde çiftçiler tarafından sulama suyu olarak da kullanılmaktadır. (Özer ve Köklü, 2019).

Bu bölümde, önceki bölümde teorik bilgileri tanıtılan Lojistik Regresyon, Robust Lojistik Regresyon, Lojistik Ridge Regresyon, Lojistik-LASSO ve Lojistik-Elastik Ağ yöntemlerini karşılaştırmak üzere Sakarya Nehri üzerinde 3 farklı bölgeden 8 adet istasyon seçilmiş ve bu istasyonlara ait veriler Devlet Su İşleri Genel Müdürlüğü'nden temin edilmiştir. 2014-2018 yılları arasında ilgili istasyonlarda su kalitesine ait 17 değişkenin düzenli ölçümü gerçekleştirilmiştir. Düzenli ölçümü yapılamayan değişkenler çalışmada yer almamaktadır. 2014-2018 yılları arasında bu istasyonlarda ilgili değişkenler için toplam 384 adet gözlem mevcuttur. Çalışmada modellerin doğru sınıflandırma yüzdelerine dayalı olarak çapraz geçerlilik sonuçlarına yer verilecektir. Bu amaçla 384 verinin 288 tanesi (%75'i) eğitim verisi olarak belirlenmiş, kalan 96 veri (%25'i) test verisi olarak ayrılmıştır. Eğitim verisi modelin belirlenmesi ve parametre tahmini aşamasında kullanılmaktadır. Model kurulduktan sonra test verisi ile modelin doğruluğu tespit edilmektedir. Çapraz geçerlilik yöntemi kullanılarak gerçekleştirilecek olan model doğruluğunun belirlenmesi aşamasında verinin %5-%33'ünün test verisi olarak ayrılması genel kabul görmüş bir orandır. Eğitim verisi kullanılarak oluşturulan modellerin genel anlamlılık sınavında lojistik regresyon, lojistik ridge regresyon ve robust lojistik regresyon için Hosmer-Lemeshow testi kullanılmıştır. Parametrelerin anlamlılık sınamaları için ise kovaryans testinden yararlanılmıştır. Uygulama sonuçlarına geçmeden önce su kalite karakteristikleri hakkında bilgi verilmiştir.

#### **4.1. Su Parametreleri Hakkında Genel Bilgiler**

Bu bölümde su kalitesi karakteristikleri hakkında bilgi verilmiştir.

##### **4.1.1. Sıcaklık (T)**

Sıcaklık, göllerde tabakalaşmanın belirlenmesinde ölçülmesi gereken en önemli parametredir. Sıcaklık su kaynağındaki biyolojik, kimyasal ve fiziksel işlemleri etkiler. Böylece pek çok parametrenin konsantrasyonu değişir. Suyun sıcaklığı arttığında, kimyasal

reaksiyonların hızı ve sudaki maddelerin buharlaşması da artar. Suyun sıcaklığının artması ayrıca O<sub>2</sub>, CO<sub>2</sub>, N<sub>2</sub>, CH<sub>4</sub> gibi gazların suda çözünürlüğünü azaltır. Sucul organizmaların metabolik hızı, sıcaklığa bağlıdır. Sıcak sularda organizmaların solunum hızının artması oksijen tüketimini artırır ve organik maddelerin bozunmasına neden olur (DSİ, 2001).

#### **4.1.2. Ph**

Su içindeki hidrojen iyonu konsantrasyonunu 10 tabanına göre negatif logaritması Ph değeri olarak tanımlanmaktadır. Ph=7 olan sular Nötr sular olarak bilinir. İçme sularındaki Ph değeri 6.5-8.5 arasında uygun görülmektedir. Yeraltı sularındaki Ph değeri, çözülmüş karbondioksit ve diğer karbonat ve bikarbonat bileşikleri arasındaki dengeye bağlı olarak değişmektedir. Bu denge, sıcaklık ve basınç değişmelerine göre kolayca değişim göstermektedir (Güler, 1997).

#### **4.1.3. Çözülmüş oksijen (DO)**

Çözülmüş oksijen, su kirlenmesi ile ilgili en önemli parametrelerden birisidir. Doğal sularda oksijen, azot ve karbondioksit gibi gazlar da erimiş halde bulunurlar. Oksijenin suda erime derecesi, suyun sıcaklık ve tuzluluk derecesine bağlıdır. Sıcaklık yükseldikçe, suda daha az oksijenin eridiği görülmektedir. Ölçülen çözülmüş oksijen konsantrasyonu; suyun kirlenme derecesini, sudaki organik madde konsantrasyonunu ve suyun kendi kendini ne derece temizleyebileceğini ifade eder (Çoban, 2007). Doğal tatlı suyun çözülmüş oksijen kapsamı, 0-300C sıcaklıklar arasında 5–15 mgL-1 düzeyleri arasında değişir.

#### **4.1.4. Elektriksel iletkenlik (EC)**

Elektriksel iletkenlik (EC), suyun elektrik akımını iletebilmesinin bir ölçüsüdür ve sularda mineral asitler olmak üzere çözülmüş katılardaki değişimi ifade eder. Suların iletkenliği, sulardaki iyon sayısı hakkında bilgi verir (Uslu ve Türkmen, 1987).

#### **4.1.5. Sülfat iyonu (SO<sub>4</sub>)**

Sülfatlar doğada bulunan ağır metal sülfürlerinin atmosferik olayların etkisiyle kısmen oksitlenerek suda çözünmesinden oluşmuşlardır. Suda yüksek sülfat konsantrasyonu dağıtma sistemlerindeki, özellikle düşük alkalinite olduğu zaman, metallerin korozyonuna neden olmaktadır (Güler, 1997).

#### **4.1.6. Amonyum azotu (NH<sub>4</sub>-N)**

Azot bileşikleri, su kirliliği açısından çeşitli etkilere sahiptir. En önemli etkileri; oksijen miktarının etkilenmesi, içme suyunun sağlıklı bir şekilde temininde oluşan hijyenik ve toksikolojik problemlerdir (Uslu ve Türkmen 1987). Amonyak yüzeysel sularda doğal olarak belirli bir miktar bulunmaktadır. Ancak evsel ve endüstriyel atık suların deşarjı ile tarımda kullanılan aşırı gübrelerin yüzey sularına karışması nedeniyle su içerisindeki oranı artmaktadır (Chapman ve Kimstach 1996).

#### **4.1.7 Biyolojik oksijen ihtiyacı (BOD)**

BOD pis sudaki organik maddelerin aerobik koşullar altında oksidasyonu ve mineralizasyonu (stabilizasyonu) için bakteriler tarafından tüketilen oksijen miktarıdır. Organik maddeler bakteriler için besin maddesidir. BOD deneyi, kanalizasyon ve sanayi artıklarının kirlenme derecesini, gerekli oksijen miktarı cinsinden tayinde kullanılır (Güler, 1997).

#### **4.1.8. Magnezyum (Mg<sup>++</sup>)**

Magnezyum, suyun sertliğini meydana getiren iyonlardan biridir. Sıcak sularda kırılğan bir kabuk meydana getirir. İnsan biyolojisinde gerekli bir mineraldir. Kemik, kas ve sinirsel dokularda bulunur. Yetişkin bir insanın günlük 35 mg magnezyuma gereksinimi

vardır. Suda bulunan karbondioksit, karbonatlı ve sislkatlı minerallerdeki magnezyumun suya geçmesinde rol oynar (Güler, 1997).

#### **4.1.9. Klor (Cl<sup>-</sup>)**

Yüzeysel sulara klorür, doğal olarak sedimentteki kayaların yıpranması ile yapay olarak ise endüstriyel ve evsel atık sular, tarımsal drenaj suları ve çeşitli amaçlarla kullanılan fazla miktardaki tuzların alıcı ortamlara taşınmasıyla girmektedir. Klorür, klorun yükseltgenme hali olup, su ve atık sularda rastlanan en önemli anorganik anyonlardandır.

#### **4.1.10. Sodyum (Na<sup>+</sup>)**

Yer kabuğunda en fazla bulunan elementlerden biri sodyum'dur. Sularda çok kolaylıkla çözünen sodyum bileşikleri suya tuzlu bir tat verir. Aşırı derecede sodyum elementi içeren sular tarımsal sulamalarda toprak ve bitkiler üzerinde zararlı etkiler yapar (Tuncay, 1983). Sodyum değerinin yüksek olması toprağın geçirgenliğini de etkilemektedir (Ertan, 2019).

#### **4.1.11. Toplam sertlik ve toplam alkalinite**

Suların sertliği, su içerisinde çözülmüş olarak bulunan toplam Ca<sup>++</sup> ve Mg<sup>++</sup> miktarının kalsiyum karbonat (CaCO<sub>3</sub>) eşdeğeri olarak tanımlanmaktadır. Bir suyun sertliği, o suyun temas etmiş olduğu topraklardaki minerallerin suda çözünmesine bağlı olarak değişiklik gösterebilmektedir (Şengün,2013).

Suyun Ph nötröle etme kapasitesi anlamına gelen alkalinite ise su içerisindeki hidroksit (OH<sup>-</sup>), karbonat (CO<sub>3</sub><sup>-2</sup>) ve bikarbonat (HCO<sub>3</sub><sup>-</sup>) iyonlarından kaynaklanmaktadır. Toplam alkalinitenin Ph üzerine tamponlayıcı etkisi bulunmaktadır. Alkalinitesi düşük olan suların Ph değişimlerine karşı dirençleri zayıftır. Böyle sularda Ph dalgalanmaları sık olmakta ve sucul canlılara zarar verebilmektedir (Tepe, 2009). Yüksek alkali sularda ise gün içerisindeki Ph değişim aralığı küçük değerlerde gözlenmektedir.



#### **4.1.12. İletkenlik ve tuzluluk**

Tuzluluk, suda erimiş toplam tuz miktarı olarak ifade edilmektedir. Sulardaki tuzluluk klorürden, esas olarak da sodyum klorürden kaynaklanır (Cirik ve Gökpınar, 1993). İç sulardaki düşük tuzluluk derecesi canlıların yaşamını etkilemektedir. Suda çözünen tuzlarının toplam konsantrasyonunun ölçüsü, iletkenlik ile ifade edilmektedir. Tuzlar suda elektrik iletkenliğine sahip yüklü iyonlar halinde çözünmektedirler. Sudaki iyon sayısı ne kadar fazla ise elektriksel kondüktivitesi o kadar yüksek olmaktadır. Elektriksel kondüktivite (iletkenlik) (EC), çözültideki atık madde miktarını ve tuzlulukla ilişkisini yaklaşık olarak gösterir.

#### **4.1.13. Toplam çözünmüş madde (TDS)**

Sudaki çözünmüş maddeler, suda çözünmüş az miktardaki organik madde ve anorganik tuzların varlığından ileri gelir. Toplam çözünmüş maddeler doğal kaynaklardan, lağım atıklarından, şehir drenaj sularından ve endüstriyel sulardan ileri gelmektedir.

#### **4.2. Çalışma Alanı**

Sakarya Nehri, Eskişehir'in Çifteler ilçesi Dikmen Mahallesi mevkiinde Seydi Çayı ile Dedemözü Deresi'nin birleşmesiyle oluşur. Sakarya Nehri, 847 metre rakımda Eskişehir, Çifteler üzerinde doğup Eskişehir, Ankara, Bilecik ve Sakarya şehirlerini geçerek Karadeniz'e dökülür. Sakarya Nehri, Kızılırmak ve Fırat nehirlerinden sonra Türkiye'nin üçüncü en uzun, Kuzeybatı Anadolu'nun ise en büyük akarsuyudur. Eskişehir'in Çifteler ilçesi yakınlarından doğan Sakarya Nehri kolları ile birlikte toplam uzunluğu 824 km.'dir (Anonim,2019). Doğduğu noktadan itibaren birçok ilden geçen Sakarya Nehri, Aşağı Sakarya Alt Havzasına gelene kadar arıtma tesislerine sahip olmayan sanayi faaliyetlerinden ve yerleşim yerlerinden kaynaklanan atık sulardan, tarım faaliyetleri sonucunda oluşan yayılı kaynaklardan, kent içi küçük sanayilerden, yağmur sularının taşımış olduğu kirliliklerden ve geçtiği jeolojik formasyonlardan etkilenmektedir. Sakarya nehri tüm bu kirleticili yüklerine

rağmen aynı zamanda geçtiği bölgelerde çiftçiler tarafından sulama suyu olarak da kullanılmaktadır (Özer ve Köklü, 2019).

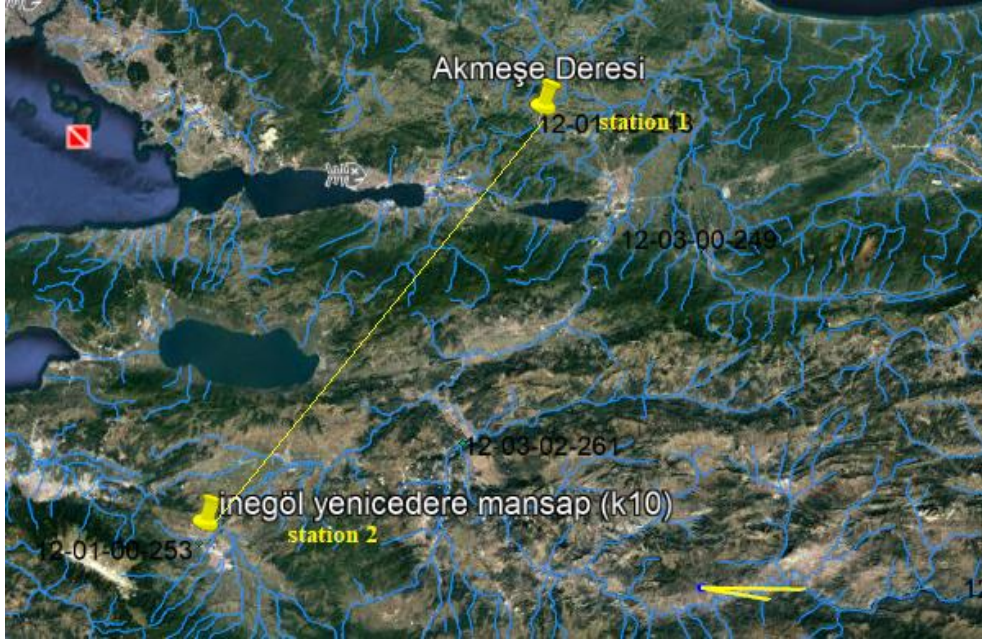
Yapılan çalışmada, birinci bölge olarak Akmeşe Deresi ve İnegöl Yenicedere Mansap (K10) istasyonu; ikinci bölge olarak adlandırılan bölge için DSİ tarafından 3. Bölge olarak belirlenmiş bölgeden seçilen istasyonlar; Mudurnu Çayı Dere Mahallesi Köprü, Sakarya Nehri Sarıyar Baraj çıkışı ve Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonu ve üçüncü bölge olarak adlandırdığımız bölge için DSİ tarafından 5. Bölge olarak belirlenmiş bölgeden seçilen istasyonlar; Çubuk çayı Çubuk 1 baraj çıkışı, Kuruçay Baraj Aksı ve Bayındır Deresi Bayındır Barajı çıkışı istasyonlarına ait 2014-2018 yıllarına ait su kalite verileri DSİ'den alınmıştır. Bu yıllar arasında, her istasyonda düzenli olarak ölçülen su kalite karakteristikleri şu şekildedir: BOD5 (Biyolojik Oksijen İhtiyacı-mg/L), Ca<sup>++</sup> (Kalsiyum-mg/L), Cl<sup>-</sup> (Klorür-mg/L), DO (Çözünmüş Oksijen- mgO<sub>2</sub> /L), EC (Elektriksel İletkenlik-Mikromhos/cm), K<sup>+</sup> (Potasyum-mg/L), M-Al (mg/L), Mg<sup>++</sup> (Magnezyum-mg/L), Na<sup>+</sup>(Sodyum-mg/L), NH<sub>4</sub>N (Amonyum Azotu-mg/L), O-PO<sub>4</sub> (Orta Fosfat-mg/L), Ph (Alkalinite), Pv (Permanganat Değeri-mgO<sub>2</sub>/L), SO<sub>4</sub> (Sülfat-mg/L), T (Sıcaklık-°C), TDS (Toplam Çözünmüş Katılar-mg/L), TH (Toplam Sertlik-mg/LcaCO<sub>3</sub>). Bu seçilen değişkenler, analizde kullanılan bağımsız değişkenlerdir.

Bağımlı değişken ise,

$$y_i = \begin{cases} 1 & , \quad \text{Akmeşe deresi} \\ 0 & , \quad \text{İnegöl Yenicedere Mansap} \end{cases}$$

şeklinde kodlanarak lojistik regresyon modelleri belirlenmeye çalışılmıştır.

Bu istasyonların konumuna ilişkin harita Şekil 4.1' de verilmiştir. Seçilen istasyonlarla yapılan ikili karşılaştırma için korelasyon matrisi hesaplanmış olup Çizelge 4.1' de verilmiştir.



Şekil 4.1: Akmeşe Deresi ve İnegöl Yenicedere Mansap istasyonlarının konumuna ilişkin harita

Benzer şekilde, ikinci bölgedeki istasyonlar ikili lojistik regresyon analizi uygulanacağı için ikiyeşerli olarak gruplandırılmıştır. İlk grup, Mudurnu Çayı Dere Mahallesi Köprü ve Sakarya Nehri Sarıyar Baraj çıkışıdır. Mudurnu Çayı Dere Mahallesi Köprü  $y_i = 1$  ve Sakarya Nehri Sarıyar Baraj çıkışı  $y_i = 0$  şeklinde kodlanarak iki istasyon arasındaki değişime neden olan değişkenler belirlenmeye çalışılmıştır. İkinci grup istasyonlardan Mudurnu Çayı Dere Mahallesi Köprü ve Kocaçay Deresi Kayaboğazı Baraj çıkışıdır. Mudurnu Çayı Dere Mahallesi Köprü  $y_i = 1$  ve Kocaçay Deresi Kayaboğazı Baraj çıkışı  $y_i = 0$  şeklinde kodlanarak iki istasyon arasındaki değişime neden olan değişkenler belirlenmeye çalışılmıştır. Üçüncü grup ise Sakarya Nehri Sarıyar Baraj çıkışı ve Kocaçay Deresi Kayaboğazı Baraj çıkışıdır. Sakarya Nehri Sarıyar Baraj çıkışı  $y_i = 1$  ve Kocaçay Deresi Kayaboğazı Baraj çıkışı  $y_i = 0$  şeklinde kodlanarak iki istasyon arasındaki değişime neden olan değişkenler belirlenmeye çalışılmıştır. İstasyonların konumlarına ilişkin harita Şekil 4.1’de verilmiştir.



Şekil 4.2: Mudurnu Çayı Dere Mahallesi Köprü, Sakarya Nehri Sarıyar Baraj çıkışı ve Kocaçay Deresi Kayaboğazi Baraj çıkışı İstasyonlarının Konumları

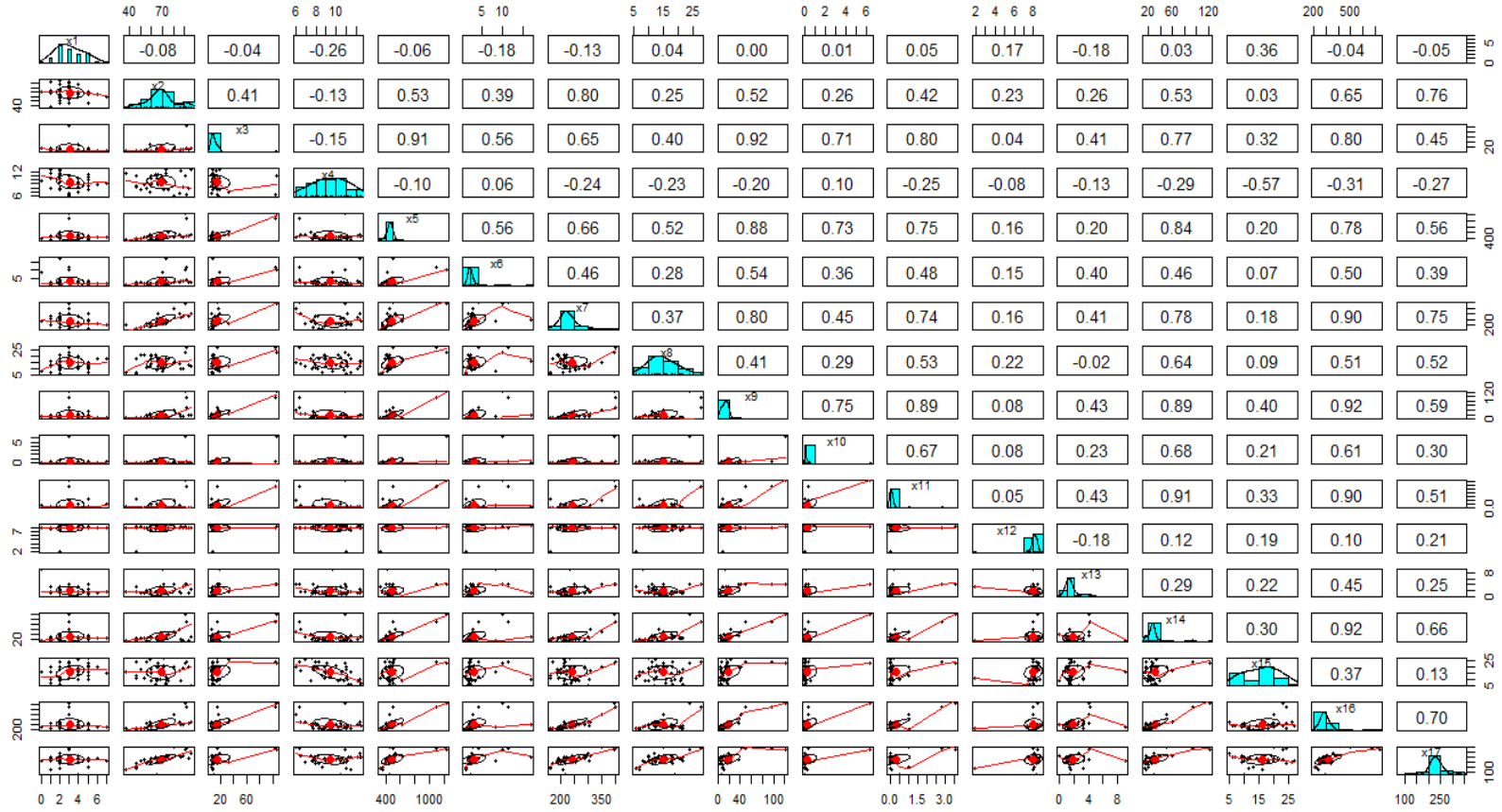
Üçüncü bölge olarak adlandırdığımız istasyon için DSİ tarafından 5. Bölge olarak adlandırılan bölgede yer almaktadır. Çubuk çayı Çubuk 1 baraj çıkışı, Kuruçay Baraj Aksı ve Bayındır Deresi Bayındır Barajı çıkışıdır ve bu istasyonlar ikişerli olarak gruplandırılarak iki istasyon arasındaki değişime neden olan değişkenler belirlenmeye çalışılmıştır. İlk grup Çubuk çayı Çubuk 1 baraj çıkışı ve Kuruçay Baraj Aksıdır. Çubuk Çayı Çubuk 1 Baraj çıkışı  $y_i = 0$  ve Kuruçay Baraj Aksı  $y_i = 1$  şeklinde kodlanmıştır. İkinci grup Çubuk çayı Çubuk 1 Baraj çıkışı ve Bayındır Deresi Bayındır Barajı çıkışıdır. Çubuk çayı Çubuk baraj çıkışı  $y_i = 0$  ve Bayındır Deresi Bayındır Barajı çıkışı  $y_i = 0$  şeklinde kodlanmıştır. Üçüncü grup ise, Kuruçay Baraj Aksı ve Bayındır Deresi Bayındır Barajı çıkışıdır. Kuruçay Baraj Aksı  $y_i = 0$  ve Bayındır Deresi Bayındır Barajı çıkışı  $y_i = 1$  şeklinde kodlanmıştır. Bu istasyonların konumlarına ilişkin harita görüntüsü Şekil 4.2' de verilmiştir.



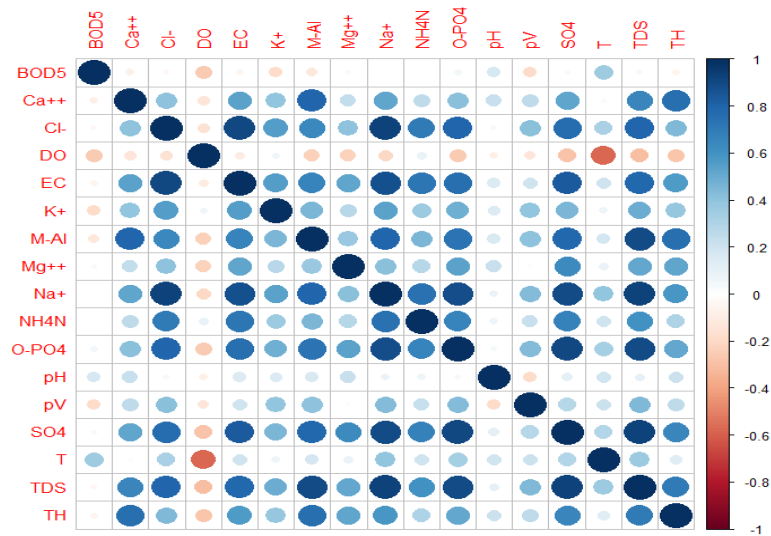
Şekil 4.3: Çubuk çayı Çubuk 1 baraj çıkışı, Kuruçay Baraj Aksı ve Bayındır Deresi Bayındır Barajı çıkışı istasyonlarının konumlarına ait harita

Hakkında bilgi verilen istasyonlar için her bölgeden seçilen istasyonlar ikişerli üçerli gruplandırılmıştır. İlk olarak 1. Bölgeden seçilen istasyonlardan Akmeşe Deresi ve İnegöl Yenicedere Mansap istasyonları için korelasyon matrisi hesaplanmış olup Çizelge 4.1' de verilmiştir. Üçüncü bölgeden seçilen Mudurnu Çayı Dere Mahallesi Köprü, Sakarya Nehri Sarıyar Baraj çıkışı ve Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonları ikişerli gruplandırılmıştır. İlk grupta, Mudurnu Çayı Dere Mahallesi Köprü ve Sakarya Nehri Sarıyar Baraj çıkışı istasyonları ikişerli gruplandırılmıştır. Korelasyon matrisi Çizelge 4.2' de verilmiştir. Mudurnu Çayı Dere Mahallesi Köprü ve Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonları için korelasyon matrisi Çizelge 4.3' de verilmiştir. Sakarya Nehri Sarıyar Baraj çıkışı ve Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonları ikili gruplandırılmış ve korelasyon matrisi Çizelge 4.4' de verilmiştir. Beşinci bölgeden seçilen Çubuk çayı Çubuk 1 baraj çıkışı, Kuruçay Baraj Aksı ve Bayındır Deresi Bayındır Barajı çıkışı istasyonları ikişerli gruplandırılmış ve korelasyon matrisleri hesaplanmıştır. Çubuk çayı Çubuk 1 baraj çıkışı ve Kuruçay Baraj Aksı istasyonları için korelasyon matrisi Çizelge 4.5' de verilmiştir. Çubuk çayı Çubuk 1 baraj çıkışı ve ve Bayındır Deresi Bayındır Barajı çıkışı istasyonları için korelasyon matrisi Çizelge 4.6' da verilmiştir. Kuruçay Baraj Aksı ve Bayındır Deresi Bayındır Barajı çıkışı istasyonu için korelasyon matrisi Çizelge 4.7' de verilmiştir.

Çizelge 4.1. Akmeşe Deresi ve İnegöl Yenicedere Mansap İstasyonu için Korelasyon Matrisi Tablosu

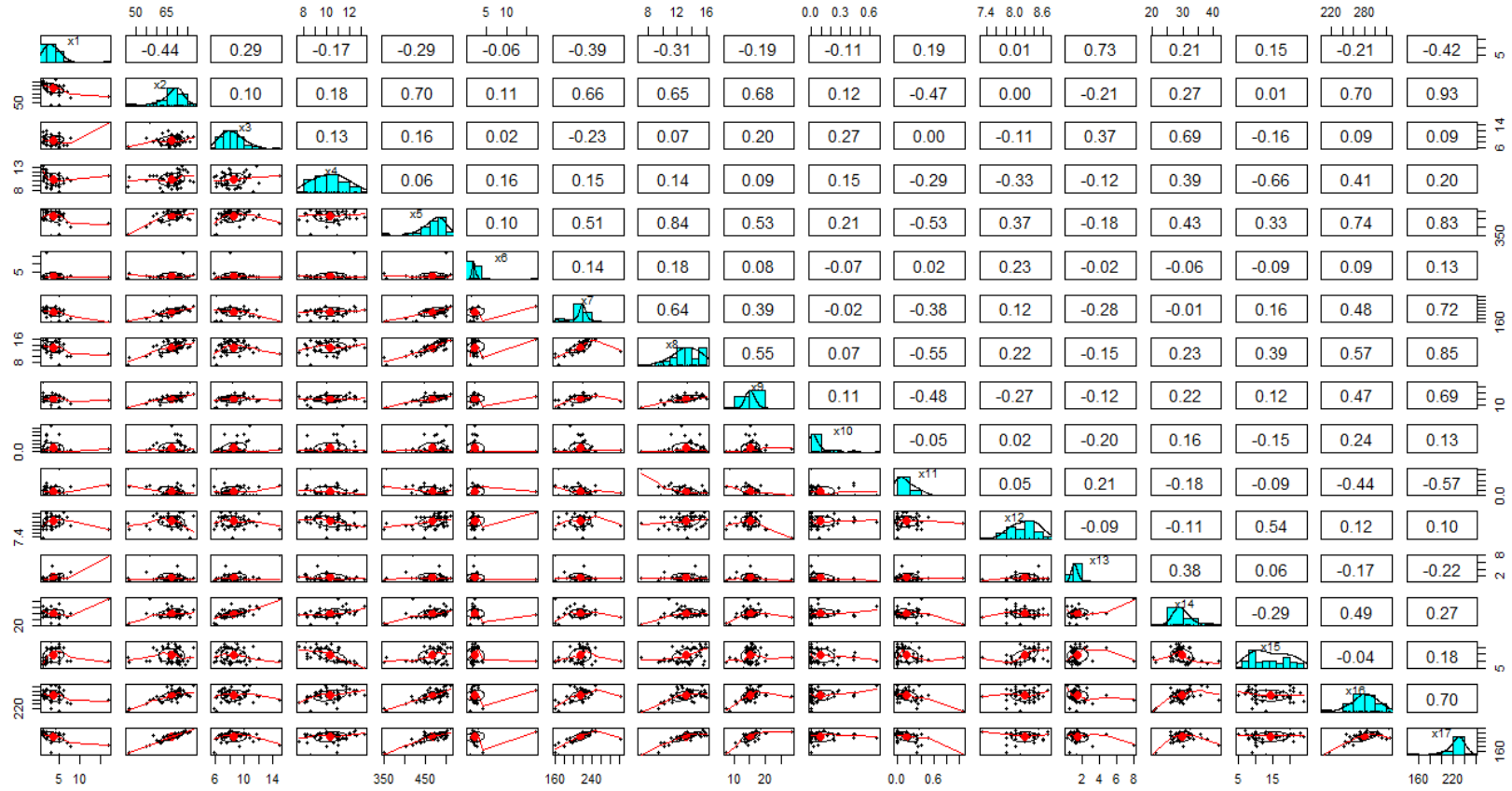


Çizelge 4.1 incelendiğinde, Ca<sup>+</sup> ile BOD<sub>5</sub> ve M-Al arasında 0,80'lik bir korelasyon söz konusudur. Bu durumda Ca<sup>+</sup> ile BOD<sub>5</sub> ve Ca<sup>+</sup> ile M-Al arasında yüksek düzeyde korelasyon olduğu tespit edilmiştir. Cl<sup>-</sup> ile EC arasında 0,91'lik korelasyon, Cl<sup>-</sup> ile Na<sup>+</sup> arasında 0,92'lik korelasyon ve Cl<sup>-</sup> ile O-PO<sub>4</sub> ve TDS arasında 0,80' lik bir korelasyon söz konusudur. Bu da Cl<sup>-</sup> ile EC, Na<sup>+</sup>, O-PO<sub>4</sub> ve TDS arasında yüksek düzeyde korelasyon olduğunu göstermektedir. EC'nin SO<sub>4</sub> ile arasında 0,84'lük ve Cl<sup>-</sup> ile arasında 0,88'lik bir korelasyon söz konusudur. Bu da EC ile SO<sub>4</sub> ve Cl<sup>-</sup> arasında yüksek düzeyde korelasyon olduğu belirlenmiştir. M-Al ile Ca<sup>+</sup> ve Na<sup>+</sup> arasında 0,80'lik korelasyon söz konusudur. Aynı zamanda M-Al ile TDS arasında 0,90'lık bir korelasyon vardır. Bu durumda M-Al ile Ca<sup>+</sup>, Na<sup>+</sup> ve TDS arasında yüksek düzeyde korelasyon vardır. Na<sup>+</sup> ile SO<sub>4</sub> ve O-PO<sub>4</sub> arasında 0,89' lık korelasyon ve Na<sup>+</sup> ile TDS arasında 0,92'lik korelasyon söz konusudur. Bu durumda Na<sup>+</sup> ile SO<sub>4</sub>, O-PO<sub>4</sub> arasında yüksek düzeyde korelasyon vardır denilmektedir. O-PO<sub>4</sub> ile Cl<sup>-</sup> arasında 0,80'lik bir korelasyon söz konusudur. Aynı şekilde O-PO<sub>4</sub>'ın SO<sub>4</sub> arasında 0,91'lik ve TDS ile arasında 0,90'lık bir korelasyon söz konusudur. Bu durumda O-PO<sub>4</sub>'ün Cl<sup>-</sup>, SO<sub>4</sub> ve TDS değişkenleri arasında yüksek düzeyde korelasyon söylenir. SO<sub>4</sub> ile TDS arasında 0,92'lik korelasyon ve TDS ile Cl<sup>-</sup> arasında 0,80'lik korelasyona sahip olduğu söylenebilir. Bu durumda SO<sub>4</sub> ile TDS arasında ve TDS ile Cl<sup>-</sup> arasında yüksek düzeyde korelasyon bulunmaktadır. Özetle, bağımsız değişkenler arasında yüksek düzeyde çoklu iç ilişkiden söz edilir. Şekil 4.4' de korelasyon grafiği verilmiştir.



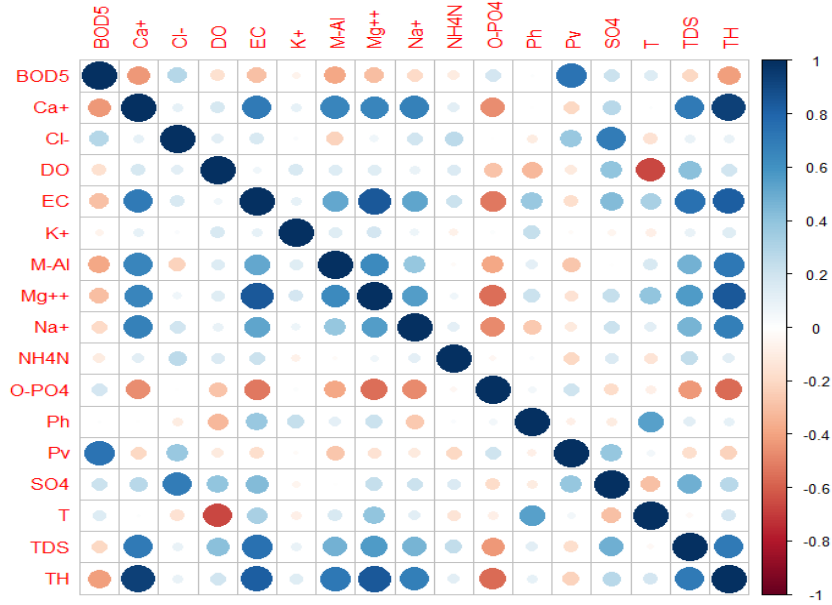
Şekil 4.4: Birinci Karşılaştırma için Korelasyon Grafiği

Çizelge 4.2: Mudurnu Çayı Dere Mahallesi Köprü ve Sakarya Nehri Sarıyar Baraj çıkışı İstasyonları için Korelasyon Matrisi



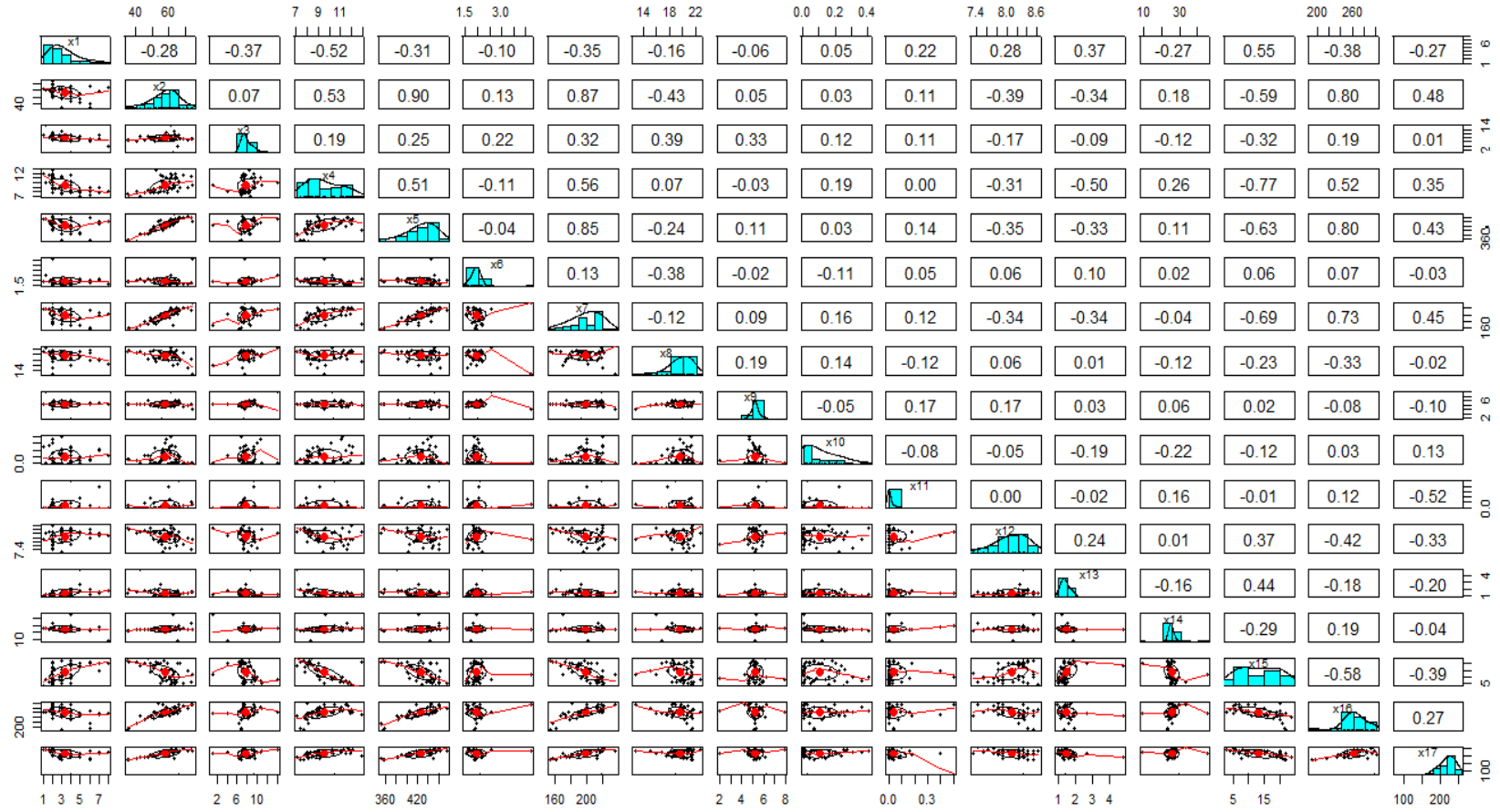


Çizelge 4.2 incelendiğinde, Ca<sup>+</sup> ile TH arasında 0,93'lük bir korelasyon söz konusudur. Bu durumda bu değişkenler arasında yüksek düzeyde korelasyon tespit edildiği ispat edilmiştir. EC ile Mg<sup>++</sup> arasında 0,84'lük bir korelasyon söz konusudur. Bu durumda bu değişkenler arasında yüksek düzeyde korelasyon olduğu tespit edilmiştir. Benzer şekilde EC ile TH arasında 0,83'lük bir korelasyon söz konusudur. Bu durumda yüksek düzeyde korelasyon tespit edildiği görülmektedir. Mg<sup>++</sup> ile TH arasında 0,85' lik bir korelasyon söz konusudur. Bu durumda bu değişkenler arasında yüksek düzeyde korelasyon tespit edilmiştir. Bu yüksek düzeyde korelasyonlar çoklu iç ilişkiye sebep olmaktadır. Şekil 4.5' de korelasyon grafiği verilmiştir.

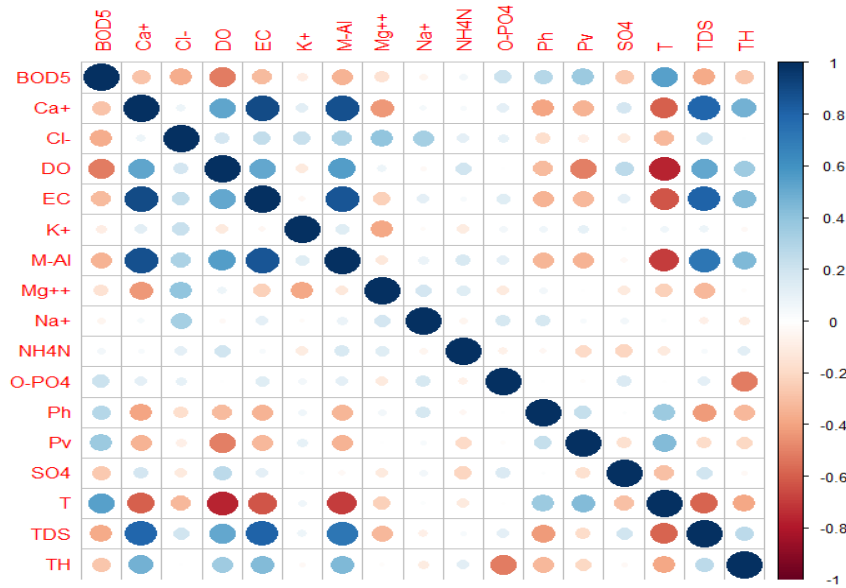


Şekil 4.5: İkinci Karşılaştırma için Korelasyon Grafiği

Çizelge 4.3: Mudurnu Çayı Dere Mahallesi Köprü ve Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonları için Korelasyon Matrisi

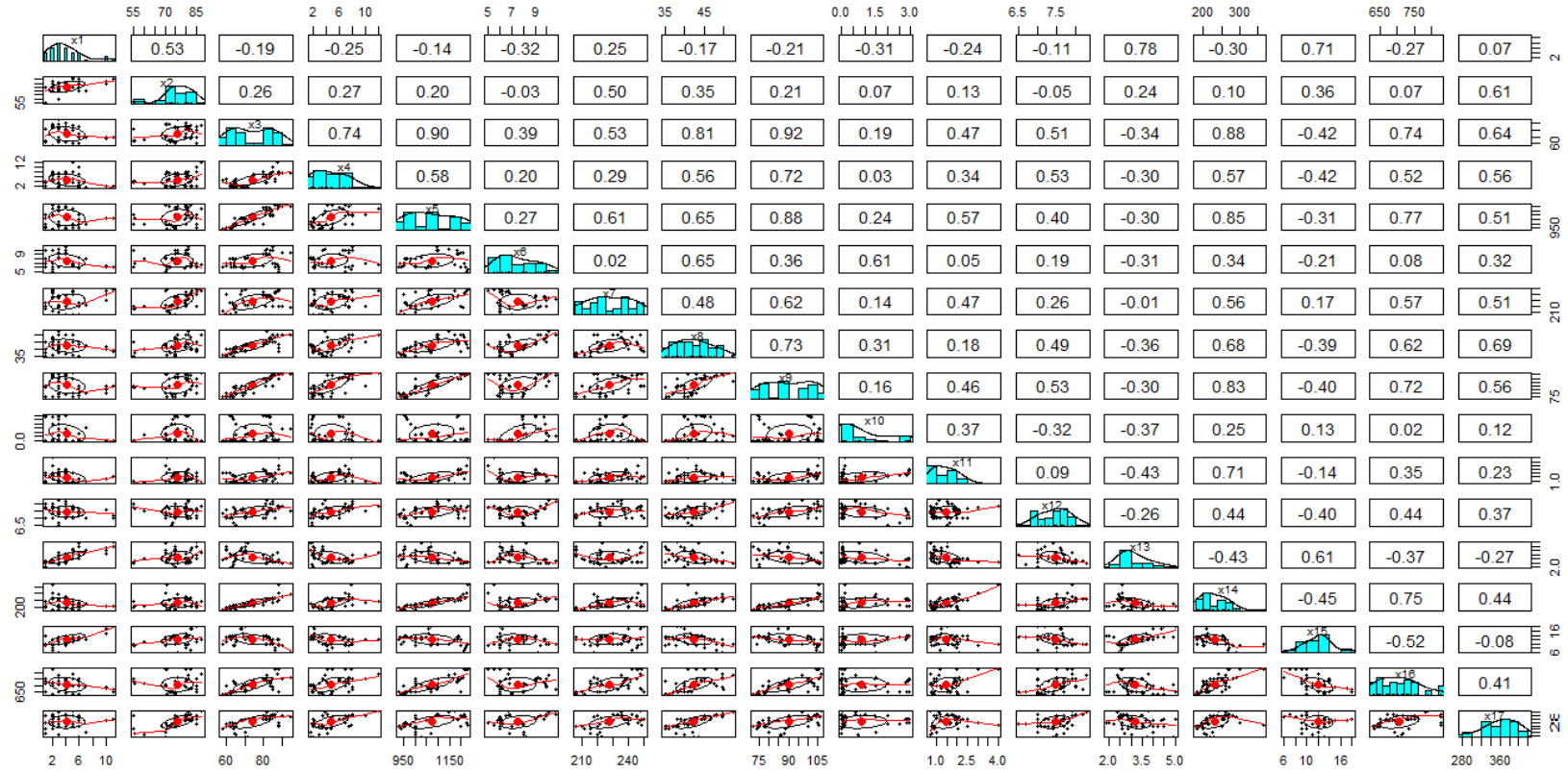


Çizelge 4.3 incelendiğinde, Ca<sup>+</sup> ile EC arasında 0,90'lık bir korelasyon söz konusudur. Aynı şekilde Ca<sup>+</sup> ile M-Al arasında 0,87'lik bir korelasyon ve Ca<sup>+</sup> ile TDS ile arasında 0,80'lik bir korelasyon söz konusudur. Bu durumda Ca<sup>+</sup> ile EC, M-Al ve TDS arasında yüksek düzeyde çoklu iç ilişki söz konusudur. EC ile Ca<sup>+</sup> ile 0,90'lık bir korelasyon, EC ile M-Al arasında 0,85'lik bir korelasyon ve EC ile TDS arasında 0,80'lik bir korelasyon söz konusudur. Bu durumda EC ile M-Al, Ca<sup>+</sup> ve TDS arasında yüksek düzeyde çoklu iç ilişkinin varlığından söz edilir. M-Al ile Ca<sup>+</sup> arasında 0,87'lik ve M-Al ile EC arasında 0,85'lik bir korelasyon söz konusudur. Bu durumda M-Al ile Ca<sup>+</sup> ve EC arasında yüksek düzeyde çoklu iç ilişki söz konusudur. Son olarak TDS ile Ca<sup>+</sup> arasında 0,80'lik ve TDS ile EC arasında 0,80'lik bir korelasyon söz konusudur. Bu durumda TDS ile Ca<sup>+</sup> ve EC arasında yüksek düzeyde çoklu iç ilişki söz konusudur. Şekil 4.6' da korelasyon grafiği verilmiştir.

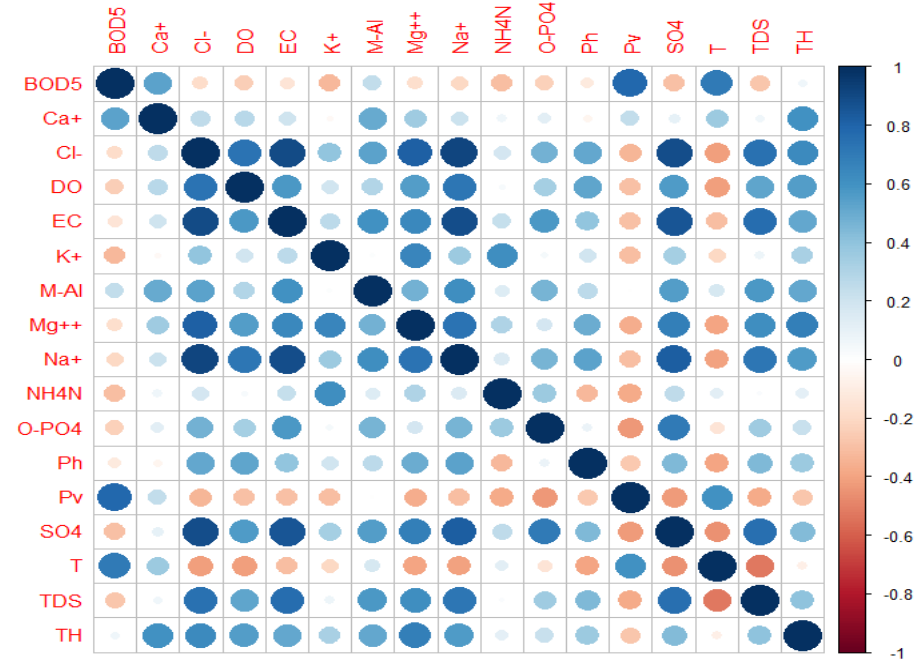


Şekil 4.6: Üçüncü Karşılaştırma için Korelasyon Grafiği

Çizelge 4.4: Sakarya Nehri Sarıyar Baraj çıkışı ve Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonları için Korelasyon Matrisi

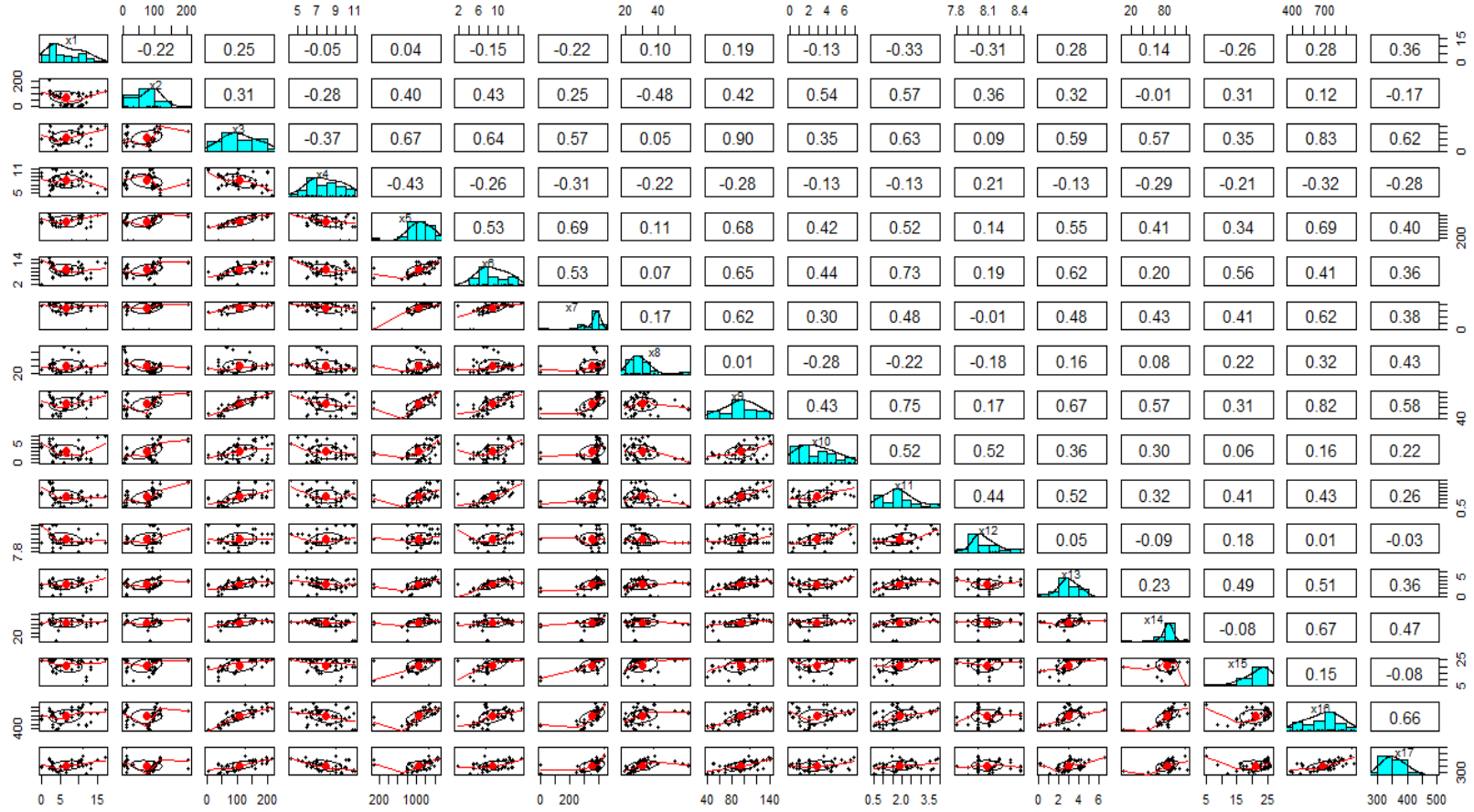


Çizelge 4.4 incelendiğinde, Cl<sup>-</sup> ile EC arasında 0,90'lık bir korelasyon söz konusudur. Aynı şekilde, Cl<sup>-</sup> ile Mg<sup>++</sup> arasında 0,81'lik, Cl<sup>-</sup> ile Na<sup>+</sup> arasında 0,92'lik ve son olarak Cl<sup>-</sup> ile SO<sub>4</sub> arasında 0,83'lük bir söz konusudur. Bu korelasyonlar Cl<sup>-</sup> ile Mg<sup>++</sup>, Na<sup>+</sup> ve SO<sub>4</sub> arasında yüksek düzeyde çoklu iç ilişki olduğunu göstermektedir. EC ile Na<sup>+</sup> arasında 0,88'lik ve EC ile SO<sub>4</sub> arasında 0,85'lik bir korelasyon söz konusudur. Bu durumda EC ile Na<sup>+</sup> ve SO<sub>4</sub> arasında yüksek düzeyde çoklu iç ilişki vardır denilmektedir. Na<sup>+</sup> ile SO<sub>4</sub> arasında 0,83'lük bir korelasyon söz konusudur. Bu durumda Na<sup>+</sup> ile SO<sub>4</sub> arasında yüksek düzeyde çoklu iç ilişkinin varlığından söz edilebilir. Şekil 4.7' de korelasyon grafiği verilmiştir.

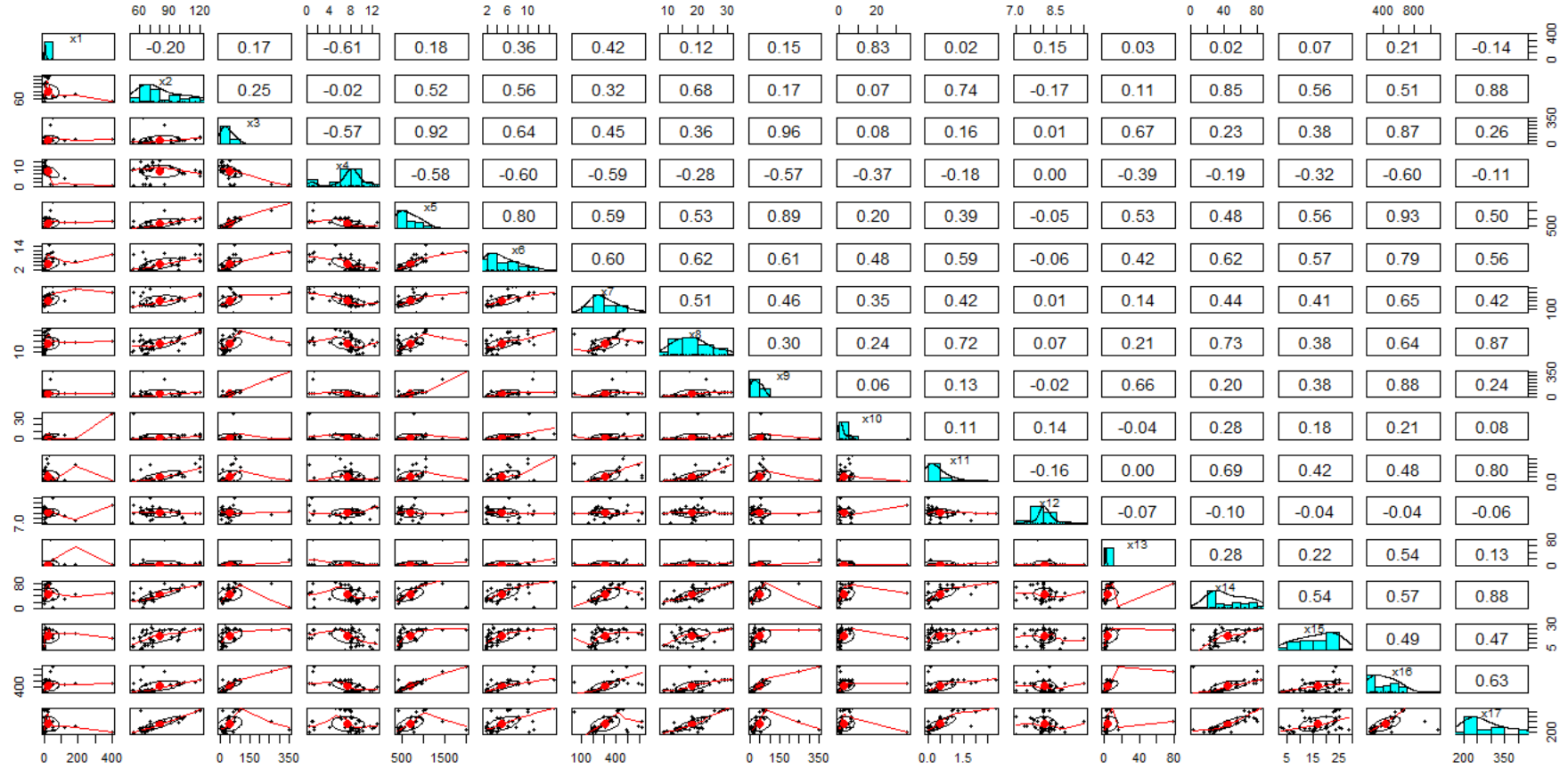


Şekil 4.7: Dördüncü Karşılaştırma için Korelasyon Grafiği

Çizelge 4.5: Çubuk Çayı Çubuk 1 Baraj Çıkışı ve Kuruçay Baraj Aksı İstasyonları İçin Korelasyon Matrisi



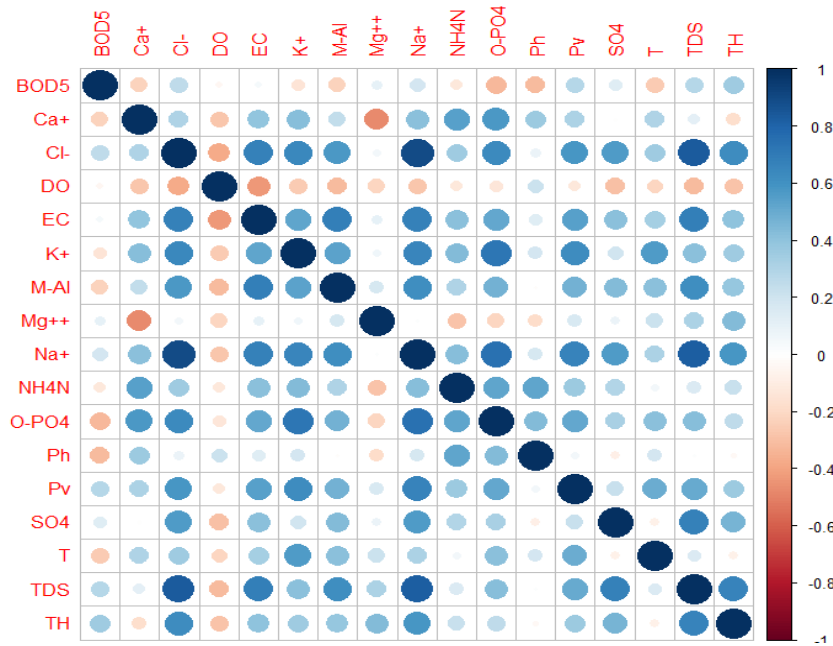
Çizelge 4.6: Çubuk Çayı Çubuk 1 Baraj çıkışı ve Bayındır Deresi Bayındır Barajı Çıkışı İstasyonları için Korelasyon Matrisi



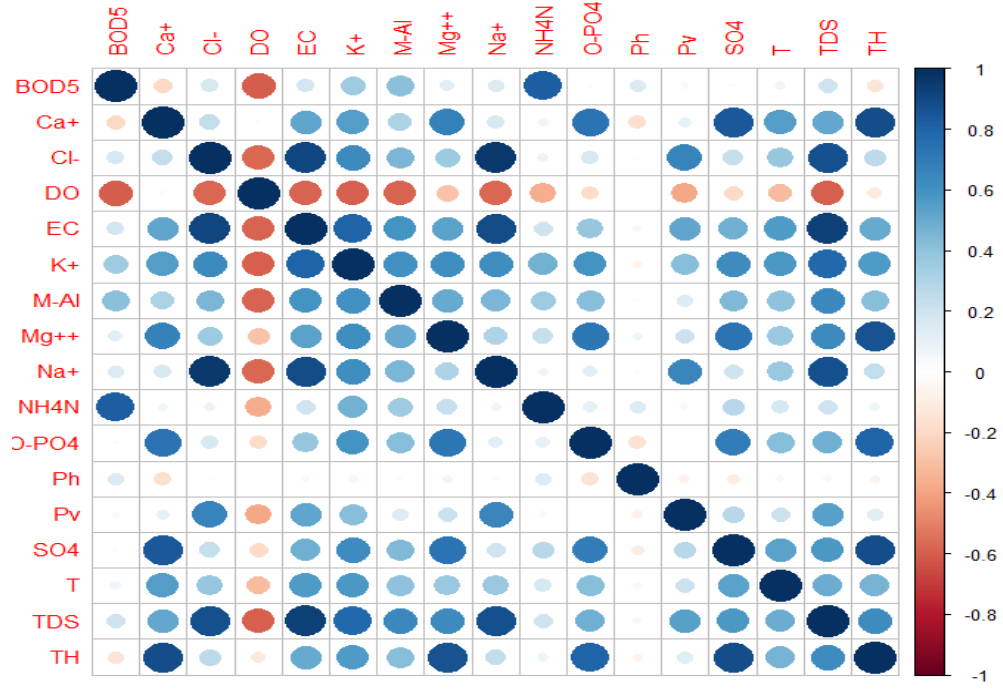


Çizelge 4.5 incelendiğinde, Cl<sup>-</sup> ile Na<sup>+</sup> arasında 0,90'lık bir korelasyon söz konusudur. Aynı şekilde Cl<sup>-</sup> ile TDS arasında 0,83'lük bir korelasyon söz konusudur. Bu durumda Cl<sup>-</sup> ile Na<sup>+</sup> ve TDS arasında yüksek düzeyde çoklu iç ilişki söz konusudur. Na<sup>+</sup> ile TDS arasında 0,82'lik bir korelasyon söz konusudur. Bu durumda Na<sup>+</sup> ile TDS arasında yüksek düzeyde çoklu iç ilişki vardır denir. Şekil 4.8' de korelasyon grafiği verilmiştir.

Çizelge 4.6 incelendiğinde, BOD<sub>5</sub>'nin NH<sub>4</sub>N ile arasında 0,83'lük bir korelasyon söz konusudur. Ca<sup>+</sup> ile TH arasında 0,88'lik bir korelasyon söz konusudur. Cl<sup>-</sup>'un EC ile arasında 0,92'lik, Na<sup>+</sup> ile arasında 0,96'lık ve TDS ile arasında 0,87'lik bir korelasyon söz konusudur. EC 'in K<sup>+</sup> ile arasında 0,80'lik, Na<sup>+</sup> ile arasında 0,89'luk, TDS ile arasında 0,93'lük bir korelasyon söz konusudur. K<sup>+</sup> ile EC arasındaki 0,80'lik korelasyon vardır. Mg<sup>++</sup> ile TH arasında 0,87'lik bir korelasyon söz konusudur. Na<sup>+</sup> ile TDS arasında 0,88'lik bir korelasyon söz konusudur. NH<sub>4</sub>N ile BOD<sub>5</sub> arasında 0,83'lük korelasyon bulunmuştur. O-PO<sub>4</sub> ile TH arasında 0,80'lik bir korelasyon söz konusudur. SO<sub>4</sub> ile TH arasında 0,88'lik bir korelasyon söz konusudur. Özetle, bağımsız değişkenler arasında yüksek korelasyon tespit edilmiştir ve çoklu iç ilişkinin varlığından söz edilebilir. Şekil 4.9' da korelasyon grafiği verilmiştir.

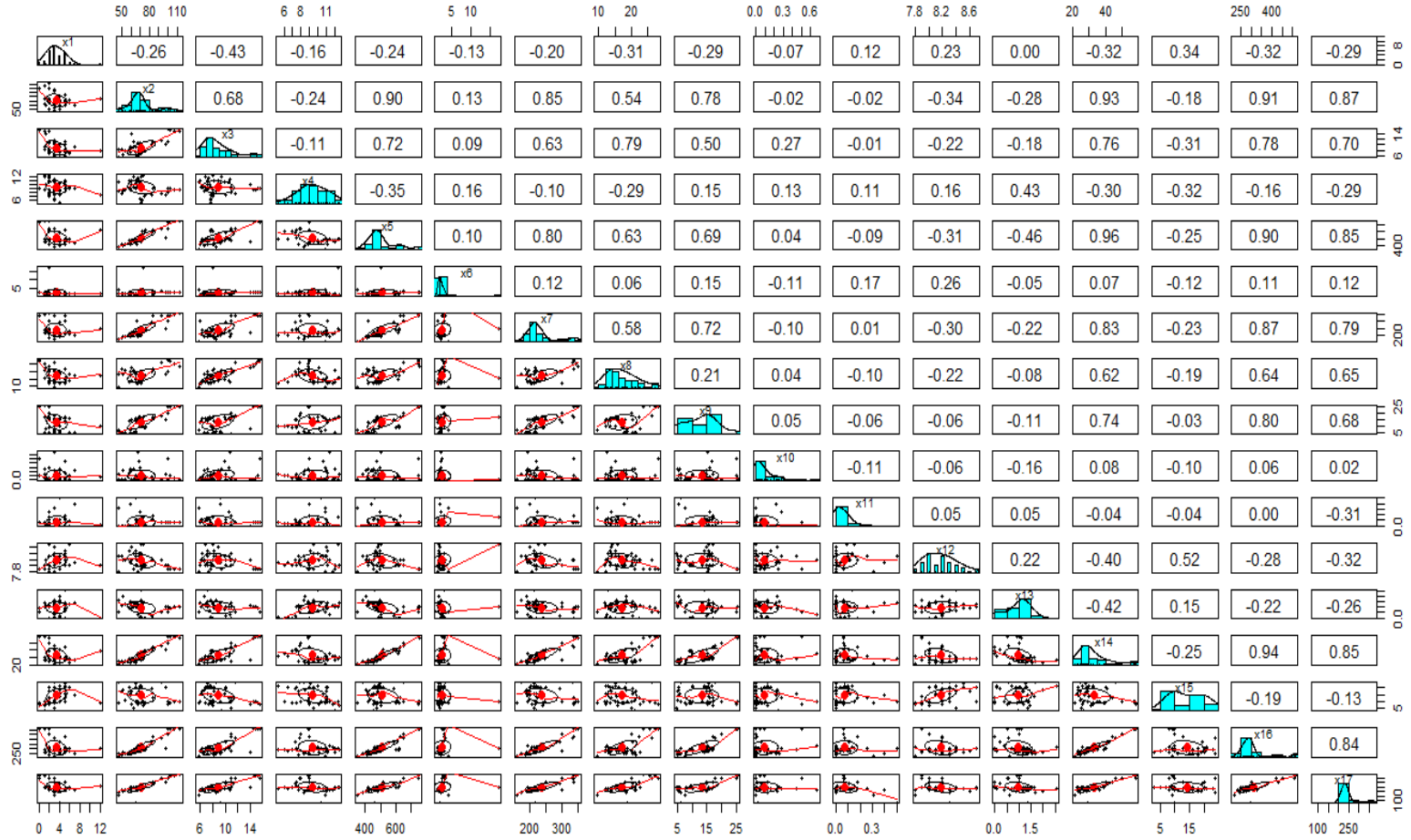


Şekil 4.8: Beşinci Karşılaştırma için Korelasyon Grafiği

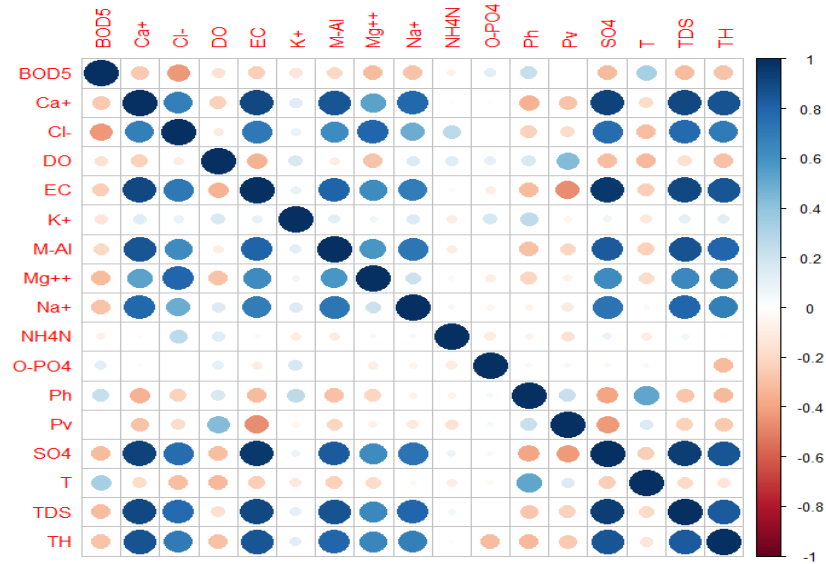


Şekil 4.9: Altıncı Karşılaştırma için Korelasyon Grafiği

Çizelge 4.7: Kuruçay Barajı Aksı ve Bayındır Deresi Bayındır Barajı Çıkışı İstasyonu İçin Korelasyon Matrisi



Çizelge 4.7 incelendiğinde, Ca<sup>+</sup> un EC ile arasında 0,90'lık, M-Al ile arasında 0,85'lik, SO<sub>4</sub> ile arasında 0,93'lük, TDS ile arasında 0,91'lik ve TH ile arasında 0,87'lik yüksek bir korelasyon söz konusudur. Aynı şekilde EC'in M-Al ile arasında 0,80'lik, SO<sub>4</sub> ile arasında 0,96'lık, TDS ile arasında 0,90'lık ve TH ile arasında 0,85'lik yüksek korelasyon söz konusudur. Na<sup>+</sup> ile TDS arasında 0,80'lik yüksek korelasyon söz konusudur. SO<sub>4</sub>'ın M-Al ile arasında 0,83'lük, TDS ile arasında 0,94'lük ve TH ile arasında 0,85'lik yüksek bir korelasyon vardır. TDS 'ın M-Al ile arasında 0,87'lik ve TH ile arasında 0,84'lük yüksek bir korelasyon söz konusudur. M-Al'nin TDS ile arasında 0,87'lik yüksek korelasyon söz konusudur. Özetle, bağımsız değişkenler arasında çoklu iç ilişki söz konusudur. Şekil 4.10' da korelasyon grafiği verilmiştir.



Şekil 4.10: Yedinci Karşılaştırma için Korelasyon Grafiği

## 5. BULGULAR VE TARTIŞMA

Bu bölümde yukarıda açıklamaları verilen istasyonlara ait veriler için Lojistik Regresyon, Robust Lojistik, Lojistik Ridge, Lojistik-LASSO ve Lojistik-Elastik Ağ yöntemleri kullanılarak regresyon modelleri kurulmuştur. Eğitim verileri kullanılarak oluşturulan modeller yardımıyla çapraz geçerlilik değerleri hesaplanmıştır. İkili lojistik regresyon modelleri kullanıldığından çapraz geçerlilik değeri olarak doğru sınıflandırma yüzdeleri hesaplanmıştır. Analizler için R programlama dili kullanılmıştır. Elde edilen analiz sonuçlarında öncelikle modelde anlamlı bulunan katsayıların yorumlanması için odds oranından yararlanılmıştır. Odd's oranı bir bağlantı ölçümüdür ve lojistik regresyon modelinin daha kolay yorumlanması için genellikle kullanılan bir değerdir. Bağımsız değişkenin 0 veya 1 olarak kodlandığı durumda  $\psi = e^{\hat{\beta}}$  şeklinde hesaplanmaktadır.

Çalışmada öncelikle birinci bölgedeki istasyonlar için lojistik regresyon analizleri yapılmış ve elde edilen sonuçlar Çizelge 5.1'de verilmiştir.

Çizelgelerde her yöntem için elde edilen tahmin değerleri, bunlara karşılık gelen odds oranları satırlarda yer almaktadır. Çizelgelerin son satırında test verisi sonucunda elde edilen doğru sınıflandırma yüzdeleri yer almaktadır. Çizelge 5.1 incelendiğinde lojistik regresyon analizi yardımıyla elde edilen modelin % 61,2'lik doğru sınıflandırmaya sahip olduğu görülmektedir.

Robust lojistik regresyon analizi ve Lojistik Ridge regresyon analizi ile elde edilen modellerin doğru sınıflandırma yüzdeleri ise sırasıyla, % 71,9 ve % 79,4'tür. Bu 3 yöntemde tüm parametreler için bir tahminde bulunmaktadır. Bir başka ifadeyle, değişken indirgemesi olmaksızın tüm model katsayılarını tahmin etmektedir. Lojistik-Lasso yönteminin doğru sınıflandırma yüzdesi % 87.5 olarak bulunmuştur. Lojistik-Elastik ağ yöntemi ile elde edilen modelin doğru sınıflandırma yüzdesi ise % 89.9 olarak elde edilmiştir.

Çizelge 5.1. Akmeşe Deresi ve İnegöl Yenicedere Mansap İst. ait Analiz Sonuçları

Parametre	Lojistik Regresyon		Robust Lojistik Regresyon		Lojistik Ridge Regresyon $k = 0,26$		Lojistik LASSO Regresyon $\lambda = 0,245$		Elastik – Lojistik Regresyon $\lambda = 0,13$ $\alpha = 0,8$	
	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds
Sabit	26,148	-	24,91	$6,58 \times 10^{10}$	26,48	-	1,045	2,84	0,463	1,59
BOD <sub>5</sub>	-0,398	0,67	-0,379	0,68	-4,286	0,014	-0,188*	0,83	- 0,125*	0,88
Ca <sup>++</sup>	0,115	1,12	0,110*	1,12	0,393*	1,480	0,061*	1,06	0,061*	1,06
Cl <sup>-</sup>	0,355*	1,43	0,338*	1,4	-2,545	0,078	-	-	-	-
DO	0,564*	1,76	0,537*	1,71	0,356*	1,427	0,061*	1,06	-0,090	0,91
EC	-0,039*	0,96	-0,038*	0,96	-2,784*	0,062	-0,001*	1	- 0,001*	0,99
K <sup>+</sup>	1,641*	5,16	1,563*	4,77	-2,229	0,11	-	-	0,014*	1,014
M-Al	0,076	1,08	0,073	1,08	-0,96	0,38	-	-	-	-
Mg <sup>++</sup>	-0,423	0,66	-0,403	0,67	-4,121*	0,017	-0,140*	0,87	- 0,134*	0,87
Na <sup>+</sup>	-0,237	0,79	-0,26	0,77	-2,948	0,052	-	-	- 0,004*	1,00
NH <sub>4</sub> N	-1,228	0,29	-1,17	0,31	-1,986*	0,14	-	-	- 0,096*	0,91
O-PO <sub>4</sub>	1,861*	6,43	1,773*	5,89	1,419*	4,13	0,134*	1,14	0,284*	1,33
Ph	-3,592	0,028	-3,422	0,033	0,834*	2,3	-0,281*	0,76	- 0,328*	0,72
Pv	-1,893	0,15	-1,899	0,15	-2,254	0,1	-	-	- 0,019*	0,98
SO <sub>4</sub>	0,207*	1,23	0,197*	1,22	-0,813	0,44	-	-	-	-
T	0,373*	1,45	0,356*	1,43	1,357*	3,88	0,047	1,05	0,058*	1,06
TDS	-0,042	0,96	-0,039	0,96	-2,133*	0,12	-	-	- 0,256*	-
TH	0,005	1,005	0,005	1,005	1,490*	4,44	0,002	1,002	0,002*	1
DSO	<b>61,2</b>		<b>71,9</b>		<b>79,4</b>		<b>87,5</b>		<b>89,9</b>	

(•0.05 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.)

Çizelge 5.1 incelendiğinde, biyolojik oksijen ihtiyacı (BOD<sub>5</sub>) sulardaki organik maddelerin mikroroganizmalar tarafından parçalanması için gerekli olan oksijen miktarını ifade etmektedir. Harcanan oksijen miktarı fazlaysa ölçüm yapılan istasyondaki organik yükün de fazla olduğunu göstermektedir. Bu durumda ölçüm yapılan istasyondaki organik kirliliğin yüksek olduğunu belirtir. Yapılan analizler sonucunda Odd's oranları hesaplandığında BOD<sub>5</sub> değerinin Akmeşe istasyonundan olma olasılığı, İnegöl Yenicedere Mansap istasyonuna olma olasılığına göre, 0,67 kat azaldığı, Lojistik-LASSO regresyon yöntemi için 0,83 kat azaldığı ve Lojistik-Elastik ağ yöntemi için 0,88 kat azaldığını

göstermektedir. Aynı zamanda BOD<sub>5</sub> değeri İnegöl Yenicedere Mansap istasyonundan Akmeşe Deresine kadar suya karışan organik yükün daha fazla olduğunu göstermektedir. Bu organik yükler karbon içeren maddelerdir ve evsel atıklardan gelmektedir. Analiz sonuçlarına bakıldığında BOD<sub>5</sub> değerinin negatif işaretli çıkması İnegöl Yenicedere Mansap (K10) istasyonunda BOD<sub>5</sub> miktarının daha fazla olduğu ve evsel atık su karışımının daha fazla olduğu söylenmektedir. Bu istasyonun çevresinde bulunan yerleşim yerleri ve fabrikaların olması bu sonucu destekler niteliktedir.

Suların sertliği, su içerisinde çözülmüş olarak bulunan toplam Ca<sup>++</sup> ve Mg<sup>++</sup> miktarının kalsiyum karbonat (CaCO<sub>3</sub>) eş değeri olarak tanımlanmıştır. Bunun sebebi, sudaki mevcut iyonlara göre daha fazla oranda Ca<sup>++</sup> ve Mg<sup>++</sup> iyonlarının bulunmasıdır (Şengün, 2013). Yapılan analizler sonucunda Ca<sup>++</sup> değerinin Odd's oranları incelendiğinde Lojistik regresyon yönteminde 1,12 kat arttığı, Lojistik-LASSO ve Lojistik-Elastik ağ yöntemi için 1,06 kat arttığı görülmektedir. Tüm yöntemler için İnegöl Yenicedere Mansap istasyonuna göre Akmeşe Deresi'nde Ca<sup>++</sup> değerinin daha fazla olduğunu göstermektedir. Analiz sonuçlarına bakıldığında kalsiyumun pozitif işaretli çıkması başka bir su kalite değişkeni olan toplam sertliği (TH) de pozitif işaretli çıkmasına sebep olmuştur. Buradan toplam sertlik (TH) için analiz sonuçlarından Odd's oranları incelendiğinde Lojistik regresyon yönteminde 1,005 kat arttığı, Lojistik-LASSO yöntemine göre 1,002 kat ve Lojistik-Elastik ağ yöntemi için ise 1,00 kat arttığı görülmektedir. Bu durumda İnegöl Yenicedere Mansap istasyonuna göre Akmeşe Deresi'nde toplam sertliğin daha fazla olduğu yargısına varılabilir. Yukarıda kalsiyum ile ilgili verilen yorumlar toplam sertliğinde arttığını destekler niteliktedir. Suların sertliğini belirleyen bir diğer parametre Mg<sup>++</sup> dur. Yapılan analiz sonuçları incelendiğinde magnezyumun her bir yöntem için Odd's oranları hesaplanmıştır. Lojistik Regresyon Analizinde, Mg<sup>++</sup> değerinin Akmeşe istasyonundan olma olasılığı, İnegöl Yenicedere Mansap istasyonuna olma olasılığına göre, 0,66 kat azaldığı, Lojistik-LASSO ve Lojistik-elastik ağ yöntemleri için ise 0,87 kat azaldığı söylenebilir.

Yüksek klorür konsantrasyonu bir kirlenme belirtisi olmakla birlikte önemli bir kalite parametresidir (Tuncay, 1983). Yapılan analizler sonucunda klorürün Odd's oranı hesaplandığında, Lojistik regresyon yöntemi için Cl<sup>-</sup> değerinin Akmeşe istasyonundan olma

olasılığı, İnegöl Yenicedere Mansap istasyonuna olma olasılığına göre, 1,43 kat arttığı görülmektedir. Lojistik-LASSO ve Lojistik-elastik ağ yöntemlerinde iki istasyon arasındaki Cl<sup>-</sup> değerinin değişiminin anlamlı olmadığı belirlenmiştir.

DO, sulardaki oksijen miktarını belirlemek için kullanılan bir parametredir. Canlı yaşamı için kritik öneme sahip olan çözünmüş oksijenin azlığı, yüzeysel sularda kirliliğin en önemli göstergesidir. Yapılan analiz sonuçlarına göre Lojistik regresyon yöntemi için Odd's oranı hesaplandığında incelendiğinde İnegöl Yenicedere Mansap istasyonuna göre Akmeşe Deresi istasyonunda çözünmüş oksijen miktarının 1,76 kat arttığı, Lojistik-LASSO yöntemi için 1,06 kat arttığı söylenmektedir. Bu sonuçlara göre, Akmeşe Deresi istasyonunda çözünmüş oksijen miktarının yüksek olduğunu ifade etmektedir ve bu da bu bölgede organik kirliliğin az olduğunun bir kanıtıdır. Lojistik-Elastik ağ yönteminde çözünmüş oksijen miktarı için elde edilen Odd's oranı değerlendirildiğinde, İnegöl Yenicedere Mansap istasyonuna göre Akmeşe Deresi istasyonunda çözünmüş oksijen miktarının 0,91 kat azaldığı söylenmektedir. Bu yöntem ile yapılan analiz sonucunda Akmeşe Deresi istasyonunda çözünmüş oksijen miktarının az olduğunu ve bu durumda bu istasyonda organik kirliliğin çok olduğunun göstermektedir.

EC, suyun elektrik akımını iletebilmesinin bir ölçüsüdür ve sulardaki mineral asitler olmak üzere çözünmüş katılardaki değişimi (TDS) ifade eder. Suyun iletkenliği sudaki iyonların toplam ve bağlı konsantrasyonlarına, hareketliliğine, değerliklerine ve ölçüm sıcaklığına bağlıdır. Suyun iletkenliği, sudaki iyon miktarı hakkında bilgi verebilir (Uslu ve Türkmen, 1987; Sawyer vd., 2013). Yapılan analiz sonuçları incelendiğinde İnegöl Yenicedere Mansap istasyonuna göre Akmeşe Deresi istasyonunda, Lojistik regresyon sonuçlarına göre elektriksel iletkenliğin 0,96 kat azaldığı, Lojistik-LASSO ve Lojistik-Elastik ağ yöntemleri için ise 0,99 kat azaldığı görülmektedir. Bu sonuçlar birlikte değerlendirildiğinde, İnegöl Yenicedere Mansap istasyonuna göre Akmeşe Deresi istasyonunda iyon miktarının daha az olduğunu göstermektedir.

K<sup>+</sup> iç sularda tuzluluğu oluşturan katyon grubundan biridir. Yapılan analiz sonuçları incelendiğinde potasyum değerinin iki istasyon arasında su kalitesi değişimi üzerinde etkili olduğu, lojistik regresyon ve lojistik-Elastik Ağ analizleri ile tespit edilmiştir. Lojistik



regresyon analizine göre, İnegöl Yenicedere Mansap istasyonu ile Akmeşe Deresi istasyonu arasında bu değerin 5,16 kat arttığı, Lojistik-Elastik Ağ yöntemine göre ise 1,01 kat arttığı belirlenmiştir.

O-PO<sub>4</sub> için yapılan analiz sonuçları incelendiğinde tüm yöntemler için iki istasyon arasında su kalitesi değişimi üzerinde bu değişkenin önemli olduğu görülmektedir. Lojistik regresyon yöntemine göre orto fosfatın 6,43 kat arttığı, Lojistik-LASSO yöntemine göre 1,14 kat arttığı ve Lojistik-Elastik ağ yöntemine göre 1,33 kat arttığı tespit edilmiştir. Sonuç olarak, Akmeşe Deresi'nde daha fazla orto fosfat bulunması ölçüm yapılan istasyon çevresinde bulunan tarlalarda fosforlu gübre kullanıldığını ve fazla gübrenin yağışla beraber dereye karıştığını göstermektedir.

NH<sub>4</sub>N genellikle çözülmüş oksijenden sonra ikinci önemli su kalite parametresidir. Sudaki NH<sub>4</sub>N, organik maddelerin bozunması, özellikle organik gübre veya inorganik amonyum kaynaklı kimyasal gübreleme sonucunda, evsel ve endüstriyel kirlenmelerden oluşabilir (Egemen ve Sunlu,1999). Yapılan analizler sonucunda, lojistik regresyon analizine göre bu değişkenin istatistiksel olarak anlamlı olmadığı ancak lojistik-elastik ağ yönteminde istatistiksel olarak anlamlı bulunduğu belirlenmiştir. Analiz sonuçlarına göre Odd's oranları incelendiğinde, İnegöl Yenicedere Mansap istasyonuna göre Akmeşe deresi istasyonunda NH<sub>4</sub>N değerinin lojistik regresyon yöntemi için 0,29 kat azaldığı ve lojistik-elastik ağ yöntemine göre 0,91 kat azaldığı görülmektedir. Her iki yönteme göre de Akmeşe Deresi istasyonunda NH<sub>4</sub>N miktarının daha az olduğu görülmektedir. Akmeşe Deresi istasyonu çevresinde evsel veya endüstriyel kirlenme yaratacak yerleşim yerleri, fabrikaların bulunmaması bu sonucu desteklemektedir.

Na<sup>+</sup> suda kolaylıkla çözünen ve suya tuzlu bir tat veren bir katyondur. Na<sup>+</sup> içeren sular tarımsal sulamalarda toprak ve bitkiler üzerinde zararlı etkiler yapmaktadır (Tuncay,1983). Yapılan analizler sonucunda iki istasyon arasındaki Na<sup>+</sup> değerindeki değişiminin Lojistik-LASSO yönteminde anlamlı olmadığı belirlenmiştir. İnegöl Yenicedere Mansap (K10) istasyonuna göre Akmeşe Deresi istasyonunda, Na<sup>+</sup> değerinin lojistik regresyon analizine göre 0,79 kat azaldığı, Lojistik-Elastik ağ yöntemine göre ise 1 kat arttığı görülmektedir. Analiz sonuçları incelendiğinde Na<sup>+</sup> değerinin lojistik regresyon

yönteminde istatistiksel olarak anlamlı bulunmadığı görülmektedir. Bu nedenle, Lojistik- elastik ağ sonuçlarına dayanarak Akmeşe Deresi istasyonunda daha fazla Na<sup>+</sup> elementi bulunduğu ve suyun daha tuzlu olduğu söylenebilir.

Çizelge 5.2 de Sakarya Nehri Sarıyar Baraj çıkışı istasyonu ve Mudurnu Çayı Dere mahallesi köprü istasyonuna ait analiz sonuçları verilmiştir.

Çizelge 5.2. Sakarya Nehri Sarıyar Baraj Çıkışı İstasyonu ve Mudurnu Çayı Dere Mah. Köprü İst. ait Analiz Sonuçları

Parametre	Lojistik Regresyon		Robust Lojistik Regresyon		Lojistik Ridge Regresyon $k = 0,13$		Lojistik LASSO Regresyon $\lambda = 0,61$		Elastik – Lojistik Regresyon $\lambda = 0,021$ $\alpha = 0,8$	
	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds
Sabit	-11,055	-	-10,045*	-4,34x10 <sup>-5</sup>	-10,021	-	6,197	-	2,325	-
BOD5	0,266*	1,30	0,312*	1,37	0,613*	1,85	0,421*	1,523	0,721*	2,056
Ca <sup>++</sup>	0,135	1,14	0,105*	1,11	-1,277	0,28	1,020*	2,773	0,936	2,550
Cl <sup>-</sup>	-0,202	0,82	-0,202*	0,82	-7,363	6,34x10 <sup>-04</sup>	-0,037*	0,96	-	0,97
DO	1,237*	3,45	0,927*	2,53	7,929*	2,78	0,302*	1,35	0,302*	1,35
EC	-0,008	0,99	-0,103*	0,90	-7,437	5,89x10 <sup>-04</sup>	-	-	-0,003	0,99
K <sup>+</sup>	-0,334	0,72	-0,412*	0,66	-4,414	0,01	-0,214	0,807	-	0,902
M-Al	0,087	1,09	0,063*	1,07	-0,343	0,71	-	-	-	-
Mg <sup>++</sup>	-0,588	0,56	-0,610	0,54	-7,325	6,59x10 <sup>-04</sup>	-0,019*	0,98	-	0,93
Na <sup>+</sup>	0,009	1,01	0,011	1,01	-7,562	5,20x10 <sup>-04</sup>	-0,036*	0,96	-	0,97
NH4N	2,331*	10,29	1,641*	0,01	-0,839	0,43	-	-	1,902*	6,699
O-PO4	-3,831	0,02	-2,138	0,12	-4,969	6,95x10 <sup>-03</sup>	-	-	-	-
Ph	4,108*	60,82	2,197*	9,00	5,873*	355,313	0,534*	1,71	1,008*	2,74
Pv	-0,373	0,69	-0,403	0,67	-4,112	0,02	-	-	-	0,98
SO4	0,108	1,11	0,097	1,10	-7,172	7,68x10 <sup>-04</sup>	-	-	-0,009	0,99
T	0,044	1,04	0,056*	1,06	1,795*	6,02	-	-	-	-
TDS	-0,056	0,95	-0,023	0,98	-7,976	344x10 <sup>-04</sup>	-0,018*	0,98	-	1,01
TH	-0,044	0,96	-0,109	0,90	-6,406	1,65x10 <sup>-03</sup>	-	-	-	1,00
DSO	48,6		63,8		50,5		67,1		80,42	

(•0.05 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.)

Lojistik regresyon modeli ve Robust Lojistik regresyon modeliyle ile yapılan analizler ayrıntılı olarak incelendiğinde birçok su karakteristiklerinin istatistiksel olarak anlamsız bulunduğu görülmektedir. Lojistik regresyon analizi sonuçlarından baz istasyon olan Sakarya Nehri Sarıyar Baraj Çıkışı istasyonundan Mudurnu Çayı Dere mahallesi Köprü istasyonuna kadar su kalitesinin değişiminde BOD<sub>5</sub>, DO, NH<sub>4</sub>N, pH ve SO<sub>4</sub> değişkenlerinin istatistiksel açıdan anlamlı olduğu bulunmuştur. Ancak tüm değişkenler için katsayı hesaplandığından Odd's oranları değişmektedir. Odd's oranları incelendiğinde, BOD<sub>5</sub> değerinin Sakarya Nehri Sarıyar Baraj çıkışı olma olasılığı, Mudurnu Çayı Dere Mahallesinden olma olasılığına göre, 1,30 kat arttığı gözlemlenmiştir. Benzer şekilde Sakarya Nehri Sarıyar Baraj çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi köprü istasyonunda Ca<sup>++</sup> değerinin 1,14 kat, DO değerinin 3,45 kat, M-Al değerinin 1,09 kat, Na<sup>+</sup> değerinin 1,01 kat, NH<sub>4</sub>N değerinin 10,29 kat, Ph değerinin 60,82 kat, SO<sub>4</sub> değerinin 1,11 kat, T değerinin 1,04 kat artarken; Cl<sup>-</sup> değerinin 0,82 kat, EC değerinin 0,99 kat, K<sup>+</sup> değerinin 0,72 kat, Mg<sup>++</sup> değerinin 0,56 kat, O-PO<sub>4</sub> değerinin 0,02 kat, Pv değerinin 0,69 kat, TDS değerinin 0,95 kat ve TH değerinin 0,96 kat azaldığı görülmektedir. Robust Lojistik regresyon analizi sonuçlarına göre ise baz istasyon olan Sakarya Nehri Sarıyar Baraj Çıkışı istasyonundan Mudurnu Çayı Dere mahallesi Köprü istasyonuna kadar su kalitesinin değişiminde BOD<sub>5</sub>, Ca<sup>++</sup>, Cl<sup>-</sup>, DO, EC, K<sup>+</sup>, M-Al, NH<sub>4</sub>N, pH ve T değişkenlerinin istatistiksel açıdan anlamlı olduğu bulunmuştur. Ancak tüm değişkenler için katsayı hesaplandığından Odd's oranları değişmektedir. Odd's oranları incelendiğinde, BOD<sub>5</sub> değerinin Sakarya Nehri Sarıyar Baraj çıkışı olma olasılığı, Mudurnu Çayı Dere Mahallesinden olma olasılığına göre, 1,37 kat arttığı gözlemlenmiştir. Benzer şekilde Sakarya Nehri Sarıyar Baraj çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi köprü istasyonunda Ca<sup>++</sup> değerinin 1,11 kat, DO değerinin 2,53 kat, M-Al değerinin 1,07 kat, Na<sup>+</sup> değerinin 1,01 kat, pH değerinin 9,00 kat, SO<sub>4</sub> değerinin 1,10 kat, T değerinin 1,06 kat artarken; Cl<sup>-</sup> değerinin 0,82 kat, EC değerinin 0,90 kat, K<sup>+</sup> değerinin 0,66 kat, NH<sub>4</sub>N değerinin 0,01 kat, Mg<sup>++</sup> değerinin 0,54 kat, O-PO<sub>4</sub> değerinin 0,12 kat, Pv değerinin 0,67 kat, TDS değerinin 0,95 kat ve TH değerinin 0,90 kat azaldığı görülmektedir.

Lojistik Ridge regresyon yöntemiyle yapılan analizler sonucunda Tablo 5.2 incelendiğinde baz istasyon olan Sakarya Nehri Sarıyar Baraj Çıkışı istasyonundan Mudurnu Çayı Dere mahallesi Köprü istasyonuna kadar BOD<sub>5</sub>, DO, pH ve T değişkenlerinin

istatistiksel açıdan anlamlı olduğu bulunmuştur. Odd's oranları incelendiğinde, Sakarya Nehri Sarıyar Baraj çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi köprü istasyonunda BOD<sub>5</sub> değerinin 1,85 kat, DO değerinin 2,78 kat, pH değerinin 355,313 kat ve T değerinin 6,02 kat artarken, Ca<sup>++</sup> değerinin 0,28 kat, Cl<sup>-</sup> değerinin 6,34x10<sup>-04</sup> kat, EC değerinin 5,89x10<sup>-04</sup> kat, K<sup>+</sup> değerinin 0,01 kat, M-Al değerinin 0,71 kat, Mg<sup>++</sup> değerinin 6,59x10<sup>-04</sup> kat, Na<sup>+</sup> değerinin 5,20x10<sup>-0,4</sup> kat, NH<sub>4</sub>N değerinin 0,43 kat, O-PO<sub>4</sub> değerinin 6,95x10<sup>-03</sup> kat, Pv değerinin 0,02 kat, SO<sub>4</sub> değerinin 7,68x10<sup>-04</sup> kat, TDS değerinin 7,68x10<sup>-04</sup> kat, TH değerinin 1,65x10<sup>-03</sup> kat azaldığı gözlemlenmiştir.

Çizelge 5.2 incelendiğinde, %48,6'lık bir doğru sınıflandırma yüzdesine sahip olan analiz, Lojistik Regresyon Analizidir. Klasik lojistik regresyon analizinde birçok katsayı istatistiksel olarak anlamsız bulunmuştur. En yüksek doğru sınıflama yüzdesi veren analiz, Lojistik Elastik Regresyon analizidir.

Lojistik LASSO yöntemiyle yapılan analizler sonucunda baz istasyon olan Sakarya Nehri Sarıyar Baraj çıkışı istasyonundan Mudurnu Çayı Dere mahallesi köprü istasyonuna kadar su kalitesinin değişiminde BOD<sub>5</sub>, Ca<sup>++</sup>, Cl<sup>-</sup>, DO, Mg<sup>++</sup>, Na<sup>+</sup>, pH ve TDS değişkenlerinin istatistiksel açıdan anlamlı olduğu bulunmuştur. Anlamlı çıkan değişkenler için Odd's oranları hesaplanmış olup Çizelge 5.2' de verilmiştir. Benzer şekilde Lojistik Elastik ağ yöntemiyle yapılan analizler sonucunda baz istasyon olan Sakarya Nehri Sarıyar Baraj çıkışı istasyonundan Mudurnu Çayı Dere mahallesi köprü istasyonuna kadar su kalitesinin değişiminde BOD<sub>5</sub>, Cl<sup>-</sup>, DO, K<sup>+</sup>, Mg<sup>++</sup>, Na<sup>+</sup>, NH<sub>4</sub>N pH, pV, TDS ve TH değişkenleri istatistiksel açıdan anlamlı bulunmuştur. Anlamlı çıkan değişkenler için Odd's oranları hesaplanmış ve Çizelge 5.2' de verilmiştir.

Lojistik LASSO yöntemiyle yapılan analizler sonucunda Sakarya Nehri Sarıyar Baraj çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi köprü istasyonunda inorganik parametrelerden olan Cl<sup>-</sup> 0,96 kat azaldığı söylenebilir. Mg<sup>++</sup> suyun sertliğini veren iyonlardan biridir. Sakarya Nehri Sarıyar Baraj çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi köprü istasyonunda Mg<sup>++</sup> değerinin 0,98 kat az olduğu yani Mudurnu Çayı Dere mahallesi Köprü istasyonundaki suyun daha yumuşak olduğu söylenebilir.

Çizelge 5.2' ye göre lojistik LASSO yöntemiyle yapılan analiz sonucunda Sakarya Nehri Sarıyar Baraj çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi köprü istasyonunda BOD<sub>5</sub> değerinin 1,253 kat ve Ca<sup>++</sup> değerinin 2,773 kat arttığı gözlemlenmiştir. Aynı zamanda K<sup>+</sup> değerinin Sakarya Nehri Sarıyar Baraj çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi köprü istasyonunda 0,807 kat azaldığı gözlemlenmektedir. Benzer şekilde lojistik Elastik ağ yöntemiyle yapılan analiz sonucunda BOD<sub>5</sub> değerinin 0,256 kat, Ca<sup>++</sup> değerinin 2,550 kat, NH<sub>4</sub>N değerinin 6,699 kat arttığı gözlenmiştir. K<sup>+</sup> değerinin ise Sakarya Nehri Sarıyar Baraj çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi köprü istasyonunda 0,902 kat azaldığı gözlenmiştir.

DO, sulardaki oksijen miktarını belirlemek için kullanılan bir parametredir. Canlı yaşamı için kritik öneme sahip olan çözünmüş oksijenin azlığı, yüzeysel sularda kirliliğin en önemli göstergesidir. Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Mudurnu Çayı dere mahallesi Köprü istasyonunda hem lojistik-LASSO hem de lojistik-Elastik ağ yöntemine göre 1,35 kat arttığı söylenmektedir. Bu durumda bu yöntemler göre Mudurnu Çayı dere mahallesi Köprü istasyonunda çözünmüş oksijen miktarının yüksek olduğunu ifade etmektedir ve bu da bu bölgede organik kirliliğin az olduğunun bir kanıtıdır.

EC, suyun elektrik akımını iletebilmesinin bir ölçüsüdür ve sulardaki mineral asitler olmak üzere çözünmüş katılardaki değişimi ifade eder. Suyun iletkenliği sudaki iyonların toplam ve bağıl konsantrasyonlarına, hareketliliğine, değerliklerine ve ölçüm sıcaklığına bağlıdır. Suyun iletkenliği, sudaki iyon miktarı hakkında bilgi verebilir (Uslu ve Türkmen, 1987; Sawyer vd., 2013). Lojistik-Elastik ağ yöntemiyle yapılan analiz sonucunda EC değeri Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Mudurnu Çayı dere mahallesi Köprü istasyonunda 0,99 kat azaldığını ve iyon miktarının daha az olduğunu göstermektedir. Bu da Mudurnu Çayı dere mahallesi Köprü istasyonunda inorganik madde miktarının daha az olduğunun bir kanıtıdır. Elektriksel iletkenliğin az olduğu Cl<sup>-</sup>, Mg<sup>++</sup>, ve Na<sup>+</sup> miktarlarının da az olması durumunu destekler niteliktedir.

Sudaki çözünmüş maddeler, suda çözünmüş az miktardaki organik madde ve anorganik tuzların varlığından ileri gelir. TDS suyun bulanıklılığını artırmaktadır. Toplam çözünmüş maddeler, doğal kaynaklardan, lağım atıklarından, şehir drenaj sularından ve

endüstriyel sulardan ileri gelmektedir. Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Mudurnu Çayı dere mahallesi Köprü istasyonunda Lojistik-LASSO yöntemiyle yapılan analiz sonucunda TDS değerinin 0,98 kat azalırken Lojistik-Elastik ağ yönteminde 1,01 kat artmaktadır. Buradan elektriksel iletkenliğin Lojistik-Elastik ağ yöntemiyle yapılan analizde Mudurnu Çayı dere mahallesi Köprü istasyonunda daha az olması toplam çözünmüş katılardaki değişimi etkilediği söylenebilir.

Suların sertliğini belirleyen parametrelerden biri  $Mg^{++}$  dur. Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Mudurnu Çayı dere mahallesi Köprü istasyonunda Lojistik-LASSO yöntemiyle yapılan analiz sonucunda 0,09 kat azaldığı ve Lojistik-Elastik-ağ yöntemiyle yapılan analiz sonucunda 0,93 kat azaldığı söylenebilir. Bu durumda Mudurnu Çayı Dere mahallesi Köprü istasyonundaki suyun daha yumuşak olduğu söylenmektedir.

$Na^+$  suda kolaylıkla çözünen ve suya tuzlu bir tat veren bir katyondur.  $Na^+$  içeren sular tarımsal sulamalarda toprak ve bitkiler üzerinde zararlı etkiler yapmaktadır (Tuncay, 1983). Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Mudurnu Çayı dere mahallesi Köprü istasyonunda Lojistik-LASSO yöntemiyle yapılan analiz sonucunda 0,02 kat azaldığı ve Lojistik-Elastik-ağ yöntemiyle yapılan analiz sonucunda 0,97 kat azaldığı söylenebilir. Bu yorumlara göre Sakarya Nehri Sarıyar Baraj Çıkışı istasyonunda daha fazla  $Na^+$  elementinin bulunduğu ve daha tuzlu olduğu söylenebilir.

Yüksek klorür konsantrasyonu bir kirlenme belirtisi olmakla birlikte önemli bir kalite parametresidir (Tuncay 1983). Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Mudurnu Çayı dere mahallesi Köprü istasyonunda Lojistik-LASSO yöntemiyle yapılan analiz sonucunda 0,96 kat azaldığı ve Lojistik-Elastik ağ yönteminde 0,97 kat azaldığı görülmektedir.

Değişik sanayilerden atılan atıklarda ve minerallerin kavrulması işleminden sonra sülfat sulara verilir. Suda sülfat genellikle yüksek konsantrasyonlarda bulunabilir. Çünkü, kayalardan çözülmeye geçen katyonlar genellikle sülfatla, çözünebilen bileşikler verirler. Lojistik-Elastik ağ yöntemiyle yapılan analizde Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Mudurnu Çayı dere mahallesi Köprü istasyonunda  $SO_4$  değerinin 0,99 kat

azaldığı söylenmektedir. Bu durumda Sakarya Nehri Sarıyar Baraj çıkışında bulunan tarlalardan gelen kayaçlardan dolayı bu istasyonda sülfat değerinin daha fazla olduğunu bir kanıtı olabilir. Özetle, hem Lojistik-LASSO hem de Lojistik-elastik ağ yöntemiyle yapılan analizde çözünmüş oksijen ve Ph değerlerinin Mudurnu Çayı Dere mahallesi Köprü istasyonunda daha fazladır.

Lojistik-LASSO yöntemiyle yapılan analizde Cl-, Mg<sup>++</sup>, Na<sup>+</sup> ve TDS değerlerinin Mudurnu Çayı Dere mahallesi Köprü istasyonunda daha az olduğu görülmektedir. Bu değişkenlere ek olarak Lojistik-Elastik ağ yönteminde EC, Pv, SO<sub>4</sub> ve TH değerleri de değerlerinin Mudurnu Çayı Dere mahallesi Köprü istasyonunda daha azdır. Buradan Lojistik-Elastik ağ yöntemi daha fazla değişken ile suyun kalitesini açıklamaktadır. Bu amaçla Lojistik-LASSO yöntemine göre daha iyi bir yöntem olduğu söylenebilir.

İkinci grup istasyonlar için ikinci karşılaştırma olarak adlandırdığımız sonuçlar Kocaçay Deresi Kayaboğazı Baraj Çıkışı ile Mudurnu Çayı Dere mahallesi Köprü istasyonları için analiz sonuçlarını göstermektedir. Bu sonuçlar Çizelge 5.3' de verilmiştir.

Çizelge 5.3. Kocaçay Deresi Kayaboğazı Baraj Çıkışı ve Mudurnu Çayı Dere Mah. Köprü İst. Analiz Sonuçları

Parametre	Lojistik Regresyon		Robust Lojistik Regresyon		Lojistik Ridge Regresyon $k = 0,29$		Lojistik LASSO Regresyon $\lambda = 0,48$		Elastik – Lojistik Regresyon $\lambda = 0,03$ $\alpha = 0,62$	
	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds
Sabit	58,33	-	51,33	-	48,01	-	7,370	-	4,796	-
BOD5	-0,127	0,88	-0,087	0,917	0,429*	1,54	0,109*	1,115	0,243*	1,28
Ca++	-0,625	0,54	-0,045	0,956	1,308*	3,7	-	-	2,013*	7,48
Cl-	-0,086	0,92	1,103	3,013	1,427*	4,17	1,263	3,53	0,985*	2,67
DO	0,719*	2,05	0,821*	2,273	3,558*	35,09	1,107	3,025	0,236*	1,27
EC	0,088*	1,09	0,976*	2,654	3,092*	22,02	2,521*	12,44	1,420*	4,13
K+	-0,094	0,91	0,854	2,349	0,944*	2,57	0,726	2,066	0,529	1,697
M-Al	-0,027	0,97	0,598	1,818	0,929*	2,53	-	-	1,012*	2,75
Mg++	-3,268	0,04	-1,026	0,358	-1,442	0,26	-0,929*	0,39	-0,874	0,42
Na+	0,503	1,653	0,521*	1,684	0,661	1,937	1,246*	3,48	1,149*	3,16
NH4N	-1,125	0,324	-1,004	0,366	-0,505	0,6	-	-	-	-
O-PO4	4,486*	88,77	2,162*	8,688	3,662*	38,94	-	-	2,539*	12,67
Ph	-1,832	0,16	-1,901	0,149	2,179*	8,84	1,243*	3,466	1,508*	4,517
Pv	2,393*	10,95	-0,634*	0,530	-0,461	0,63	3,261*	26,07	2,513*	12,34
SO4	-0,254	0,78	1,529	4,614	2,277*	9,75	-	-	2,534*	12,60
T	-0,533	0,59	1,231	3,425	2,019*	7,53	-	-	1,428*	4,170
TDS	-0,107	0,90	-0,984	0,374	-1,129	0,32	-0,012*	0,99	-0,018*	0,98
TH	-0,005	1	0,632	1,881	0,536*	1,71	-0,008	0,99	-0,0002*	1
DSO	32,5		47,6		62,8		65,26		84,42	

(•0.05 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.)

Çizelge 5.3 incelendiğinde Lojistik regresyon analizi ile elde edilen modelin doğru sınıflandırma oranının %32,5 olduğu tespit edilmiştir. Robust Lojistik regresyon analizi ile elde edilen modelin doğru sınıflandırma oranlarının %47,6 olduğu görülmektedir. Lojistik Ridge regresyon analizi ile edilen model % 62,8 doğru sınıflandırma oranına sahiptir. Lojistik Elastik Ağ yönteminin doğru sınıflandırma yüzdesinin yüksek olduğu görülmektedir.



Lojistik regresyon analizi yöntemiyle yapılan analizler sonucunda baz istasyon olan Kocaçay Deresi Kayaboğazı Baraj Çıkışı istasyonundan Mudurnu Çayı Dere Mahallesi Köprü istasyonuna kadar su kalitesinin değişiminde DO, EC, Na<sup>+</sup>, O-PO<sub>4</sub> ve pV değişkenlerinin istatistiksel olarak anlamlı bulunmuştur. Robust lojistik regresyon analizi yöntemiyle yapılan analiz sonucunda DO, EC, Na<sup>+</sup>, O-PO<sub>4</sub> ve pV değişkenleri istatistiksel olarak anlamlı bulunmuştur. Hem Lojistik regresyon analizi hem de robust lojistik regresyon analizi yöntemleri için tüm değişkenlerin modelde bulunması gereklidir. Bu amaçla tüm değişkenler için Odd' s oranları hesaplanmış olup Çizelge 5.3' te verilmiştir. Lojistik regresyon yöntemiyle yapılan analiz sonucunda Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonu olma olasılığı Mudurnu Çayı Dere Mahallesinden olma olasılığına göre BOD<sub>5</sub> değerinin 0,88 kat, Ca<sup>++</sup> değerinin 0,54 kat, Cl<sup>-</sup> değerinin 0,92 kat, K<sup>+</sup> değerinin 0,91 kat, M-Al değerinin 0,97 kat, Mg<sup>++</sup> değerinin 0,04 kat, NH<sub>4</sub>N değerinin 1,29x10<sup>-05</sup> kat, pH değerinin 0,16 kat, SO<sub>4</sub> değerinin 0,78 kat, T değerinin 0,59 kat, TDS değerinin 0,90 kat azalırken, DO değerinin 2,05 kat, EC değerinin 1,09 kat, Na<sup>+</sup> değerinin 152,47 kat, O-PO<sub>4</sub> değerinin 88,77 kat, Pv değerinin 10,95 kat ve TH değerinin 1 kat arttığı gözlemlenmiştir.

Robust lojistik regresyon yöntemiyle yapılan analiz sonucunda ise odds oranları incelenecek olursa BOD<sub>5</sub> değerinin 0,917 kat, Ca<sup>++</sup> değerinin 0,956 kat, Mg<sup>++</sup> değerinin 0,358 kat, NH<sub>4</sub>N değerinin 0,366 kat, pH değerinin 0,149 kat, Pv değerinin 0,530 kat, SO<sub>4</sub> değerinin 0,78 kat, TDS değerinin 0,374 kat azalırken, Cl<sup>-</sup> değerinin 4,614 kat, DO değerinin 2,273 kat, EC değerinin 2,654 kat, K<sup>+</sup> değerinin 2,349 kat, Na<sup>+</sup> değerinin 1,684 kat, M-Al değerinin 1,818 kat, O-PO<sub>4</sub> değerinin 8,688 kat, T değerinin 3,425 kat, TH değerinin 1,881 kat arttığı gözlemlenmiştir

Lojistik-Ridge regresyon analizi sonucunda BOD<sub>5</sub>, Ca<sup>++</sup>, Cl<sup>-</sup>, DO, EC, K<sup>+</sup>, M-Al, Na<sup>+</sup>, O-PO<sub>4</sub>, pH, SO<sub>4</sub>, T ve TH değişkenlerinin istatistiksel açıdan anlamlı bulunduğu Çizelge 5.3' de görülmektedir. Bu anlamlı çıkan değişkenler için Odd's oranları hesaplanmıştır. Odd's oranları incelendiğinde, Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi köprü istasyonunda BOD<sub>5</sub> değerinin 1,54 kat, Ca<sup>++</sup> değerinin 3,70 kat, Cl<sup>-</sup> değerinin 4,17 kat, DO değerinin 35,09 kat, EC değerinin

22,02 kat, K+ değerinin 2,57 kat, M-Al değerinin 2,53 kat, O-PO<sub>4</sub> değerinin 38,94 kat artarken, Mg<sup>++</sup> değerinin 0,67 kat, NH<sub>4</sub>N değerinin 0,60 kat azaldığı gözlemlenmiştir.

Çizelge 5.3'e göre Lojistik-LASSO yöntemi ile yapılan analiz sonucunda baz istasyon olan Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonundan Mudurnu Çayı Dere mahallesi Köprü istasyonuna kadar su kalitesinin değişiminde BOD<sub>5</sub>, EC, Mg<sup>++</sup>, Na<sup>+</sup>, pH, Pv ve TDS değişkenlerinin istatistiksel olarak anlamlı olduğu görülmektedir. Lojistik-Elastik ağ yöntemi ile yapılan analiz sonucunda ise baz istasyon olan Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonundan Mudurnu Çayı Dere mahallesi Köprü istasyonuna kadar su kalitesinin değişiminde BOD<sub>5</sub>, Ca<sup>++</sup>, Cl, DO, EC, M-Al, Na<sup>+</sup>, O-PO<sub>4</sub>, pH, Pv, SO<sub>4</sub>, T, TDS ve TH değişkenleri istatistiksel olarak anlamlıdır.

Anlamlı çıkan değişkenler için Odd's oranları hesaplanmış olup Çizelge 5.3' de verilmiştir. Çizelge 5.3'e göre Kocaçay Deresi Kayaboğazı Baraj Çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi Köprü istasyonunda Lojistik-Elastik ağ yöntemine göre DO 1,27 kat arttığı görülmektedir. Bu durum Mudurnu Çayı Dere mahallesi Köprü istasyonunda çözünmüş oksijen miktarının yüksek olduğunu ifade etmektedir ve bu da bu bölgede organik kirliliğin az olduğunu bir kanıttır

Lojistik LASSO yöntemiyle yapılan analiz sonucunda Kocaçay Deresi Kayaboğazı Baraj Çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi Köprü istasyonunda inorganik parametrelerden ve suyun sertliğini veren iyonlardan biri olan Mg<sup>++</sup> değerinin 0,39 kat azalmıştır. Benzer şekilde Lojistik-Elastik ağ yöntemiyle yapılan analizde de Kocaçay Deresi Kayaboğazı Baraj Çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi Köprü istasyonunda Mg<sup>++</sup> 0,42 kat azaldığı şeklinde yorum yapılabilir. Burada hareketle her iki yöntemle yapılan analiz sonucunda Mudurnu Çayı Dere mahallesi Köprü istasyonundaki suyun daha yumuşak olduğunu kanıtlar niteliktedir. Aynı zamanda Mg<sup>++</sup> değerinin Kocaçay Deresi Kayaboğazı Baraj Çıkışı istasyonunda daha az olması TH değerinin de az olmasını etkilemiştir.

Kocaçay Deresi Kayaboğazı Baraj Çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi Köprü istasyonunda Na<sup>+</sup> değerinin Lojistik-LASSO yöntemiyle yapılan analiz

sonucunda 3,48 kat artmıştır. Benzer şekilde Lojistik Elastik-ağ yöntemiyle yapılan analiz sonucunda 3,16 kat arttığı söylenebilir. Buradan hareketle Mudurnu Çayı Dere mahallesi köprü istasyonunda daha fazla Na<sup>+</sup> elementinin bulunduğu ve suyun daha tuzlu olduğu söylenebilir.

Sudaki çözünmüş maddeler, suda çözünmüş az miktardaki organik madde ve anorganik tuzların varlığından ileri gelir. Toplam çözünmüş katılar (TDS) suyun bulanıklılığını artırmaktadır. Toplam çözünmüş maddeler, doğal kaynaklardan, lağım atıklarından, şehir drenaj sularından ve endüstriyel sulardan ileri gelmektedir. Kocaçay Deresi Kayaboğazı Baraj Çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi Köprü istasyonunda Lojistik-LASSO yöntemiyle yapılan analiz sonucunda TDS değerinin 0,99 kat azalırken Lojistik-Elastik ağ yönteminde 0,98 kat azalmaktadır. Buradan her iki yöntem ile yapılan analize göre Mudurnu Çayı Dere mahallesi Köprü istasyonundaki suyun daha az bulanık olduğu şeklinde yorum yapılabilir.

Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonuna göre Mudurnu Çayı Dere mahallesi köprü istasyonunda O-PO<sub>4</sub> miktarının Lojistik-Elastik ağ yöntemiyle yapılan analiz sonucunda 12,67 kat arttığı gözlemlenmiştir. Mudurnu Çayı Dere mahallesi köprü istasyonunda daha fazla O-PO<sub>4</sub> bulunması ölçüm yapılan istasyon çevresinde bulunan tarlalarda fosforlu gübre kullanıldığını ve fazla gübrenin yağışla beraber dereye karıştığını göstermektedir.

Özetle, Lojistik-LASSO yöntemiyle yapılan analizler sonucunda Mg<sup>++</sup>, TDS, TH ve Na<sup>+</sup> değişkenleri istatistiksel olarak anlamlı çıkarken Lojistik-Elastik ağ yöntemiyle yapılan analiz sonucunda Lojistik-LASSO yöntemiyle anlamlı çıkan değişkenlere ek olarak DO ve O-PO<sub>4</sub> değişkenleri de istatistiksel olarak anlamlı çıkmıştır. Bunun nedeni lojistik-Elastik ağ yönteminin, lojistik regresyon kadar tüm değişkenlerin modelde yer almasını gerektirmeyen, ancak lojistik-LASSO yöntemi kadar da sert bir değişken seçim kriteri uygulamayan bir yöntem olmasıdır.

Üçüncü grup istasyonlardan Sakarya Nehri Sarıyar Baraj çıkışı ve Kocaçay Deresi Kayaboğazı Baraj çıkışına ait analiz sonuçları Çizelge 5.4'de verilmiştir.

Çizelge 5.4. Sarıyar Baraj Çıkışı ve Kayaboğazı Baraj Çıkışı İst. ait Analiz Sonuçları

Parametre	Lojistik Regresyon		Robust Lojistik Regresyon		Lojistik Ridge Regresyon $k = 0,52$		Lojistik LASSO Regresyon $\lambda = 0,32$		Elastik – Lojistik Regresyon $\lambda = 0,06$ $\alpha = 0,25$	
	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds
<b>Sabit</b>	14,78	-	10,12	-	13,92	-	-8,835	-	-6,446	-
<b>BOD5</b>	0,201*	1,22	0,020	1,02	0,015	1,02	0,312*	1,366	0,258*	1,294
<b>Ca++</b>	0,149*	1,16	1,149*	3,155	3,042	20,95	-	-	0,002*	1,002
<b>Cl-</b>	0,215*	1,24	0,605*	1,831	0,722*	2,058	-	-	0,002*	1,002
<b>DO</b>	-1,512	0,22	0,912*	2,489	-0,733*	2,081	-0,246*	0,78	-0,244	0,78
<b>EC</b>	0,023	1,02	0,504	1,655	0,733*	2,081	-	-	0,002	1,002
<b>K+</b>	0,330*	1,39	0,420	1,522	0,654*	1,923	-	-	0,229*	1,26
<b>M-Al</b>	-0,141	0,87	1,056	2,875	2,122*	8,35	-	-	0,005*	1,005
<b>Mg++</b>	-0,130	0,88	-0,051*	0,950	0,683*	1,979	0,211*	1,235	0,061*	1,06
<b>Na+</b>	0,440*	1,55	0,600*	1,822	0,774*	2,168	0,085*	1,09	0,002*	1,002
<b>NH4N</b>	-0,439	0,64	0,125*	1,133	0,730	2,08	0,481*	1,617	0,506*	1,658
<b>O-PO4</b>	0,326*	1,385	0,441*	1,554	0,552*	1,738	1,762*	5,824	1,583*	4,870
<b>Ph</b>	-1,831	0,16	-2,681	0,068	-4,439	0,012	-0,105*	0,90	-	0,400
<b>Pv</b>	-0,183*	0,83	0,307	1,359	0,593*	1,81	-	-	0,518*	1,681
<b>SO4</b>	-0,109	0,90	0,804*	2,234	0,708*	2,03	-	-	0,007	1,007
<b>T</b>	-0,384	0,68	-0,771	0,463	-0,830	0,44	0,048*	1,04	-	-
<b>TDS</b>	0,062	1,06	0,662*	1,939	0,782*	2,185	-0,016*	0,98	0,004*	1,004
<b>TH</b>	0,023	1,02	0,823*	2,278	5,891	361,77	0,107*	1,11	0,008*	1,008
<b>DSO</b>	<b>34,1</b>		<b>45,5</b>		<b>69,4</b>		<b>77,3</b>		<b>91,3</b>	

(•0.05 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.)

İlk olarak Çizelge 5.4 incelendiğinde doğru sınıflandırma yüzdesi bakımından en geçerli modelin lojistik elastik ağ yöntemiyle elde edilen model olduğu görülmektedir.

Lojistik regresyon yöntemiyle yapılan analizler sonucunda baz istasyon olan Kocaçay Deresi Kayaboğazı Baraj Çıkışı istasyonundan Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna kadar su kalitesinin değişiminde BOD<sub>5</sub>, Ca<sup>++</sup>, Cl<sup>-</sup>, K<sup>+</sup>, Na<sup>+</sup>, O-PO<sub>4</sub> ve pV değişkenleri istatistiksel açıdan anlamlı bulunmuştur. Ancak lojistik regresyon analizinde tüm değişkenlerin modelde bulunması gerekir. Bu amaçla tüm değişkenler için Odd's oranları hesaplanmış olup Çizelge 5.4'te verilmiştir. Odd's oranları incelendiğinde, Kocaçay Deresi Kayaboğazı Baraj Çıkışı istasyonuna göre Sakarya Nehri Sarıyar Baraj Çıkışı

istasyonunda BOD<sub>5</sub> deęerinin 1,22 kat, Ca<sup>++</sup> deęerinin 1,16 kat, Cl<sup>-</sup> deęerinin 1,24 kat, EC deęerinin 1,02 kat, K<sup>+</sup> deęerinin 1,39 kat, Na<sup>+</sup> deęerinin 1,55 kat, O-PO<sub>4</sub> deęerinin 1,385 kat, TDS deęerinin 1,06 kat, TH deęerinin 1,02 kat artarken, DO deęerinin 0,22 kat, M-Al deęerinin 0,87 kat, Mg<sup>++</sup> deęerinin 0,88 kat, NH<sub>4</sub>N deęerinin 0,64 kat, pH deęerinin 0,16 kat, Pv deęerinin 0,83 kat, SO<sub>4</sub> deęerinin 0,90 kat, T deęerinin 0,68 kat azaldığı gözlemlenmiştir.

Robust lojistik regresyon yöntemiyle yapılan analiz sonucunda istasyonunda BOD<sub>5</sub> deęerinin 1,02 kat, Ca<sup>++</sup> deęerinin 3,155 kat, Cl<sup>-</sup> deęerinin 1,831 kat, EC deęerinin 1,655 kat, K<sup>+</sup> deęerinin 1,522 kat, Na<sup>+</sup> deęerinin 1,822 kat, NH<sub>4</sub>N deęerinin 1,133 kat, O-PO<sub>4</sub> deęerinin 1,554 kat, TDS deęerinin 1,939 kat, TH deęerinin 2,278 kat, M-Al deęerinin 2,875, kat, Pv deęerinin 1,359 kat, SO<sub>4</sub> deęerinin 2,234 kat artarken, DO deęerinin 0,22 kat, Mg<sup>++</sup> deęerinin 0,950 kat, pH deęerinin 0,068 kat, T deęerinin 0,463 kat azaldığı gözlemlenmiştir.

Lojistik Ridge yöntemiyle yapılan analizler sonucunda baz istasyon olan Kocaçay Deresi Kayaboęazı Baraj Çıkışı istasyonundan Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna kadar Cl<sup>-</sup>, DO, EC, K<sup>+</sup>, M-Al, Mg<sup>++</sup>, Na<sup>+</sup>, O-PO<sub>4</sub>, pV, SO<sub>4</sub> ve TDS deęişkenlerinin istatistiksel olarak anlamlı olduęu Çizelge 5.4' da görölmektedir. Odd's oranları incelendiğinde, BOD<sub>5</sub> deęerinin 1,02 kat, DO deęerinin 2,081 kat, Cl<sup>-</sup> deęerinin 2,058 kat, EC deęerinin 2,081 kat, K<sup>+</sup> deęerinin 1,923 kat, M-Al deęerinin 8,35 kati Mg<sup>++</sup> 1,979 kat, Na<sup>+</sup> deęerinin 2,168 kat artarken, pH deęerinin 0,012 kat, T deęerinin 0,44 kat azaldığı gözlemlenmiştir.

Lojistik LASSO analizi sonucunda, baz istasyon olan Kocaçay Deresi Kayaboęazı Baraj Çıkışı istasyonundan Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna kadar su kalitesinin deęişiminde BOD<sub>5</sub>, DO, Mg<sup>++</sup>, Na<sup>+</sup>, NH<sub>4</sub>N, O-PO<sub>4</sub>, pH, T, TDS ve TH deęişkenleri istatistiksel olarak anlamlı bulunmuştur. Benzer şekilde lojistik Elastik ağ yöntemiyle yapılan analiz sonucunda baz istasyon olan Kocaçay Deresi Kayaboęazı Baraj Çıkışı istasyonundan Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna kadar su kalitesinin deęişiminde Lojistik LASSO yönteminde anlamlı çıkan deęişkenlere ek olarak Ca<sup>++</sup>, Cl<sup>-</sup>, EC, K<sup>+</sup>, M-Al, Mg<sup>++</sup>, Na<sup>+</sup>, pV ve SO<sub>4</sub> istatistiksel olarak anlamlı olduęu bulunmuştur.

Bunun nedeni lojistik-Elastik ağ yönteminin, lojistik regresyon kadar tüm değişkenlerin modelde yer almasını gerektirmeyen, ancak lojistik-LASSO yöntemi kadar da sert bir değişken seçim kriteri uygulamayan bir yöntem olmasıdır.

Anlamli çıkan deęişkenler için Odd's oranları hesaplanmış olup Çizelge 5.4' de verilmiştir. Odd's oranları incelendiğinde, Lojistik-LASSO ve Lojistik-Elastik Ağ yöntemiyle yapılan analiz sonucunda Kocaçay Deresi Kayaboęazı Baraj çıkışı istasyonuna göre Sakarya Nehri Sarıyar Baraj Çıkışı istasyonunda DO deęerinin 0,78 kat azalmıştır. Bu durumda Kocaçay Deresi Kayaboęazı Baraj Çıkışı istasyonundaki suda organik kirlilięin daha az olduęu şeklinde yorum yapılabilir.

Suların sertlięi, su içerisinde çözünmüş olarak bulunan toplam  $Ca^{++}$  ve  $Mg^{++}$  miktarının kalsiyum karbonat ( $CaCO_3$ ) eş deęeri olarak tanımlanmıştır. Bunun sebebi, sudaki mevcut iyonlara göre daha fazla oranda  $Ca^{++}$  ve  $Mg^{++}$  iyonlarının bulunmasıdır (Şengün, 2013). Yapılan analizler sonucunda  $Ca^{++}$  deęerinin Lojistik-Elastik Ağ yöntemiyle yapılan analizler sonucunda  $Ca^{++}$  deęerinin Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Kocaçay Deresi Kayaboęazı Baraj çıkışı istasyonunda 1,002 kat arttıęı ve  $Mg^{++}$  deęerinin 1,06 kat arttıęı söylenmektedir. Yani, Kocaçay Deresi Kayaboęazı Baraj Çıkışı istasyonundaki suda  $Ca^{++}$  ve  $Mg^{++}$  iyonlarının daha fazla olduęu söylenir ve bu istasyondaki suyun Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre daha sert olduęu şeklinde yorum yapılabilir. Analiz sonuçlarına bakıldığında kalsiyumun pozitif işaretli çıkması başka bir su kalite deęişkeni olan TH pozitif işaretli çıkmasına sebep olmuştur. Buradan TH için Lojistik-Elastik ağ yöntemiyle yapılan analiz sonuçlarına bakıldığında Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Kocaçay Deresi Kayaboęazı Baraj çıkışı istasyonunda TH deęerinin 1,008 kat arttıęı şeklinde yorum yapılır. Yukarıda kalsiyum için yapılan yorum toplam sertliğinde arttıęını destekler niteliktedir.

Yüksek klorür konsatrasyonu bir kirlenme belirtisi olmakla birlikte önemli bir kalite parametresidir (Tuncay 1983). Lojistik-Elastik ağ yöntemiyle yapılan analiz sonucunda Cl- deęerinin Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Kocaçay Deresi Kayaboęazı Baraj çıkışı istasyonunda 1,002 kat artmıştır. Bu da Kocaçay Deresi Kayaboęazı Baraj çıkışı istasyonun daha kirli olduęunun bir göstergesidir.

Çizelge 5.4'e göre Lojistik-LASSO yöntemiyle yapılan analiz sonucunda EC değeri Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonunda 1,002 kat artmıştır. Bu da Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonunda iyon miktarının daha fazla olduğunu göstermektedir. Elektriksel iletkenliğin fazla olduğu Cl<sup>-</sup>, Mg<sup>++</sup>, ve Na<sup>+</sup> miktarlarının da fazla olduğunu destekler niteliktedir. Sodyum (Na<sup>+</sup>) suda kolaylıkla çözünen ve suya tuzlu bir tat veren bir katyondur. Na<sup>+</sup> içeren sular tarımsal sulamalarda toprak ve bitkiler üzerinde zararlı etkiler yapmaktadır (Tuncay,1983). Buradan Lojistik-Elastik ağ yöntemiyle yapılan analiz sonucunda Na<sup>+</sup> değerinin Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonunda 1,002 kat arttığı söylenir. Lojistik-LASSO yöntemiyle yapılan analiz sonucunda Na<sup>+</sup> değerinin Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonunda 1,09 kat arttığı görülmektedir. Buradan her iki yöneme göre de Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonundaki suyun daha tuzlu olduğunu göstermektedir. Sudaki çözülmüş maddeler, suda çözülmüş az miktardaki organik madde ve anorganik tuzların varlığından ileri gelir. TDS suyun bulanıklılığını artırmaktadır. Toplam çözülmüş maddeler, doğal kaynaklardan, lağım atıklarından, şehir drenaj sularından ve endüstriyel sulardan ileri gelmektedir. Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonunda Lojistik-Elastik ağ yöntemiyle yapılan analiz sonucunda TDS değerinin 1,004 kat artarken Lojistik-LASSO yönteminde 0,98 kat azaldığı şeklinde yorumlanır. Buradan elektriksel iletkenliğin toplam çözülmüş katılardaki değişimi etkilediğinin bir kanıtıdır.

O-PO<sub>4</sub> için analiz sonuçları incelendiğinde Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonunda bu değişkenin Lojistik-Elastik ağ yöntemi için 1,79 kat arttığı görülmektedir. Yani, Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonunda daha fazla orto fosfat bulunması ölçüm yapılan istasyon çevresinde bulunan tarlalarda fosforlu gübre kullanıldığını ve fazla gübrenin yağışla beraber dereye karıştığını göstermektedir. Lojistik-Elastik ağ yöntemiyle yapılan analizde Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonunda SO<sub>4</sub> değerinin 1,007 kat arttığı söylenebilir.

K<sup>+</sup> iç sularında tuzluluğu oluşturan katyon grubundan biridir. Yapılan analiz sonuçları incelendiğinde potasyum değerinin Lojistik-Elastik ağ yöntemiyle yapılan analiz sonucunda Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonunda 1,26 kat arttığı söylenmektedir. M-Al değerinin Sakarya Nehri Sarıyar Baraj Çıkışı istasyonuna göre Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonunda 1,005 kat arttığı söylenir.

Su içindeki hidrojen iyonu konsantrasyonunu 10 tabanına göre negatif logaritması Ph değeri olarak tanımlanır. Ph=7 olan sular nötr olarak bilinir. Ph değeri 7'nin altındaysa su asidik karakter kazanırken Ph 7'nin üzerinde ise su bazik karakter kazanır. Sakarya Nehri Sarıyar Baraj Çıkışı istasyonu ve Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonu karşılaştırılırken Lojistik-LASSO yöntemiyle yapılan analiz sonucunda Kocaçay Deresi Kayaboğazı Baraj çıkışı istasyonunda Ph değerinin 0,90 kat daha az iken Lojistik-Elastik ağ yöntemiyle yapılan analiz sonucunda 0,40 kat daha az olduğu söylenmektedir.

Özetle, Lojistik-LASSO yöntemiyle yapılan analiz sonucunda ile iki istasyon arasındaki değişime neden olan değişkenler DO, Na<sup>+</sup>, Ph ve TDS karakteristikleridir. Bu karakteristiklerin yanı sıra Lojistik-Elastik ağ yöntemiyle yapılan analiz sonucunda Ca<sup>++</sup>, Cl<sup>-</sup>, EC, K<sup>+</sup>, M-Al, Mg<sup>++</sup>, Na<sup>+</sup>, O-PO<sub>4</sub>, Pv, SO<sub>4</sub> ve TH karakteristikleri de iki istasyon arasındaki değişime neden değişkenlerdendir. Doğru sınıflandırma yüzdesi bakımından Lojistik-Elastik ağ yönteminin iki istasyon arasındaki değişimi en iyi açıklayan model olduğu söylenebilir.

Üçüncü grup istasyonlardan seçilen Çubuk Çayı Çubuk-1 Baraj Çıkışı ve Kuruçay Baraj Aksı'na ait analiz sonuçları Çizelge 5.5'te verilmiştir.



Çizelge 5.5. Çubuk 1 Baraj Çıkışı ve Kuruçay Baraj Aksı İst. ait Analiz Sonuçları

Parametre	Lojistik Regresyon		Robust Lojistik Regresyon		Lojistik Ridge Regresyon $k = 0, 17$		Lojistik LASSO Regresyon $\lambda = 0, 29$		Elastik – Lojistik Regresyon $\lambda = 0, 16$ $\alpha = 0, 23$	
	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds
Sabit	-188,7	-	-129,7	-	-192,41	-	7,764*	-	-2,215	-
BOD <sub>5</sub>	-1,260*	0,28	-0,912*	0,402	-0,691*	0,501	-0,205*	0,81	-0,396*	0,67
Ca <sup>++</sup>	0,007	1,01	0,012	1,013	3,965*	52,72	-	-	0,020*	1,02
Cl <sup>-</sup>	0,289*	1,34	0,319*	1,376	-3,732	0,02	-	-	-0,009	0,99
DO	-3,110	0,05	-1,114	0,328	-0,197*	0,08	-2,174*	0,113	-0,826	0,438
EC	-0,025	0,98	-0,507*	0,602	-0,671*	0,511	-	-	-0,004*	1,00
K <sup>+</sup>	-1,592*	0,20	-0,665	0,514	-0,718	0,487	-0,044*	0,96	-0,340*	0,72
M-Al	0,077	1,08	1,217*	3,378	3,060	21,33	-	-	0,010*	1,01
Mg <sup>++</sup>	0,991*	2,69	0,521*	1,684	-0,591*	0,553	-0,063*	0,94	-0,135*	0,87
Na <sup>+</sup>	-1,048	0,35	-0,955	0,385	-0,847	0,429	-0,099*	0,91	-0,049*	0,95
NH <sub>4</sub> N	-0,602	0,547	-0,754	0,470	-0,884*	0,413	-1,045*	0,35	-1,332	0,26
O-PO <sub>4</sub>	0,559*	1,748	0,781*	2,184	-0,868*	0,419	-0,029*	0,97	-2,066*	0,13
Ph	3,334*	28,05	2,916*	18,47	2,469*	11,81	1,251*	3,493	2,502*	12,21
Pv	3,289*	26,82	1,245*	3,473	-0,613*	0,541	-0,541	0,582	-0,963*	0,38
SO <sub>4</sub>	-0,216	0,81	-2,223	0,108	-4,267	0,01	-	-	-0,027*	0,97
T	-1,371	0,25	-0,947	0,388	-0,469	0,625	-0,302*	0,740	-0,136*	0,87
TDS	0,065*	1,07	0,341*	1,041	-0,403*	0,668	-	-	-0,003	1,00
TH	0,048	1,05	-0,219	0,803	-0,439	0,644	-	-	-0,007*	0,99
DSO	45,7		65,2		74,8		71,0		83,8	

(•0.05 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.)

Lojistik regresyon analizi sonucunda elde edilen modelin en düşük doğru sınıflandırma yüzdesine sahip olduğu, lojistik elastik ağ yöntemiyle elde edilen modelin % 83,8'lik doğru sınıflandırma oranı ile en yüksek doğru sınıflandırma oranına sahip olduğu görülmektedir.

Çizelge 5.5 incelendiğinde Lojistik regresyon yöntemiyle ile yapılan analiz sonucunda Odds oranları incelendiğinde Çubuk 1 Baraj çıkışı olma olasılığının Kuruçay Baraj Aksı olma olasılığına göre, BOD<sub>5</sub> değerinin 0,28 kat, DO değerinin 0,05 kat, EC değerinin 0,98 kat, K<sup>+</sup> değerinin 0,20 kat, Na<sup>+</sup> değerinin 0,35 kat, NH<sub>4</sub>N değerinin 0,547

kat, SO<sub>4</sub> deęerinin 0,81 kat, T deęerinin 0,25 kat azalırken, Ca<sup>++</sup> deęerinin 1,01 kat, Cl<sup>-</sup> deęerinin 1,34, M-Al deęerinin 1,08 kat, Mg<sup>++</sup> deęerinin 2,69 kat, O-PO<sub>4</sub> deęerinin 269,35 kat, Ph deęerinin 28,05 kat, Pv deęerinin 26,82 kat, TDS deęerinin 1,07 kat ve TH deęerinin 1,05 kat arttıęı řeklinde yorumlanır.

Saęlam lojistik regresyon yöntemiyle yapılan analiz sonucunda BOD<sub>5</sub> deęerinin 0,402 kat, DO deęerinin 0,328 kat, EC deęerinin 0,602 kat, K<sup>+</sup> deęerinin 0,514 kat, Na<sup>+</sup> deęerinin 0,385 kat, NH<sub>4</sub>N deęerinin 0,470 kat, SO<sub>4</sub> deęerinin 0,108 kat, T deęerinin 0,388 kat, TH deęerinin 0,803 kat azalırken, Ca<sup>++</sup> deęerinin 1,013 kat, Cl<sup>-</sup> deęerinin 1,34, M-Al deęerinin 3,378 kat, Mg<sup>++</sup> deęerinin 1,684 kat, O-PO<sub>4</sub> deęerinin 2,184 kat, Ph deęerinin 18,47 kat, Pv deęerinin 3,473 kat ve TDS deęerinin 1,041 kat arttıęı řeklinde yorumlanır

Lojistik-Ridge regresyon yöntemiyle yapılan analizler sonucunda Odds oranları incelendięinde BOD<sub>5</sub> deęerinin 0,501 kat azaldıęı, Ca<sup>++</sup> deęerinin 52,72 kat arttıęı, Cl<sup>-</sup> deęerinin 0,02 kat azaldıęı, DO deęerinin 0,08 kat azaldıęı, EC deęerinin 0,511 kat azaldıęı, K<sup>+</sup> deęerinin 0,487 kat azaldıęı, M-Al deęerinin 21,33 kat arttıęı, Mg<sup>++</sup> deęerinin 0,553 kat azaldıęı, Na<sup>+</sup> deęerinin 0,429 kat azaldıęı, NH<sub>4</sub>N deęerinin 0,413 kat azaldıęı, O-PO<sub>4</sub> deęerinin 0,419 kat azaldıęı, Ph deęerinin 11,81 kat arttıęı, Pv deęerinin 0,541 kat azaldıęı, SO<sub>4</sub> deęerinin 1,01 kat arttıęı, T deęerinin 0,625 kat azaldıęı, TDS deęerinin 0,668 kat azaldıęı ve TH deęerinin 0,644 kat azaldıęı řeklinde yorumlanabilir.

BOD<sub>5</sub> sulardaki organik maddelerin mikroroganizmalar tarafından parçalanması için gerekli olan oksijen miktarını ifade etmektedir. Harcanan oksijen miktarı fazlaysa ölçüm yapılan istasyondaki organik yükün de fazla olduęunu göstermektedir. Bu durumda ölçüm yapılan istasyondaki organik kirlilięin yüksek olduęunu belirtir. Çubuk Çayı Çubuk 1 Baraj giriři istasyonu olma olasılıęının Kuruçay Baraj aksı istasyonu olma olasılıęına göre, BOD<sub>5</sub> deęerinin Lojistik-LASSO yöntemiyle yapılan analiz sonuçları göz önüne alındıęında 0,81 kat azaldıęı ve Lojistik-Elastik Ağ yöntemiyle yapılan analiz sonucuna göre ise 0,67 kat azaldıęı yani her iki yönteme göre de Kuruçay Baraj Aksı istasyonundaki suda organik yükün daha az olduęu ve Çubuk Çayı Çubuk 1 Baraj giriři istasyonundaki suya göre organik

kirliliğin az olduğu söylenmektedir. Çubuk Çayı Çubuk 1 Baraj girişi istasyonu çevresinde bulunan kanalizasyonlar buradaki suyun daha kirli olduğunun bir kanıtıdır.

İç sularda tuzluluk dört katyon grubu olan sodyum ( $\text{Na}^+$ ), kalsiyum ( $\text{Ca}^{++}$ ), magnezyum ( $\text{Mg}^{++}$ ), potasyumdan ( $\text{K}^+$ ) oluşur. Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj aksı istasyonunda  $\text{K}^+$  değerinin Lojistik-LASSO yöntemiyle yapılan analiz sonuçlarına göre 0,96 kat azaldığı ve Lojistik-Elastik Ağ yöntemiyle yapılan analiz sonucuna göre ise 0,67 kat azaldığı görülmektedir. Sodyum ise suda kolaylıkla çözünen ve suya tuzlu bir tat veren bir katyondur.  $\text{Na}^+$  içeren sular tarımsal sulamalarda toprak ve bitkiler üzerinde zararlı etkiler yapmaktadır (Tuncay,1983). Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj aksı istasyonunda  $\text{Na}^+$  değerinin Lojistik-LASSO yöntemiyle yapılan analiz sonuçlarına göre 0,91 kat azaldığı ve Lojistik-Elastik Ağ yöntemiyle yapılan analiz sonucuna göre ise 0,95 kat azaldığı görülmektedir. Kalsiyum ( $\text{Ca}^{++}$ ) için analiz sonuçları incelendiğinde Lojistik-Elastik ağ yöntemine göre  $\text{Ca}^{++}$  değerinin Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj aksı istasyonunda 1,02 kat arttığı görülmektedir. Magnezyum aynı zamanda suların sertliğini belirleyen iyonlardan biridir. Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj Aksı istasyonunda  $\text{Mg}^{++}$  değerinin Lojistik-LASSO yöntemiyle yapılan analiz sonuçlarına göre 0,94 kat azaldığı ve Lojistik-Elastik Ağ yöntemiyle yapılan analiz sonucuna göre ise 0,87 kat azaldığı görülmektedir. Buradan her iki yönteme göre de Kuruçay baraj Aksı istasyonunda ölçülen suyun daha yumuşak olduğu yorumu yapılabilir. Bu durum suyun TH karakteristiğini de etkilemektedir. Toplam sertlik için analiz sonuçları incelendiğinde Lojistik-Elastik ağ yöntemine göre, Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj Aksı istasyonunda 0,99 kat azaldığı görülmektedir. Bu durum yukarıda magnezyum için yapılan yorumu destekler niteliktedir.

Su içindeki hidrojen iyonu konsantrasyonunu 10 tabanına göre negatif logaritması Ph değeri olarak tanımlanır. Ph=7 olan sular nötr olarak bilinir. Ph değeri 7'nin altındaysa su asidik karakter kazanırken Ph 7'nin üzerinde ise su bazik karakter kazanır. Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj Aksı istasyonunda Ph değeri 12,21 kat artmıştır.

NH<sub>4</sub>N genellikle çözülmüş oksijenden sonra ikinci önemli su kalite parametresidir. Sudaki NH<sub>4</sub>N, organik maddelerin bozunması, özellikle organik gübre veya inorganik amonyum kaynaklı kimyasal gübreleme sonucunda, evsel ve endüstriyel kirlenmelerden oluşabilir (Egemen ve Sunlu,1999). Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj aksı istasyonunda NH<sub>4</sub>N değerinin Lojistik-LASSO yöntemiyle yapılan analiz sonuçlarına göre 0,35 kat azaldığı ve Lojistik-Elastik Ağ yöntemiyle yapılan analiz sonucuna göre ise 0,26 kat azaldığı görülmektedir. Bu durumda Kuruçay Baraj Aksı istasyonunun daha temiz olduğu şeklinde yorum getirilebilir. Çubuk Çayı Çubuk 1 Baraj girişi istasyonu çevresinde bulunan yerleşim yerleri ve kanalizasyonlar bu istasyondaki suyun daha kirli olduğunun bir kanıtıdır. Aynı zamanda yüksek klorür konsantrasyonu bir kirlenme belirtisi olmakla birlikte önemli bir kalite parametresidir (Tuncay 1983). Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj aksı istasyonunda Cl<sup>-</sup> değerinin Lojistik-Elastik Ağ yöntemiyle yapılan analiz sonucuna göre ise 0,13 kat azaldığı görülmektedir. Yine Kuruçay Baraj Aksındaki suyun daha temiz olduğu yorumu Cl<sup>-</sup> karakteristiği ile de desteklenmektedir.

Orto fosfat (O-PO<sub>4</sub>) için analiz sonuçları incelendiğinde Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj aksı istasyonunda O-PO<sub>4</sub> değerinin Lojistik-LASSO yöntemiyle yapılan analiz sonuçlarına göre 0,97 kat azaldığı ve Lojistik-Elastik Ağ yöntemiyle yapılan analiz sonucuna göre ise 0,13 kat azaldığı görülmektedir. Yani, Kuruçay Baraj aksı istasyonunda daha az orto fosfat bulunması ölçüm yapılan istasyon çevresinde bulunan tarlalarda fosforlu gübrenin daha az kullanıldığı yorumunu desteklemektedir.

Çizelge 5.5'e göre Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj aksı istasyonunda EC değerinin Lojistik-Elastik Ağ yöntemiyle yapılan analiz sonucuna göre ise 1 kat arttığı görülmektedir. Yani Kuruçay Baraj Aksındaki suda iyon miktarının daha fazla olduğunu göstermektedir. Buradan elektriksel iletkenliğinin daha fazla olduğu şeklinde yorum yapılabilir. Sudaki çözülmüş maddeler, suda çözülmüş az miktardaki organik madde ve anorganik tuzların varlığından ileri gelir. TDS suyun bulanıklılığını artırmaktadır. Toplam çözülmüş maddeler, doğal kaynaklardan, lağım atıklarından, şehir drenaj sularından ve endüstriyel sulardan ileri gelmektedir. Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj aksı istasyonunda TDS değerinin Lojistik-Elastik Ağ

yöntemiyle yapılan analiz sonucuna göre ise 1 kat arttığı görülmektedir. Yukarıda belirtildiği gibi elektriksel iletkenlik toplam çözünmüş katılardaki değişimi ifade ettiğinden dolayı Kuruçay Baraj aksındaki EC'nin fazla olması TDS'nin fazla olmasını da etkilemiştir. Aynı zamanda Kuruçay baraj Aksı istasyonundaki suda TDS değerinin daha fazla olması Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre daha bulanık olduğunun bir kanıtıdır.

Değişik sanayilerden atılan atıklarda ve minerallerin kavrulması işleminden sonra sülfat sulara verilir. Suda sülfat genellikle yüksek konsantrasyonlarda bulunabilir. Çünkü, kayalardan çözülmeye geçen katyonlar genellikle sülfatla, çözünebilir bileşikler verirler. Lojistik-Elastik ağ yöntemiyle yapılan analizde Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj aksı istasyonunda  $SO_4$  değerinin 0,97 kat azaldığı söylenmektedir. Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Kuruçay Baraj aksı istasyonunda Lojistik-Elastik ağ yöntemiyle yapılan analizde T değerinin 0,87 kat azaldığı görülmektedir.

Özetle, Çubuk Çayı Çubuk 1 Baraj girişi istasyonu ve Kuruçay Baraj aksı istasyonundaki arasındaki değişime neden su karakteristikleri belirlenirken kullanılan Lojistik-LASSO ve Lojistik-Elastik ağ yöntemleri karşılaştırılırken Lojistik-LASSO yöntemi ile  $BOD_5$ ,  $Ca^{++}$ , DO,  $K^+$ ,  $Mg^+$ ,  $Na^+$ ,  $NH_4N$ , pH, T ve  $O-PO_4$  karakteristikleri ile açıklamaktadır. Lojistik-Elastik ağ yönteminde ise  $BOD_5$ ,  $Ca^{++}$ , EC,  $K^+$ , M-Al,  $Mg^{++}$ ,  $Na^+$ , Pv,  $SO_4$ , T, TDS, TH,  $Ca^{++}$ , M-Ali  $O-PO_4$  ve pH karakteristikleri de anlamlı çıkmıştır.

Çubuk 1 Baraj Çıkışı ve Bayındır Barajı çıkışı istasyonlarına ait analiz sonuçları Çizelge 5.6' da verilmiştir.

Çizelge 5.6. Çubuk 1 Baraj Çıkışı ve Bayındır Barajı Çıkışı İst. Analiz Sonuçları

Parametre	Lojistik Regresyon		Robust Lojistik Regresyon		Lojistik Ridge Regresyon $k = 0,42$		Lojistik LASSO Regresyon $\lambda = 0,47$		Elastik – Lojistik Regresyon $\lambda = 0,29$ $\alpha = 0,41$	
	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds
<b>Sabit</b>	-46,93	-	-39,43	-	-42,21	-	-3,081	-	-3,656	-
<b>BOD5</b>	0,164	1,18	0,964	2,622	2,184	8,88	0,019	0,02	0,074	1,08
<b>Ca++</b>	0,182*	1,20	0,452*	1,571	0,533*	1,704	0,047*	1,05	0,118*	1,13
<b>Cl-</b>	0,187*	1,21	0,217*	1,242	-0,950	0,39	0,021*	1,02	0,040*	1,04
<b>DO</b>	-1,413*	0,24	-0,841*	0,431	-0,513*	0,598	-0,498*	0,61	-0,953*	0,39
<b>EC</b>	-0,060	1,06	-0,984*	0,374	-3,451*	0,03	-	-	-	-
<b>K+</b>	-2,112	0,12	-1,962	0,141	-3,255*	0,04	-0,377*	0,69	-0,658*	0,52
<b>M-Al</b>	0,083	1,09	1,657	0,191	2,921	18,56	0,012	1,01	0,015*	1,02
<b>Mg++</b>	-0,491*	1,633	-0,542*	0,582	-0,506*	1,658	-0,573*	0,56	-0,606	0,55
<b>Na+</b>	-0,889	0,41	-2,224*	0,108	-3,786*	0,02	-0,073	0,93	-0,093*	0,91
<b>NH4N</b>	0,573*	1,773	0,728	2,071	0,892	2,44	0,401*	1,49	1,069*	2,91
<b>O-PO4</b>	-2,651	14,161	-1,971	1,139	-0,661	0,516	-3,630	0,03	-4,522	0,01
<b>Ph</b>	0,561	1,752	-0,912	0,402	-0,896	0,41	0,971*	2,64	1,724*	5,61
<b>Pv</b>	0,851	2,341	2,761*	15,82	3,732*	41,76	1,212*	3,36	1,083*	2,95
<b>SO4</b>	-1,006*	0,37	-0,925*	0,396	-0,629*	0,533	-	-	-	-
<b>T</b>	-0,181	0,83	0,048	1,049	0,302*	1,35	-	-	0,007*	1,01
<b>TDS</b>	0,197*	1,22	0,216*	1,241	-2,218	0,11	-	-	-	-
<b>TH</b>	0,445	1,56	0,613	1,846	0,933	2,54	-	-	0,059*	1,06
<b>DSO</b>	<b>53,4</b>		<b>66,1</b>		<b>77,2</b>		<b>81,4</b>		<b>94,3</b>	

(•0.05 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.)

İlk olarak Çizelge 5.6 incelendiğinde lojistik LASSO ve lojistik-elastik ağ yöntemlerinin doğru sınıflandırma yüzdelerinin oldukça yüksek olduğu görülmektedir. Lojistik regresyon yöntemiyle yapılan analiz sonucunda Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda BOD<sub>5</sub> değerinin 1,18 kat, Ca<sup>++</sup> değerinin 1,20 kat, Cl<sup>-</sup> değerinin 1,21 kat, EC değerinin 1,06 kat, M-Al değerinin 1,09 kat, NH<sub>4</sub>N değerinin 1,773 kat, Ph değerinin 1,752 kat, Pv değerinin 2,341 kat, TDS değerinin 1,22 kat ve TH değerinin 1,56 kat, Mg<sup>++</sup> değerinin 1,633 kat, O-PO<sub>4</sub> değerinin 14,161 kat artarken, DO değerinin 0,24 kat, K<sup>+</sup> değerinin 0,12 kat, Na<sup>+</sup> değerinin 0,41 kat, SO<sub>4</sub> değerinin 0,37 kat, T değerinin 0,83 kat azaldığı şeklinde yorum yapılabilir.

Sağlam lojistik regresyon analizi yöntemiyle yapılan analiz sonucunda, BOD<sub>5</sub> değerinin 2,622 kat, Ca<sup>++</sup> değerinin 1,571 kat, Cl<sup>-</sup> değerinin 1,242 kat, NH<sub>4</sub>N değerinin 2,071 kat, Pv değerinin 15,82 kat, TDS değerinin 1,241 kat ve TH değerinin 1,846 kat, O-PO<sub>4</sub> değerinin 1,139 kat, T değerinin 1,049 kat artarken, DO değerinin 0,431 kat, EC değerinin 0,374 kat, K<sup>+</sup> değerinin 0,141 kat, Mg<sup>++</sup> değerinin 0,582 kat, M-Al değerinin 0,191 kat, Na<sup>+</sup> değerinin 0,108 kat, Ph değerinin 0,402 kat, SO<sub>4</sub> değerinin 0,396 kat azaldığı şeklinde yorum yapılabilir

Çizelge 5.6' da Lojistik-Ridge regresyon yöntemiyle yapılan analiz sonuçlarına göre Odds oranları incelendiğinde, Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda BOD<sub>5</sub> değerinin 8,88 kat, Ca<sup>++</sup> değerinin 208,30 kat, M-Al değerinin 18,56 kat, NH<sub>4</sub>N değerinin 2,44 kat, , Pv değerinin 41,76 kat, T değerinin 1,35 kat, TH değerinin 2,54 kat artarken, Cl<sup>-</sup> değerinin 0,39 kat, DO değerinin  $5,87 \times 10^{-3}$  kat, EC değerinin 0,03 kat, K<sup>+</sup> değerinin 0,04 kat, Mg<sup>++</sup> değerinin  $3,69 \times 10^{-3}$  kat, Na<sup>+</sup> değerinin 0,02 kat, O-PO<sub>4</sub> değerinin  $1,35 \times 10^{-3}$  kat, Ph değerinin 0,41 kat, SO<sub>4</sub> değerinin  $1,85 \times 10^{-3}$  kat, TDS değerinin 0,11 kat azaldığı şeklinde yorum yapılabilir.

Çizelge 5.6 incelendiğinde, Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda BOD<sub>5</sub> değerinin Lojistik-LASSO yöntemine göre 0,02 kat azalırken Lojistik-Elastik ağ yöntemine göre 1,08 kat arttığı sonucuna ulaşılmıştır. Buradan Lojistik LASSO yöntemine göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda BOD<sub>5</sub> değerinin daha az olduğu yani bu istasyonda organik kirliliğin daha az olduğunu göstermektedir. Lojistik-Elastik ağ yöntemiyle yapılan analize göre de Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda BOD<sub>5</sub> değerinin daha fazla olduğu yani bu istasyonda organik kirliliğin daha fazla olduğunu göstermektedir. Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda Lojistik-LASSO yöntemine göre DO değerinin 0,61 kat azaldığı ve Lojistik-Elastik Ağ yöntemine göre de 0,39 kat azaldığı görülmektedir. Her iki yönteme göre de Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda DO değerinin daha az olduğu yani organik kirliliğin az olduğunun bir göstergesidir. Yukarıda BOD<sub>5</sub> değeriyle ilgili yapılan yorumlar DO değeriyle ilgili yapılan yorumları destekler niteliktedir. Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda Lojistik-

LASSO yöntemine göre EC değerinin 1 kat arttığı ve Lojistik-Elastik Ağ yöntemine göre de 0,99 kat azaldığı görülmektedir. Lojistik-LASSO yöntemine göre Çubuk Çayı Çubuk 1 Baraj girişi istasyonundaki suda iyon miktarının daha az olduğunu göstermektedir. Lojistik-Elastik ağ yöntemine göre de EC değerinin Bayındır Deresi Bayındır Barajı Çıkışı istasyonundaki suda iyon miktarının az olduğunu göstermektedir. Elektriksel iletkenliğin az olması  $Mg^{++}$ , ve  $Na^{+}$  miktarlarının da az olduğunu göstermektedir.

Çizelge 5.6' de Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda  $Na^{+}$  değerinin Lojistik-LASSO yöntemine göre 0,93 kat azaldığı ve Lojistik-Elastik ağ yöntemine göre de 0,91 kat azaldığı görülmektedir. Her iki yöntemine göre de Bayındır Deresi Bayındır Barajı Çıkışı istasyonundaki suyun daha az tuzlu olduğu söylenebilir.  $K^{+}$  iç sularda tuzluluğu oluşturan katyon grubundan biridir. Analiz sonuçları incelendiğinde Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda Lojistik-LASSO yöntemine göre 0,69 kat azaldığı ve Lojistik-Elastik ağ yöntemine göre de 0,52 kat azaldığı görülmektedir. Buradan Bayındır Deresi Bayındır Barajı Çıkışı istasyonundaki suyun daha az tuzlu olduğu söylenebilir. Yukarıda  $Na^{+}$  ile ilgili yorum  $K^{+}$  değeriyle ilgili yorumu destekler niteliktedir.

Suların sertliği, su içerisinde çözülmüş olarak bulunan toplam  $Ca^{++}$  ve  $Mg^{++}$  miktarının kalsiyum karbonat ( $CaCO_3$ ) eş değeri olarak tanımlanmıştır. Bunun sebebi, sudaki mevcut iyonlara göre daha fazla oranda  $Ca^{++}$  ve  $Mg^{++}$  iyonlarının bulunmasıdır (Şengün, 2013). Suların sertliğini belirleyen parametrelerden biri  $Mg^{++}$  dur. Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda  $Mg^{++}$  değerinin Lojistik-LASSO yöntemine göre 0,56 kat azaldığı Lojistik-Elastik ağ yöntemine göre de 0,55 kat azaldığı görülmektedir. Buradan her iki yöntemine göre de Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda  $Mg^{++}$  değerinin daha az olduğu ve bu istasyondaki suyun daha yumuşak olduğunun bir kanıtıdır. Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda  $Ca^{++}$  değerinin Lojistik-LASSO yönteminde 1,05 kat arttığı ve Lojistik-Elastik ağ yönteminde ise 1,13 kat arttığı görülmektedir. Burada kalsiyumun pozitif işaretli çıkması başka bir su kalite değişkeni olan toplam sertliği (TH) de pozitif işaretli çıkmasına sebep olmuştur. Bu durumda TH için analiz sonuçları incelendiğinde Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre



Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda TH değerinin her iki yöntemle göre de 1,06 kat arttığı gözlemlenmektedir.  $Ca^{++}$  ile ilgili yapılan yorum burdaki yorumu destekler niteliktedir.

Sudaki çözünmüş maddeler, suda çözünmüş az miktardaki organik madde ve anorganik tuzların varlığından ileri gelir. TDS suyun bulanıklılığını artırmaktadır. Toplam çözünmüş maddeler, doğal kaynaklardan, lağım atıklarından, şehir drenaj sularından ve endüstriyel sulardan ileri gelmektedir. Lojistik-Elastik ağ yöntemiyle yapılan analiz sonucunda Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda TDS değerinin 1,011 kat arttığı gözlemlenmiştir. Buradan Bayındır Deresi Bayındır Barajı Çıkışı istasyonundaki suyun daha bulanık olduğu söylenebilir.

$NH_4N$  genellikle çözünmüş oksijenden sonra ikinci önemli su kalite parametresidir. Sudaki  $NH_4N$ , organik maddelerin bozunması, özellikle organik gübre veya inorganik amonyum kaynaklı kimyasal gübreleme sonucunda, evsel ve endüstriyel kirlenmelerden oluşabilir (Egemen ve Sunlu,1999). Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda  $NH_4N$  değerinin Lojistik-LASSO yöntemiyle yapılan analiz sonuçlarına göre 1,49 kat arttığı ve Lojistik-Elastik ağ yöntemiyle yapılan analiz sonuçlarına göre de 2,91 kat arttığı görülmektedir. Bu durum Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda evsel veya endüstriyel bir kirlenmenin daha fazla olduğunu göstermektedir.

Su içindeki hidrojen iyonu konsantrasyonunu 10 tabanına göre negatif logaritması Ph değeri olarak tanımlanır. Ph=7 olan sular nötr olarak bilinir. Ph değeri 7'nin altındaysa su asidik karakter kazanırken Ph 7'nin üzerinde ise su bazik karakter kazanır. Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda Ph değerinin Lojistik-LASSO yöntemiyle yapılan analiz sonucunda 2,64 kat arttığı ve Lojistik-Elastik ağ yöntemine göre de 5,61 kat arttığı görülmektedir.

Değişik sanayilerden atılan atıklarda ve minerallerin kavrulması işleminden sonra sülfat ( $SO_4$ ) sulara verilir. Suda sülfat genellikle yüksek konsantrasyonlarda bulunabilir.

Çünkü, kayalardan çözeltilmeye geçen katyonlar genellikle sülfatla, çözünebilen bileşikler verirler. Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda  $SO_4$  değerinin Lojistik-LASSO yöntemine göre 0,92 azaldığı ve Lojistik-Elastik ağ yöntemine göre de 0,85 kat azaldığı görülmektedir. Buradan Bayındır Deresi Bayındır Barajı Çıkışı istasyonu çevresinde bulunabilecek tarlalardan gelen kayalardan suya  $SO_4$  karışımının daha az olduğu söylenebilir.

Yüksek klorür konsantrasyonu bir kirlenme belirtisi olmakla birlikte önemli bir kalite parametresidir (Tuncay 1983). Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda  $Cl^-$  değerinin Lojistik-LASSO yöntemiyle yapılan analiz sonucunda 1,02 kat arttığı ve Lojistik-Elastik ağ yöntemine göre de 1,04 kat arttığı görülmektedir. Bu durumda Bayındır Deresi Bayındır Barajı Çıkışı istasyonundaki suyun her iki yöntemine göre de daha kirli olduğu söylenebilir.

Orto fosfat ( $O-PO_4$ ) değerleri incelendiğinde Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda Lojistik-LASSO yöntemine göre 0,03 kat azaldığı ve Lojistik-Elastik ağ yöntemine göre de 0,01 kat azaldığı görülmektedir. Bu durumda her iki yöntemine göre de Bayındır Deresi Bayındır Barajı Çıkışı istasyonu çevresinde bulunan tarlalarda fosforlu gübre kullanıldığını ve gübrenin yağışla suya karıştığı şeklinde bir yorum yapılabilir.

Analiz sonuçlarına göre M-Al değeri Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda Lojistik-LASSO yöntemine göre 1,01 kat arttığı ve Lojistik-Elastik ağ yöntemine göre de 1,02 kat arttığı görülmektedir. P<sub>v</sub> değeri ise Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda Lojistik-LASSO yöntemine göre 3,36 kat arttığı ve Lojistik-Elastik ağ yöntemine göre de 2,35 kat arttığı görülmektedir. Son olarak sıcaklığın (T) Çubuk Çayı Çubuk 1 Baraj girişi istasyonuna göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda Lojistik-Elastik ağ yöntemine göre 1,01 kat arttığı görülmektedir.

Üçüncü grup istasyonlardan seçilen Kuruçay Baraj Aksı ile Bayındır Deresi Bayındır Barajı Çıkışı istasyonlarına ait analiz sonuçları Çizelge 5.7'de verilmiştir.

Çizelge 5.7. Kuruçay Baraj Aksı ile Bayındır Barajı Çıkışı İst. ait Analiz Sonuçları

Parametre	Lojistik Regresyon		Robust Lojistik Regresyon		Lojistik Ridge Regresyon $k = 0,57$		Lojistik LASSO Regresyon $\lambda = 0,079$		Elastik – Lojistik Regresyon $\lambda = 0,051$ $\alpha = 0,45$	
	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds
<b>Sabit</b>	5,430	-	4,955	-	4,012	-	6,708	-	7,381	-
<b>BOD5</b>	0,004	1	0,014	1,014	1,095	2,99	-	-	-	0,98
<b>Ca<sup>++</sup></b>	-0,248	0,78	-0,226	0,798	-1,034	0,36	0,015*	1,02	0,044*	1,04
<b>Cl<sup>-</sup></b>	0,673*	1,96	0,614*	1,848	6,705*	816,48	0,203*	1,23	0,169*	1,18
<b>DO</b>	-0,823	0,44	-0,317*	0,728	-3,539	0,03	-0,425*	0,65	-	0,64
<b>EC</b>	-0,071	0,93	-0,054*	0,947	-0,674	0,51	-0,009	0,99	-0,010	0,99
<b>K<sup>+</sup></b>	-0,267*	0,77	1,043*	2,838	2,708*	15	-0,014*	0,99	-	0,96
<b>M-Al</b>	-0,048	0,95	0,507*	1,660	0,631*	1,88	-0,006	0,99	-	0,99
<b>Mg<sup>++</sup></b>	-1,567	0,21	-0,143	0,867	-0,446	0,640	-0,227	0,80	-0,206	0,81
<b>Na<sup>+</sup></b>	0,351*	1,42	0,320*	1,377	0,385*	1,470	0,039*	1,04	0,081*	1,08
<b>NH<sub>4</sub>N</b>	0,566*	1,761	2,205	9,070	0,412*	1,510	0,626*	1,87	1,134*	3,11
<b>O-PO<sub>4</sub></b>	0,978	2,659	2,887	17,939	3,192*	34,337	2,740*	15,49	3,689*	40
<b>Ph</b>	2,355	10,54	2,149*	8,576	-2,530	0,08	-	-	-	-
<b>Pv</b>	0,501*	1,650	1,579*	4,850	1,877*	6,53	0,821*	2,27	0,748*	2,11
<b>SO<sub>4</sub></b>	-0,289	0,75	-0,023*	0,977	0,940*	2,56	-0,109	0,90	-0,139	0,87
<b>T</b>	-0,638*	0,53	-0,582	0,559	-0,493	0,61	-0,136	0,87	-	0,87
<b>TDS</b>	-0,066	0,94	-0,060*	0,942	1,673*	5,33	-0,004*	1	-	0,99
<b>TH</b>	0,362*	1,44	0,033*	1,034	0,327*	1,39	0,031	1,03	0,035*	1,04
<b>DSO</b>	<b>54,7</b>		<b>86,2</b>		<b>74,8</b>		<b>72,8</b>		<b>88,9</b>	

(•0.05 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.)

Çizelge 5.7 incelendiğinde ilk olarak lojistik yöntemiyle elde edilen model % 54,7 doğru sınıflandırma oranına sahiptir. İkinci sırada lojistik ridge regresyon yöntemiyle elde edilen modelin % 74,8 doğru sınıflandırma oranına sahip olduğu görülmektedir. Son olarak lojistik elastik ağ yöntemiyle elde edilen modelin % 88,9 doğru sınıflandırma oranına sahiptir. Lojistik regresyon yöntemiyle yapılmış analiz sonuçlarına göre Kuruçay Baraj Aksı olma olasılığının Bayındır Deresi Bayındır Barajı Çıkışı istasyonu olma olasılığına göre, BOD<sub>5</sub> değerinin 1 kat, Cl<sup>-</sup> değerinin 1,96 kat, Na<sup>+</sup> değerinin 1,42 kat, NH<sub>4</sub>N değerinin 1,761 kat, O-PO<sub>4</sub> değerinin 2,659 kat, Ph değerinin 10,54 kat, Pv değerinin 151,11 kat, ve TH değerinin 1,44 kat artarken, Ca<sup>++</sup> değerinin 0,78 kat, DO değerinin 0,44 kat, EC değerinin

0,93 kat, K<sup>+</sup> değerinin 0,77 kat, M-Al değerinin 0,95 kat, Mg<sup>++</sup> değerinin 0,21 kat, SO<sub>4</sub> değerinin 0,75 kat, T değerinin 0,53 kat, TDS değerinin 0,94 kat azaldığı gözlemlenmiştir.

Sağlam lojistik regresyon yöntemiyle yapılmış analiz sonuçlarına göre, Kuruçay Baraj Aksı olması olasılığı, Bayındır Deresi Bayındır Barajı Çıkışı istasyonu olması olasılığına göre, BOD<sub>5</sub> değerini 1,014 kat, Cl<sup>-</sup> değerini 1,848 kat, Na<sup>+</sup> değerini 1,377 kat, K<sup>+</sup> değerini 2,838 kat, M-Al değerini 1,660 kat, NH<sub>4</sub>N değerini 9,070 kat, O-PO<sub>4</sub> değerini 17,939 kat, Ph değerini 8,576 kat, Pv değerini 4,850 kat ve TH değerini 1,034 kat, arttırırken, Ca<sup>++</sup> değerini 0,80 kat, DO değerini 0,728 kat, EC değerini 0,947 kat, Mg<sup>++</sup> değerini 0,867 kat, SO<sub>4</sub> değerini 0,977 kat, T değerini 0,559 kat, TDS değerini 0,942 kat azalttığı gözlemlenmiştir.

Lojistik Ridge regresyon yöntemiyle yapılmış analiz sonuçlarına göre Kuruçay Baraj Aksı olması olasılığı, Bayındır Deresi Bayındır Barajı Çıkışı istasyonu olma olasılığına göre, BOD<sub>5</sub> değerini 2,99 kat, Cl<sup>-</sup> değerini 816,48 kat, K<sup>+</sup> değerini 15 kat, M-Al değerini 1,88 kat, Na<sup>+</sup> değerini 1,470 kat, NH<sub>4</sub>N değerini 1,510 kat, O-PO<sub>4</sub> değerini 34,337 kat, Pv değerini 6,53 kat, SO<sub>4</sub> değerini 2,56 kat, TDS değerini 5,33 kat ve TH değerini 1,39 kat arttırırken, Ca<sup>++</sup> değerini 0,36 kat, DO değerini 0,03 kat, EC değerini 0,51 kat, Mg<sup>++</sup> değerini 0,640 kat, Ph değerini 0,08 kat ve T değerini 0,61 kat azalttığı görülmektedir.

BOD<sub>5</sub> sulardaki organik maddelerin mikroorganizmalar tarafından parçalanması için gerekli olan oksijen miktarını ifade etmektedir. Harcanan oksijen miktarı fazlaysa ölçüm yapılan istasyondaki organik yükün de fazla olduğunu göstermektedir. Bu durumda ölçüm yapılan istasyondaki organik kirliliğin yüksek olduğunu belirtir. Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda Lojistik-Elastik ağ yöntemiyle yapılan analize göre BOD<sub>5</sub> değerinin 0,98 kat azaldığı görülmektedir. Buradan Kuruçay Baraj Aksı'nda organik kirliliğin daha fazla olduğunu göstermektedir.

Suların sertliği, su içerisinde çözülmüş olarak bulunan toplam Ca<sup>++</sup> ve Mg<sup>++</sup> miktarının kalsiyum karbonat (CaCO<sub>3</sub>) eş değeri olarak tanımlanmıştır. Bunun sebebi, sudaki mevcut iyonlara göre daha fazla oranda Ca<sup>++</sup> ve Mg<sup>++</sup> iyonlarının bulunmasıdır

(Şengün, 2013). Suların sertliğini belirleyen parametrelerden biri  $Mg^{++}$  dur. Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda  $Mg^{++}$  değerinin Lojistik-LASSO yöntemine göre 0,80 kat azaldığı ve Lojistik-Elastik Ağ Yöntemine göre de 0,81 kat azaldığı görülmektedir. Her iki yöntemine göre de Kuruçay Baraj Aksı istasyonundaki suyun daha sert olduğunu göstermektedir.  $Ca^{++}$  değerinin ise Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda Lojistik-LASSO yöntemine göre 1,02 kat arttığı ve Lojistik-Elastik Ağ Yöntemine göre de 1,04 kat arttığı görülmektedir. Başka bir su kalite karakteristiği olan toplam sertlik (TH) değerinin Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda Lojistik-LASSO yöntemine göre 1,03 kat arttığı ve Lojistik-Elastik ağ yöntemine göre de 1,04 kat arttığı görülmektedir.

DO, sulardaki oksijen miktarını belirlemek için kullanılan bir parametredir. Canlı yaşamı için kritik öneme sahip olan çözünmüş oksijenin azlığı, yüzeysel sularda kirliliğin en önemli göstergesidir. Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda DO değerinin Lojistik-LASSO yöntemine göre 0,65 kat azaldığı ve Lojistik-Elastik ağ yöntemine göre de 0,64 kat azaldığı görülmektedir. Buradan her iki yöntemine göre de Bayındır Deresi Bayındır Barajı Çıkışı istasyonunun suda organik kirliliğin daha az olduğunu göstermektedir.

Yüksek klorür konsantrasyonu bir kirlenme belirtisi olmakla birlikte önemli bir kalite parametresidir (Tuncay 1983). Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda  $Cl^-$  değerinin Lojistik-LASSO yöntemine göre 1,23 kat arttığı ve Lojistik-Elastik Ağ yöntemine göre de 1,18 kat arttığı görülmektedir. Bayındır Deresi Bayındır Barajı Çıkışı İstasyonunun daha kirli olduğunu göstermektedir. Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda EC değerinin her iki yöntemine göre de 0,99 kat azaldığı görülmektedir. Bu da Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda iyon miktarının daha az olduğunu göstermektedir. TDS suyun bulanıklılığını artırmaktadır. Toplam çözünmüş maddeler, doğal kaynaklardan, lağım atıklarından, şehir drenaj sularından ve endüstriyel sulardan ileri gelmektedir. Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda TDS değerinin

Lojistik-LASSO yöntemine göre 1 kat arttığı ve Lojistik-Elastik ağ yöntemine göre 0,99 kat azaldığı görülmektedir.

Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda Na<sup>+</sup> değerinin Lojistik-LASSO yöntemine göre 1,04 kat arttığı ve Lojistik-Elastik ağ yöntemine göre de 1,08 kat arttığı görülmektedir. Her iki yöntemine göre de Bayındır Deresi Bayındır Barajı Çıkışı istasyonundaki suyun daha tuzlu olduğu söylenebilir. Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda K<sup>+</sup> değerinin Lojistik-LASSO yöntemine göre 0,99 kat azaldığı ve Lojistik-Elastik ağ yöntemine göre de 0,96 kat azaldığı görülmektedir. Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda SO<sub>4</sub> değerinin Lojistik-LASSO yöntemine göre 0,90 kat azaldığı ve Lojistik-Elastik Ağ yöntemine göre de 0,87 kat azaldığı görülmektedir. Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda NH<sub>4</sub>N değerinin Lojistik-LASSO yöntemine göre 1,87 kat arttığı ve Lojistik-Elastik ağ yöntemine göre de 3,11 kat arttığı görülmektedir.

O-PO<sub>4</sub> değeri için Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda O-PO<sub>4</sub> değerinin Lojistik-LASSO yöntemine göre 15,49 kat arttığı ve Lojistik-Elastik Ağ yöntemine göre 40 kat arttığı görülmektedir. Buradan Bayındır Deresi Bayındır Barajı Çıkışı istasyonu çevresinde bulunan tarlalarda fosforlu gübre kullanıldığını ve bu gübrenin yağışla birlikte bu istasyona karıştığı söylenebilir.

Kuruçay Baraj Aksına göre Bayındır Deresi Bayındır Barajı Çıkışı istasyonunda her iki yöntemine göre de sıcaklığı 0,87 kat azalmaktadır. Buradan Bayındır Deresi Bayındır Barajı Çıkışı istasyonundaki sıcaklığın daha az olduğu söylenebilir.

## 5. SONUÇ VE ÖNERİLER

Bu tez çalışmasında lojistik regresyon modelinde karşılabilecek iki problem ele alınmıştır. İlki çoklu iç ilişki problemidir. İkinci olarak ele alınan problem ise aykırı değer problemidir. Aykırı değer problemin üstesinden gelmek için sağlam parametre tahmini yapan Bianco ve Yohai tahmin edicisine dayalı sağlam lojistik regresyon tanıtılmıştır.

Çoklu iç ilişkinin çözümüne yönelik yanlı ve boyut indirgeyen yöntemler tanıtılmıştır. Lojistik ridge regresyon ele alınan tüm değişkenlerin katsayılarını tahmin etmeye çalışan yanlı bir regresyon yöntemidir. LASSO ve Elastik Ağ aynı anda hem boyut indirgeyen hem de model parametrelerini tahmin eden yöntemlerdir. Son dönemlerde bu yöntemler literatürde yaygın olarak kullanılmaktadır.

Çoklu doğrusal regresyon modelinde olduğu gibi lojistik regresyon modelinde de çoklu iç ilişki problemiyle başa çıkmanın bir başka yolu da değişken seçimidir. Bu amaçla hem boyut indirgeyen hem de model parametrelerini tahmin eden lojistik LASSO yöntemi incelenmiştir. Bu yöntemin sıfıra büzülen tüm değişkenleri modelden atması bunun bir dezavantajıdır.

Lojistik-LASSO'nun dezavantajlarını gidermek için, yeni bir kısıt eklenerek Elastik ağ yöntemi önerilmiştir.  $L_1$  ve  $L_2$  normunu birleştiren Elastik Ağ yöntemi Lojistik regresyon modeline uyarlanmıştır.

Model seçimi için, çapraz geçerlilik testi kullanılmıştır. Seçilen modelin istatistiksel olarak anlamlılığının testi için kovaryans testi hakkında bilgiler verilmiştir. Modelin anlamlılık sınavasından sonra katsayıların anlamlılık sınavası için kullanılacak yaklaşımlar açıklanmıştır.

Çalışmanın uygulama bölümünde, su kalitesinin tespitinde kullanılan değişkenler yardımıyla yukarıda ifade edilen yöntemlerin performansı belirlenmeye çalışılmıştır. Öncelikle ele alınan su kalite karakteristikleriyle ilgili genel bilgiler verilmiştir. Yapılan uygulamada DSİ'den temin edilen Sakarya Nehri'nden seçilmiş 8 istasyona ait su kalite

verilerine lojistik regresyon, lojistik Ridge, lojistik LASSO, lojistik Elastik Ağ ve robust lojistik yöntemleri kullanılarak analiz yapılmıştır. Analizler için R programlama dili kullanılmıştır. Seçilen istasyonlar ikiyeşerli gruplandırılarak iki istasyon arasındaki su kalitesi ölçümlerinin değişimine sebep olan değişkenler belirlenmeye çalışılmıştır. Tüm bu analizler sonucunda hangi yöntemin daha iyi olduğuna karar verebilmek için doğru sınıflandırma oranı kullanılmıştır.

Bu bağlamda, Lojistik Elastik ağ yönteminin daha az parametreyle daha doğru sınıflandırma verdiğinden su kalitesi verileri kullanılarak yapılan analizlerde bu yöntemin kullanılması tavsiye edilir.

Çizelge 6.1. İstatistiksel olarak Anlamlı olan Değişkenler

Yöntem	Lojistik Elastik Ağ
Akmeşe ve İnegöl Yenicedere İstasyonu	BOD <sub>5</sub> , Ca <sup>++</sup> , EC, K <sup>+</sup> , Mg <sup>++</sup> , Na <sup>+</sup> , NH <sub>4</sub> N, O-PO <sub>4</sub> , pH, pV, T, TDS, TH
Sarıyar Baraj Çıkışı ve Dere mah. Köprü İstasyonu	BOD <sub>5</sub> , Cl <sup>-</sup> , DO, K <sup>+</sup> , Mg <sup>++</sup> , Na <sup>+</sup> , NH <sub>4</sub> N, pH, pV, TDS, TH
Kayaboğazı Baraj Çıkışı ve Dere mah. Köprü İstasyonu	BOD <sub>5</sub> , Ca <sup>++</sup> , Cl <sup>-</sup> , DO, EC, M-Al, Na <sup>+</sup> , O-PO <sub>4</sub> , pH, pV, T, TDS, TH
Sarıyar Baraj Çıkışı ve Kayaboğazı Baraj Çıkışı	BOD <sub>5</sub> , Ca <sup>++</sup> , Cl <sup>-</sup> , K <sup>+</sup> , M-Al, Mg <sup>++</sup> , Na <sup>+</sup> , O-PO <sub>4</sub> , NH <sub>4</sub> N, pH, pV, TDS, TH
Çubuk 1 Baraj Çıkışı ve Kuruçay Baraj Aksı	BOD <sub>5</sub> , Ca <sup>+</sup> , DO, K <sup>+</sup> , M-Al, Mg <sup>++</sup> , Na <sup>+</sup> , NH <sub>4</sub> N, O-PO <sub>4</sub> , pH, pV, SO <sub>4</sub> , T, TDS, TH
Çubuk 1 Baraj Çıkışı ve Bayındır Barajı Çıkışı	BOD <sub>5</sub> , Ca <sup>++</sup> , Cl <sup>-</sup> , DO, EC, K <sup>+</sup> , M-Al, Na <sup>+</sup> , NH <sub>4</sub> N, O-PO <sub>4</sub> , pV, T, TDS, TH
Kuruçay Baraj Aksı ve Bayındır Barajı Çıkışı	Ca <sup>++</sup> , Cl <sup>-</sup> , DO, K <sup>+</sup> , M-Al, Na <sup>+</sup> , NH <sub>4</sub> N, pH, pV, SO <sub>4</sub> , T, TH

Lojistik Elastik Ağ analiz sonuçları incelendiğinde; her istasyon için anlamlı çıkan değişkenler, Çizelge 6.1 de verilmiştir. Buradan hareketle, suyun kalitesi ölçmek için bu değişkenler kullanılabilir.



## KAYNAKLAR DİZİNİ

- Aguilera, A. M., Escabias, M., Valderrama, M. J., 2006, Using Principal Components For Estimating Logistic Regression With High- Dimensional Multicollinear Data. *Computational Statistics & Data Analysis*, 50: 1905-1924.
- Akdi, Y., 2014, *Matematiksel İstatistiğe Giriş*, Gazi Kitabevi, s.504-524.
- Akın, M. & Akın, G., 2007, Suyun Önemi, Türkiye’de Su Potansiyeli, Su Havzaları ve Su Kirliliği. *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi*, 47(2), s.105-118.
- Akın, H.B., Şentürk, E., 2012, Bireylerin Mutluluk Düzeylerinin Ordinal Lojistik Regresyon Analizi İle İncelenmesi, 10, 37, s.83-93.
- Arıcan, E., 2010, ‘Nitel Yanıt Değişkene Sahip Regresyon Modellerinde Tahmin Yöntemleri’, Yüksek Lisans Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, s.105
- Akyol, K., Bayır, Ş., Şen, B., 2020, Importance of Attribute Selection for Parkinson Disease, *Academic Platform Journal of Engineering and Science* 8-1, s.175-180.
- Algamal, Z. Y., Lee, M. H., 2015, Penalized Logistic Regression with the adaptive LASSO for Selection in high dimensional cancer classification, *Expert Systems With Applications*.
- Algamal, Z. Y., Lee, M. H., 2015, High Dimensional Logistic Regression Model using Adjusted Elastic Net Penalty, *Expert Systems With Applications*.
- Alkhamisi, M., Shukur, G., 2008, Developing ridge parameters for SUR model, *Communications in Statistics Theory and Methods* 37(4):544–564.
- Anderson, J.A., Richardson, S.C., 1979, Logistic discrimination and bias correction in maximum likelihood estimation. *Technometrics*, 21, p.71-78.
- Anderson, J.A., 1983, Robust Inference Using Logistic Models, *Bulletion of International Statistical Institute*, 48, 35-53.
- Aranda-Ordaz, F. J., 1981, On Two Families of Transformations to Additivity for Binary Response Data, *Biometrika*, 68, p.357–363.
- Arı, A., Önder, H., 2013, Farklı Veri Yapılarında Kullanılabilecek Regresyon Yöntemleri, *Anadolu Tarım Bilimleri Dergisi*, 28, 3, s.168-174.
- Arıcan, E., 2010, Nitel yanıt değişkenine sahip regresyon modellerinde tahmin yöntemleri, Yüksek Lisans tezi, ÇU, s.105.
- Arıcan, E., 2016, Lojistik Regresyon Modelinde Çoklu iç İlişki Problemi, Doktora tezi, ÇU, s.99..

### KAYNAKLAR DİZİNİ (devam)

- Baba, A., M., Maibulangu, S., Bishir, A., 2014, Adopting Elastic Net Penalization in Logistics Regression to Achieve Stability: A Case Study of University of UMTHBorno State Nigeria
- Barker, L., Brown, C., 2001, Logistic Regression When Binary Predictor Variables Are Highly Correlated, *Statistics In Medicine*, 20: 1431-1442
- Bayram, N., 2004, Multinomial Lojistik Regresyon Analizinin İstihdamdaki İşgücüne Uygulanması, *İstanbul Üniversitesi İktisat Fakültesi Mecmuası*, 54, 2, 61-76.
- Berkson, J., 1944, Application of the Logistic Function to Bio-Assay. *Journal of the American Statistical Association*, 9, 357-365.
- Bianco, A.M., Yohai V.J., 1996, *Robust Statistics, Data Analysis, and Computer Intensive Methods*, Springer New York, 109, 17-34.
- Bielza, C., Robles, V., Larrañaga, P., 2011, Regularized Logistic Regression Without A Penalty Term: An Application To Cancer Classification With Microarray Data. *Expert Systems With Applications*, 38(5): 5110-5118.
- Bilgin, M., 2012, Türetilmiş İkili Heterojen Veri Yapılarında Genel, Sağlam ve Kesin Lojistik Regresyon Yöntemlerinin Karşılaştırılması, Yüksek Lisans, ESOGÜ, 59 s.
- Bircan, H., 2004. Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama, *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 2, 185-208.
- Bonney, G. E., 1987, Logistic Regression For Dependent Binary Observations, *Biometrics*, 43, 951-973.
- Buescher, P. A., Larson, L. C., Nelson, M. D., Lenihan, A. J., 1993, Prenatal WIC Participation Can Reduce Low Birth Weight and Newborn Medical Costs: A Cost Benefit Analysis of Wic Participation in North Carolina. *Journal of the American Dietetic Association*, 93, 163-166
- Bulut and Kubilay, 2019, "Eğirdir Gölü (Isparta/Türkiye) su kalitesinin mevsimsel değişimi", *Ege Journal of Fisheries and Aquatic Sciences*, 36(1), 13-23.
- Brooks, W., Corsi, S., Fienen, M., Carvin, R., 2016, Predicting Recreational water Quality Advisories: A comparison of Statistical Methods, *Environmental Modelling & Software* 76 (2016) p.81-94
- Bruce, P., Bruce, A., 2017, *Practical Statistics for Data Scientists*, O'Reilly Media.
- Candia, J., Tsang, J. S., 2019, eNetXplorer: an R package for the quantitative exploration of elastic net families for generalized linear models, *BMC Bioinformatics*, 20:189.
- Chang, L., Roberts, S., Welsh, A., 2018, Robust Lasso Regression Using Tukey's Biweight Criterion, *Technometrics*, VOL. 60, NO. 1, 36-47

### KAYNAKLAR DİZİNİ (devamı)

- Chapman, D., Kimstach, V. 1996. Chapter 3. Selection of Water Quality Variables. Water Quality and Assessments-A Guide to Use of Biota, Sediments and Water in Enviromental Monitoring, Second Edition, Chapman, D. (ed), pp 1-56, UNESCO / WHO/ UNEP.
- Cirik, S. and Ş. Gökpinar, 1993. Plankton and Plankton Culture, (in Turkish). Ege Üniv. Basımevi, Bornova, İzmir, 274s
- Comber, A., Harris, P., 2018, Geographically weighted elastic net logistic regression, Journal of Geographical Systems, 317-341 (2018)
- Coskun, S., Kartal, M., Coskun, A., Bircan, H., 2004, Lojistik Regresyon Analizinin İncelenmesi ve Dış Hekimliğinde Bir Uygulama, Cumhuriyet Üniversitesi Dış Hekimliği Fakültesi Dergisi, 7, 1, 41-50
- Cox, D. R., 1970, The Analysis of Binary Data, Methuen, Journal of the Royal Statistical Society. Series C (Applied Statistics) ,21, 2, 113-120.
- Croux, C. and Haesbroeck, G., 2003, Implementing the Bianco and Yohai estimator for Logistic Regression, Computational Statistics and Data Analysis, 44(1-2):273-295
- Çoban, F., 2007, Hazar göülü su kalitesinin araştırılması, Fırat Üniversitesi, Yüksek Lisans Tezi, s.96.
- Demir, R. E., 2017, Cezalı Regresyon Yöntemleri ve Uygulanması, Yüzüncüyıl Üniversitesi, Yüksek Lisans Tezi, s.53
- Dempster, A. P., Schatzoff, M., & Wermuth, N. 1977, A simulation study of alternatives to ordinary least squares. Journal of the American Statistical Association, 72, 77-91.
- Duffy, D. E., Santner, T. J., 1989, On the Small Sample Properties of Norm-Restricted Maximum Likelihood Estimators for Logistic Regression Models, Communs Statist, Theory Meth., 18, 959-980.
- Duffy, D.E., 1990, On Continuity-corrected Residuals in Logistic Regression, Biometrika, 77: 287-293.
- Dorugade, A.V., 2013, New ridge parameters for ridge regression, Journal of the Association of Arab Universities for Basic and Applied Sciences, 2014, (15), 94-99
- Devlet Su İşleri Müdürlüğü Enerji ve Tabii Kaynakları Bakanlığı., 2001, Eğrekkaya Baraj Gölü ve Havzasında Kirlilik Araştırması Raporu., Ankara.
- Egemen Ö, Sunlu U, 1999, Su Kalitesi (Ders Kitabı) III. Baskı, Ege Üniversitesi Basımevi, İzmir
- Emiroğlu, E., 2014, Optimizasyon Uygulaması Olarak Regresyon Parametrelerinin Seçimi, İstanbul Üniversitesi, Doktora Tezi, s.96

## KAYNAKLAR DİZİNİ (devamı)

- Erar, A., 2013, Doğrusal Regresyon Analizine Giriş, Nobel Yayınları, s.645
- Ergül, B., Yavuz, A., A., Gündoğan, E., 2016, Avrupa Bayanlar Voleybol Şampiyonlar Ligi Takımlarının Lojistik, Robust Lojistik ve Lojistik-Ridge Regresyon Analizi ile Sınıflandırılması, X<sup>th</sup> International Statistics Days Conference, Giresun
- Ergül, B., Yavuz, A., A., 2016, Eskişehir İli Otomobil Tercihlerinden Memnuniyet: Bir İstatistiksel Bakış, X<sup>th</sup> International Statistics Days Conference, Giresun
- Ertan, G., 2019, Niğde Bölgesi İçme ve Sulama Suyu Kaynaklarının Hidrojeokimyasal, Su Kalitesi ve İçilebilirlik Açısından Değerlendirilmesi, Niğde Ömer Halisdemir Üniversitesi, Yüksek Lisans Tezi, s.133.
- Ertuğrul, S., 2018, Farklı Tuzluluğa Sahip Göllerde Popülasyon Büyüklüğünün Alansal ve Zamansal Modellenmesi, Hacettepe Üniversitesi, Yüksek Lisans Tezi, s.145.
- Fidanoğlu, I., 2009, İstatistiksel Daraltıcı (shrinkage) Model ve Uygulamaları, Çukurova Üniversitesi, Yüksek Lisans Tezi ,
- Garcia-Carretero, R., Vigil-Medina, L., Barquero-Perez, O., Mora-Jimenez, I., 2020, Logistic LASSO and Elastic Net to Characterize Vitamin D Deficiency in a Hypertensive Obese Population, Metabolic Syndrome and Related Disorders, Vol. 18, No.2
- Garside, P. S., Glueck, C. J., 1995, The Important Role of Modifiable Dietary and Behaviour Characteristic in the Causation and Prevention of Coronary Heart Disease Hospitalization and Mortality. Journal of American College of Nutrition, 14, p.71- 79.
- Genç, M., 2020, Ridge, LASSO And Elastic Net Methods In High Dimensional Data, Çukurova Üniversitesi, Doktora Tezi, s.160.
- Gibbons, D. G., 1981, A simulation study of some ridge estimators, Journal of the American Statistical Association, 76:131–139.
- Gölbaşı, S., Şen B., 2019, Atatürk Baraj Gölü'ne dökülen Kahta Çayı'nın (Adıyaman) su kalitesi, Ege Journal of Fisheries and Aquatic Sciences, 36(4): 337-346.
- Göv, A., 2019, Cezalı Tahmincilere Dayalı Granger Nedensellik Analizi Ve Uygulamaları, İnönü Üniversitesi, Doktora Tezi, s.111.
- Güler, Ç., 1997, Su Kalitesi, Çevre Sağlığı Temel Kaynak Dizisi, 43, 1-93 s.
- Gültay, B., 2009, Çoklu İç İlişki Ve Ekolojik Regresyon, Çukurova Üniversitesi, Yüksek Lisans Tezi, s.35-40.
- Gündoğan, E., 2015, Çoklu İç İlişki ve Aykırı Değer Olduğu Durumda Lojistik Regresyon Yaklaşımları, Eskişehir Osmangazi Üniversitesi, Yüksek Lisans Tezi, 51 s.

### KAYNAKLAR DİZİNİ (devamı)

- Gündoğan Aşık, E., A. Altın Yavuz, 2018, "Investigation Of Life Satisfaction In Oecd Countries With Multivariate Analysis Methods," *Journal Of Social And Humanities Sciences Research (Jshsr)*, Vol. 5, No. 26, Pp. 0–0.
- Hacıoğlu N., Dulger B., 2009, Monthly variation of some physico-chemical and microbiological parameters in Biga Stream (Biga, Canakkale, Turkey), *African Journal of Biotechnology*, 8(9), 1929-1937.
- Hadi A.S., Imon R.A.H.M., 2008, Identification of Multiple Outliers in Logistic Regression, *Communications in Statistics, Theory and Methods*, 37, 1697–1709.
- Hastie, T., Tibshirani, R., Wainwright, M., 2015, *Statistical Learning with Sparsity* Taylor & Francis Group, LLC, p.335
- Hobza, T., Pardo, L., Vajda, I., 2012, Robust median estimator for generalized linear models with binary responses, *Kybernetika*, 48, 4, 768-794.
- Hoerl, A. E., Kennard, R. W., 1970, Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12:55–67.
- Johnson, W., 1985, Influence Measures for Logistic Regression. Another Point of View, *Biometrika*, 72, 1, 59–65.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013, *An introduction to statistical learning: with Applications in R.*, New York: Springer, Vol. 112, p. 18.
- Kalkan, E., 2020, Linner Regresyonda Küçültme Yöntemlerinin Karşılaştırılması, *Dicle Üniversitesi, Yüksek Lisans Tezi*, s.78.
- Karacasu, M., Yavuz, A., A., Ergül, B., Akalın, K., B., 2016, Traffic Perception in Eskişehir Province, *World Multidisciplinary Earth Sciences Symposium (WMESS 2016)*, IOP Conf. Series: Earth and Environmental Science 44 (2016) 052063.
- Khalaf, G., Alkhamisi, M., 2006, Some modifications for choosing ridge parameters, *Communications in Statistics—Theory and Methods* 35:2005–2020.
- Kibria, B. M. G., 2003, Performance of some new ridge regression estimators. *Communications in Statistics—Theory and Methods* 32:419–435.
- Kibria, B. M. G., Mansson, K., Shukur, G., 2011, Performance of Some Logistic Regression, *Comput Econ* (2012) 40:401–414
- Kim, S., M., Kim, Y., Jeong, K., Jeong, H., Kim, J., 2018, Logistic LASSO regression for the diagnosis of the breast cancer using clinical demographic data and the BI-RADS lexicon for ultrasonography, *Ultrasonography* 2018;37:36-42, ISSN: 2288-5919

### KAYNAKLAR DİZİNİ (devamı)

- Kloiber, L. L., Winn, N. J., Shaffer, S. G., Hassanein, R. S., 1996, Late Hyponatremia in Very Low Birth Weight Infants: Incidence and Associated Risk Factors. *Journal of the American Dietetic Association*, 96, 880-884
- Küçük, A., 2020, Doğrusal Regresyonda Ridge, Liu ve LASSO tahmin edicileri üzerine bir çalışma, Hacettepe Üniversitesi Yüksek Lisans Tezi, s.57.
- Lang, K., 1995, Newsweeder: Learning to filter netnews., in *Proceedings of the Twelfth International Conference on Machine Learning (ICML)*, pp. 331-339.
- Lawless, J. F., Wang, P., 1976, A simulation study of ridge and other regression estimators. *Communications in Statistics-Theory and Methods*, 5(4), 307-323.
- Lee, C.T., 1984, Logistic Models for Cross-over Designs. *Biometrika*, 71: 216-217.
- Lecessie, S., Van Houwelingen J. C., 1992, Ridge Estimators in Logistic Regression. *Applied Statistics*, 41, 1, 191-201.
- Lesaffre, E., 1986, Logistic Discriminant Analysis With Applications In Electrocardiography, PhD thesis, Katholieke Universiteit Leuven, Belgium, 354 p. (unpublished).
- Lesaffre, E., Albert, A., 1989, A Multiple Group Logistic Regression Diagnostics, *Applied Statistics*, 38, 3, 425-440
- Liu, Y., Liang, Y., Liu, S., Rosenblum, D. S., Fellow, IEEE, Zheng, Y., 2016, Predicting Urban Water Quality with Ubiquitous Data, [arXiv:1610.09462v1](https://arxiv.org/abs/1610.09462v1)
- Lockhart, R., Taylor, J., Tibshirani, R., Tibshirani, R. J., 2014, A Significance Test for the LASSO, *The Annals of Statistics*, vol.42, No.2, p.413-468.
- Mansson, K., Shukur G., 2013, On Ridge Parameters in Logistic Regression, *Communications in Statistics - Theory and Methods*, 40, 18, 3366-3381.
- Melkumova, L.E, Shatskikh, S. Ya., 2017, ‘Comparing Ridge And Lasso Estimators For Data Analysis’, *Procedia Engineering*, 201, 746-755.
- Montgomery D.C., Peck A.E., and Vining G.G., 2001, *Introduction to linear regression analysis*, third edition, John Wiley & Sons.Inc., p. 641
- Muniz, G., Kibria, B. M. G., 2009, On Some Ridge Regression Estimators: An Empirical Comparisons, *Communications in Statistics-Simulation and Computation*, 38:3,621-630
- Myers, R., H., Montgomery, D., C., Goffrey Vining, G., Robinson, T., J., 2012, *Generalized Linear Models with Applications in Engineering and the Sciences*, Wiley Series Probability and Statistics, p.123-127.
- O’Neil, T. J., Barry, S. C., 1995, Truncated Logistic Regression, *Biometrics*, 51, 533-54

### KAYNAKLAR DİZİNİ (devamı)

- Özer, Ç ve Köklü, R (2019), "Aşağı Sakarya Nehri Su Kalitesinin Sulama Suyu Açısından Değerlendirilmesi" Artvin Çoruh Üniversitesi Doğal Afetler Uygulama ve Araştırma Merkezi, Doğal Afetler ve Çevre Dergisi, 2019; 5(2): 237-246
- Pereira, J. M., Basto, M., Silva, A. F., 2015, The logistic LASSO and Ridge Regression in Predicting Corporate Failure, *Procedia Economics and Finance* 39 (2016)634-641.
- Peoples, M. D., Siegel, E., Suchindran, C. M., Origasa, H., Ware, A., Barakat, A., 1991, Characteristics of Maternal Employment During Pregnancy: Effects on Low Birthweight, *American Journal of Public Health*, 81, 007-1012.
- Pregibon, D., 1981, Logistic Regression Diagnostics, *The Annals of Statistics*, 9, 4, 705-724.
- Qiu, D., 2015, An Applied Analysis of High Dimensional Logistic Regression, Yüksek Lisans Tezi, University of British Columbia.
- Rençber, Ö. F., Bağcı, H., 2019, Determination of Factors Affecting Capital Adequacy Using the Elastic Net Regression Method, *Uluslararası Toplum Araştırmaları Dergisi*
- Sabbe, N., Thas, O., Ottoy, J-P., 2012, EMLasso: logistic lasso with missing data, *Statistics in Medicine*, Volume32, Issue 18, p,3143-3157
- Sable, M. R., Herman, A. A., 1997, The Relationship Between Prenatal Health Behaviour Advice and Low Birthweight. *Public Health Reports*. 112, 332-339.
- Santos, I.S., Victoria, C.G., Huttly, S., Carvalhal, J.B. (1998), Caffeine Intake and Low Birth Weight: A Population Based Case Control Study. *American Journal of M.* (1988), *The Retreat From Class: A New True Socialism*, London: Verso.
- Schaefer, R. L., Roi, L. D., Wolfe, R. A., 1984, A Ridge Logistic Estimator, *Communications in Statistics Theory and Methods*, 13, p. 99-113.
- Sawyer C., McCarty P., Parkin G., 2013, Çevre Mühendisliği ve Bilimi İçin Kimya, Çeviri Editörü: İsmail TORÖZ.
- Şengün, E., 2013, Aksu Deresi Su kalitesi ve Kirlilik Düzeyinin Belirlenmesi, Giresun Üniversitesi, Yüksek Lisans Tezi, s.81.
- Sneidarman, R., 2020, From Linear Regression to Ridge Regression, The LASSO and the ELastic Net.
- Tepe, Y., 2009, Reyhanlı Yenişehir Gölü (Hatay) Su Kalitesinin Belirlenmesi, *Ekoloji* 18, 70, 38-46.
- Tibshirani, R., 1996, Regression Shrinkage and Selection via the LASSO, *Journal of Royal Statistical Society B*, 58:267–288
- Tuncay, H., Su Kalitesi, Ders Notları, Ege Üniversitesi Ziraat Fakültesi Toprak Bölümü, İzmir, 1983.
- Tütmez, B., 2020, Air Quality Assesment by Statistical Learning-Based Regularization, Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 35(2), ss. 271-278.

### KAYNAKLAR DİZİNİ (devamı)

- Uslu, O., Türkman, A., 1987, Su Kirliliği ve Kontrolü., T.C. Başbakanlık Çevre Genel Müd. Yayınları Ankara, No: 1, 344–345 s.
- Vorlickova, J., 2017, Least Absolute Shrinkage and Selection Operator Method, Charles University, Lisans Tezi, p.39
- Weissfeld, L. A., Sereika, S. M., 1991. A Multicollinearity Diagnostic For Generalized Linear Models. *Communications In Statistics Theory And Methods*, 20(4): 1183-1198.
- Wu, T., T., Chen, Y. F., Hastie, T., Sobel, E., Lange K., 2009, Genome-wide Association Analysis by Penalized Logistic Regression, *Bioinformatics*, Vol. 25 no. 6 2009, pages 714–721.
- Yaman, A., 2017, Çok değişkenli Bernoulli Lojistik Modellerde LASSO Tahmincileriyle Model Seçimi, On Dokuz Mayıs Üniversitesi, Yüksek Lisans Tezi, s.56.
- Yanık, T., Aras, N.M. & Çiltaş, A. K. (2001). Su Kalitesi Ders Kitabı. Atatürk Üniversitesi Ziraat Fakültesi Ders Yayınları No:225, Atatürk Üniversitesi Ziraat Fakültesi Ofset Tesisi, Erzurum, 142 s.
- Yavuz, A., Çilengiroğlu, Ö., V., 2020, Lojistik Regresyon ve CART Yöntemlerinin Tahmin Edici Performanslarının Yaşam Memnuniyeti Verileri için Karşılaştırılması, *Avrupa Bilim ve Teknoloji Dergisi Sayı 18, S. 719-727.*
- Yavuzkanat, M. S., 2011, ‘Lojistik Regresyonda Çoklu Aykırı Gözlemlerin Belirlenmesi Ve Etkililiklerinin İncelenmesi’, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, 91
- Yıldırım, N., Kayaalp, T., 2010, En Küçük Kareler, Ridge Regresyon ve Robust Regresyon Yöntemlerinde Analiz Sonuçlarına Aykırı Değerlerin Etkilerinin Belirlenmesi, s.24-29
- Zeny, Z.F., 2012, The LASSO and sparse least squares regression methods for SNP selection in Predicting quantitative traits, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9, 629-636.
- Zeydan, Ö., Özdoğan, N., Taştepe, Ş.P. ve Demirtaş, D., 2019, Kozlu Deresinde (Zonguldak) Su Kalitesinin İncelenmesi, *Artvin Çoruh Üniversitesi Doğal Afetler Uygulama ve Araştırma Merkezi, Doğal Afetler ve Çevre Dergisi*, 2019; 5(2): 187-197.
- Zou, H., Hastie, T., 2005, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, p.301-320.