

SU KALİTESİ DEĞİŞİMİNE ETKİ EDEN DEĞİŞKENLERİN LOJİSTİK REGRESYON, LOJİSTİK-RIDGE VE LOJİSTİK-LASSO YÖNTEMLERİ İLE TESPİTİ

Simge Sevim Uysal^{1,a,*}, Arzu Altın Yavuz^{2,b}, Cansu Filik İşçen^{3,c}

¹Eskisehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü İstatistik Bölümü, 26480, Eskisehir, Türkiye




²Eskisehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 26480, Eskisehir, Türkiye

³Eskisehir Osmangazi Üniversitesi, Eğitim Fakültesi, Matematik ve Fen Bilimleri Eğitimi Bölümü, 26480, Eskisehir, Türkiye

*Corresponding Author:

E-mail: simgeuysal93@gmail.com

(Received 17th April 2021; accepted 02th June 2021)

a:  ORCID 0000-0002-5301-1470, b:  ORCID 0000-0002-3277-740X, c:  ORCID 0000-0001-5463-8825

ÖZET. Günümüzde yer üstü su kaynaklarının kalitesinin belirlenmesi ve yönetimi tüm dünyada en önemli araştırma konuları arasında yer almaktadır. Su kalitesi ile ilgili yapılan çalışmalar, suyun kullanım amacının belirlenmesinde önemli olmasının yanı sıra, yüzey su kaynaklarının sürdürülebilir kullanımı açısından da önemlidir. Su kaynakları çeşitli faaliyetler sonucu kirlilik tehdidi ile karşı karşıya kalmaktadır. Sürdürülebilir bir su yönetimi, bu kaynaklarda meydana gelen fiziksel ve kimyasal değişimlerin sürekli izlenmesi ve değerlendirilmesini gerektirmektedir. Tüm lokasyonlardan alınan örneklerin sürekli olarak ölçümünün yapılması zaman ve maliyet gerektirmektedir. Ancak makine öğrenme yöntemleri yardımıyla İki istasyon arasında hangi su karakteristiklerinin değişime neden olduğunun belirlenmesi sağlanabilmektedir. Sonrasında bu değişkenler yardımıyla sürekli izleme sensörler yardımıyla gerçekleştirilebilmektedir. Bu çalışmada Sakarya nehri üzerindeki iki istasyon arasındaki su değişimine neden olan karakteristikler makine öğrenme yöntemleri olan lojistik regresyon, lojistik Ridge ve lojistik-LASSO yöntemleri ile belirlenmiştir. En etkili değişkenlerin belirlenmesinde lojistik LASSO yönteminin daha güvenilir sonuçlar verdiği belirlenmiştir. Sonuç olarak, Lojistik-LASSO yöntemine göre iki istasyon arasında değişime neden olan su kalite karakteristikleri Ca^{++} , $O-PO_4$, T, TH, BOD_5 , DO, EC, Mg^{++} , pH'tır.

Anahtar kelimeler: su kalitesi, sakarya nehri, lojistik regresyon, lojistik-ridge, lojistik-lasso

GİRİŞ

Su tüm canlılar için yaşam boyu biyolojik ve fiziksel faaliyetlerindeki en önemli ihtiyaçtır [1]. Canlıların yaşam ortamını oluşturduğu gibi kendisi de bir yaşam ortamıdır. Bu nedenle suyun kalitesi, yaşamında kalitesini etkilemektedir. Yeryüzünün 0.75 i sularla kaplı olmasına rağmen, içilebilir tatlı su kaynaklarının oranı % 0.74 düzeyindedir [1, 2, 3]. Dere, çay ve nehir gibi akarsular başlıca tatlı su kaynaklarıdır. Tatlı su kaynakları içme suyu başta olmak üzere, tarımsal alanların sulanmasında ve endüstriyel faaliyetlerin gerçekleştirilmesinde kullanılmaktadır. Ayrıca, nehirler üzerine kurulan tesislerle enerji üretimine olanak sağlamaktadır [4]. Tüm bu faaliyetler sonucunda akarsular çevre kirliliğinden en çok etkilenen su kaynaklarıdır [5].

Günümüzde yerüstü su kaynaklarının su kalitesinin belirlenmesi ve buna bağlı olarak oluşturulacak su kalite yönetimi tüm dünyada önemli araştırma konuları arasında yer almaktadır. Su kalitesi ile ilgili yapılan çalışmalar, suyun kullanım amacının

belirlenmesinde önemli olmasının yanı sıra, yüzey su kaynaklarının sürdürülebilir kullanımı açısından da önemlidir. Su kaynakları endüstriyel ve tarımsal faaliyetler sonucu özellikle alıcı ortam olarak kullanılmaları nedeniyle hızlı bir 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 gerekmektedir [6, 7].

İstasyonlar arasındaki su kalitesinin değişiminde etkili olan faktörlerin tespiti suyun korunması açısından önemlidir ve uzun yıllardır bu alandaki çalışmalar devam etmektedir. Su kalitesinin değerlendirildiği çalışmalar incelendiğinde genellikle faktör analizi, kümeleme analizi, diskriminant analizi ve ANOVA gibi istatistiksel yöntemlerin kullanıldığı görülmektedir. Bu analizler içerisinde faktör analizi birbiriyle yüksek derecede ilişkili su kalite karakteristiklerinin, faktör adı verilen bağımsız yapılara dönüştürülmesinde kullanılmaktadır [8]. Ancak, boyut indirgemedeki bilgi kaybı yaşanmaktadır. Diskriminant analizinde ise, ait oldukları grupların belli olduğu gözlemleri çeşitli özellikleri bakımından ayırmayı sağlayacak ayırıcı bir fonksiyon belirlenmesi amaçlanmaktadır. Ancak her iki analiz içinde verilerin çok değişkenli normal dağılıma sahip olması gerektirir [9]. Yapılan bir bilimsel çalışmanın geçerlilik ve güvenilirliği, uygulanan yöntemin dayandığı varsayımların sağlanması ile doğru orantılıdır.

Lojistik regresyon analizi normallik varsayımının sağlanmaması durumunda diskriminant analizinin alternatifi olarak kullanılmaktadır. Hangi değişkenlerin su kalitesinin değişimi üzerinde etkili olduğunun belirlenmesine olanak sağlamaktadır. Lineer regresyon analizinde olduğu gibi, lojistik regresyon analizinde de bağımlı değişken ile bağımsız değişkenler arasındaki fonksiyonel ilişki kurulmaya çalışılmaktadır. Ayrıca, lojistik regresyon analizinde de bağımsız değişkenler arasında yüksek derecede korelasyonların olması problem yaratmaktadır. Su kalite karakteristikleri gibi birbiriyle doğal korelasyonlu olan değişkenlerin analizi için bu yöntemlerin kullanılması, su kalitesinin sınıflandırılmasında elde edilecek ayırıcı fonksiyonlar üzerinde olumsuz etkiye neden olmaktadır. Bu nedenle son dönemlerde lojistik regresyon analizinde hem değişken seçimi yapan hem de model başarısı yüksek olan yöntemler kullanılmaya başlanmıştır. Bu çalışmada lojistik regresyon, Lojistik-Ridge ve Lojistik-LASSO yöntemleri kullanılarak, su kalitesinin değişimini etkileyen değişkenlerin belirlenmesi amaçlanmıştır. Çalışma alanı olarak Sakarya Nehri üzerinde bulunan gözlem istasyonlarından 2 istasyon ele alınmıştır. 2014-2018 yılları arasındaki veriler kullanılarak iki istasyon arasındaki su kalitesi ölçümlerinin değişimine sebep olan değişkenler belirlenmeye çalışılmıştır. İlgili değişkenlerin belirlenmesi su kalitesinin sürdürülebilir olmasına, ilgili karakteristikler bakımından daha sıklıkla analiz yapılmasına ve çevre kirliliğinin azaltılmasında alınacak tedbirler bakımından faydalı olacaktır.

MATERYAL VE METOD

Çalışma Alanı ve Veri Toplama

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, Kuzeypbatı Anadolu'nun ise en büyük akarsuyudur. Eskişehir'in Çifteler ilçesi yakınlarından doğan Sakarya Nehri kolları ile birlikte toplam uzunluğu 824 km.'dir [10]. 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 sularından, tarım faaliyetleri sonucunda oluşan atıklardan, 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 [6]. Bu çalışmada Akmeşe Deresi ve İnegöl Yenicedere Mansap (K10) istasyonlarına ait 2014-2018 yıllarına ait su kalite verileri Devlet Su İşleri'nden (DSİ) alınmıştır. Bu yıllar arasında her iki istasyonda düzenli olarak ölçülen su kalite karakteristikleri şu şekildedir: BOD5 (Biyolojik Oksijen İhtiyacı-mg/L), Ca⁺⁺ (Kalsiyum-mg/L), Cl⁻ (Klorür-mg/L), DO (Çözünmüş Oksijen- mgO₂ /L), EC (Elektriksel İletkenlik- Mikromhos/cm), K⁺ (Potasyum-mg/L), M-Al (mg/L), Mg⁺⁺ (Magnezyum-mg/L), Na⁺(Sodyum-mg/L), NH₄N (Amonyum Azotu-mg/L), O-PO₄ (Orta Fosfat-mg/L), pH (Alkalinite), pV (Permanganat Değeri-mgO₂/L), SO₄ (Sülfat-mg/L), T (Sıcaklık-°C), TDS (Toplam Çözünmüş Katılar-mg/L), TH (Toplam Sertlik-mg/LCaCO₃). Akmeşe Deresi 1 ve İnegöl Yenicedere Mansap istasyonu 0 şeklinde kodlanarak iki istasyon arasındaki değişime neden olan değişkenler belirlenmeye çalışılmıştır. Bu istasyonların konumuna ilişkin harita Şekil 1'de verilmiştir.



Fig. 1. Akmeşe Deresi ile İnegöl Yenicedere Mansap İstasyonlarına ait Konum

Metot

Bu çalışmada Sakarya Nehri üzerinde yer alan Akmeşe Deresi ve İnegöl Yenicedere Mansap istasyonlara ait veriler kullanılarak, Lojistik Regresyon, Lojistik Ridge ve Lojistik-LASSO yöntemleri yardımıyla iki istasyon arasında su kalitesi değişimine etki eden değişkenler belirlenmeye çalışılmıştır. Analizler R programlama dili kullanılarak yapılmıştır.

Lojistik Regresyon Analizi

Lojistik regresyon, sınıflandırma problemlerinde diskriminant analizinin alternatifi olarak sıklıkla kullanılan bir analizdir yöntemidir. Ayrıca son yıllarda makine öğrenme yöntemleri arasında popüleritesi artmaktadır. Lojistik regresyon modeli, bağımlı değişkenin ikili veya daha fazla kategorik değer aldığı durumda kullanılmaktadır. 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 [11]. Lojistik regresyon analizinde, bağımsız değişkenlerin doğrusal kombinasyonları ile bağımlı değişken doğrusal olmayan ilişkiye sahiptir [12,13]. İkili lojistik regresyon analizinde, bağımlı değişkenin düzeylerini ifade edilirken ilgilenilen durum için 1 ve diğer durum için 0 değerleri kullanılır. Bu nedenle bağımlı değişken Bernoulli dağılımına sahiptir. $\beta' = [\beta_0, \beta_1, \dots, \beta_p]$ bilinmeyen parametreler vektörünü, x_j , $p \times 1$ boyutlu bağımsız değişkenler vektörünü göstermek üzere, $E(Y|X = x_i) = \pi(x_i)$ koşullu olasılığı olmak üzere, lojistik regresyon modelinin koşullu olasılığı Eşitlik (1)' de verilmiştir.

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

Eqn. 1

Eşitlik (1) için log-olabilirlik fonksiyonu;

$$L(\beta_0, \beta) = \sum_{i=1}^n \{y_i \ln(x_{ij}) + (1 - y_i) \ln(1 - \pi(x_{ij}))\} \quad j = 1, 2, \dots, p$$

Eqn. 2

Eşitlik (2) ile tanımlanır. Parametre tahminleri log-olabilirlik fonksiyonunun maksimize edilmesiyle bulunur. Nümerik çözüm için Newton-Raphson algoritması kullanılır. Lojistik regresyon modelinin lojit dönüşümü Eşitlik (3)' de verilmiştir.

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

Eqn. 3

Lojistik regresyon analizinde anlamlı bulunan katsayıların yorumlanmasını kolaylaştırmak için Odd's oranı kullanılabilir. Odds oranı, bağımsız değişken değerindeki bir birim değişime karşılık gelen başarı olasılığındaki artış olarak da tanımlanabilir [14]. Odds oranı Eşitlik (4)' de verilmiştir.

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

Eqn.4

Regresyon modellerinde bağımsız değişkenler arasında yüksek derecede korelasyon olması çoklu iç ilişki problemi olarak nitelendirilir [15]. Lojistik regresyon analizinde çoklu iç ilişki problemi, tahminlerin etkinliğini azaltan bir problemdir. Bu durum yanlış sınıflandırmaya neden olmaktadır. Su kalitesi ile ilgili çalışmalarda, kalite değişkenlerinin birçoğu birbiriyle yüksek derecede ilişkilidir. Çoklu iç ilişki probleminin giderilmesi için çeşitli yöntemler vardır. Bunlar arasında Lojistik Ridge regresyon ilk sırada gelmektedir.

Lojistik-Ridge Regresyon

Doğrusal regresyonda çoklu iç ilişki probleminin giderilmesi için belirlenen yöntemler bazı araştırmacılar tarafından lojistik regresyon modeline uyarlanmıştır [16, 17]. Bağımlı değişken iki düzeyli olduğu durumda lojistik ridge tahmin edicisi kullanılabilir. İlk kez Duffy ve Santer (1989) [18] tarafından önerilen lojistik ridge regresyonu Eşitlik (5)' de verilmiştir.

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

Eqn. 5

Lojistik regresyon modelinde matris gösterimiyle ridge regresyonun En Çok Olabilirlik tahmin edicisi Eşitlik (6)' da verilmiştir.

$$\hat{\beta}_{E\check{C}O} = (X' \hat{V} X)^{-1} X' \hat{V} Z$$

Eqn. 6

Burada, \hat{V} ağırlık matrisi $\hat{V} = \operatorname{diag}[\hat{\pi}(1 - \hat{\pi})]$ şeklindedir.

Lojistik ridge regresyonda, ayar parametresi λ büzülme miktarını kontrol eder, ancak hiçbir zaman bağımsız değişken katsayılarını tam olarak sıfır yapmaz. Bu durum, model katsayılarının yorumlanmasını zorlaştırmakta ve her parametre için tahmini zorunlu kıldığından tahmin değerinin yanlışlığını arttırmaktadır. 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) [19] tarafından önerilmiştir. Bağımlı değişkenin ikili olduğu durumlarda kullanılan lojistik regresyon modeli LASSO yöntemine uyarlanmıştır [20].

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 (7)'de verilmiştir.

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

Eqn. 7

Eşitlik (7)' de λ terimi tahminleri cezalandıran ceza terimidir. Burada, λ 'nın değeri çapraz geçerlilik (Cross-validation) yöntemiyle belirlenir. Sınıflandırma probleminde, λ parametresi yanlış sınıflandırma hatasını en aza indirmek için sapma ve varyans arasında doğru dengeyi sağlamaktadır [12, 13]. LASSO cezasının özelliğine bağlı olarak, bazı katsayılar tam olarak sıfıra eşit olacaktır.

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 [21]. Çapraz geçerlilik katsayısı Eşitlik (8)' de verilmiştir.

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

Eqn. 8

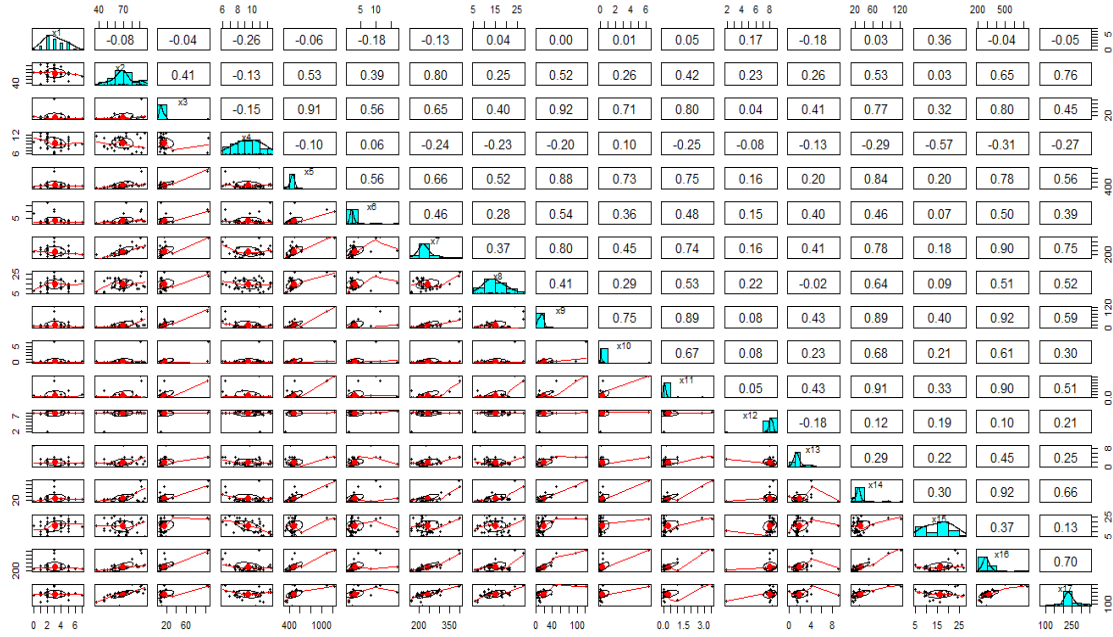
Çapraz geçerlilik, bir modelin bağımsız bir veri kümesine ne kadar iyi genelleştirilebileceğini değerlendiren bir yöntem olarak ifade edilebilir [20].

BULGULAR VE TARTIŞMA

Bu bölümde yukarıda açıklamaları verilen istasyonlara ait veriler için Lojistik Regresyon, Lojistik Ridge ve Lojistik-LASSO yöntemleri kullanılarak regresyon modelleri oluşturulmuştur. 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 sınavları için ise kovaryans testinden yararlanılmıştır. 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.

Şekil 2 de verilen haritada gösterilen istasyonların su kalite karakteristikleri arasındaki ilişkilerin belirlenmesi için korelasyon matrisi hesaplanmış ve Tablo 1'de verilmiştir.

Tablo 1. Akmeşe Deresi ve İnegöl Yenicedere Mansap İstasyonu için Korelasyon Matrisi Tablosu



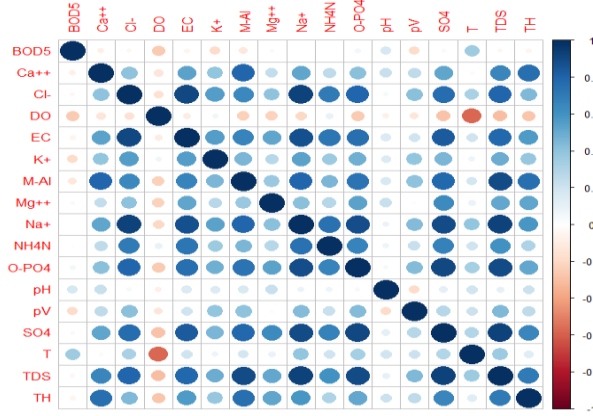


Fig. 2. Birinci Karşılaştırma için Korelasyon Grafiği

İki istasyon arasında su kalitesinin değişimine neden olan parametrelerin tespiti için öncelikle lojistik regresyon analizi uygulanmıştır. Analiz sonuçları Tablo 2’de verilmiştir. Tablo 2’de ele alınan yöntemler yardımıyla tahmin edilen parametre değerleri ve odd’s oranları yer almaktadır. Ayrıca her bir yöntem yardımıyla elde edilen sınıflandırıcı fonksiyonların doğru sınıflandırma oranları tablonun son satırında verilmiştir.

Tablo 2 incelendiğinde lojistik regresyon analizi yardımıyla elde edilen modelin % 61,2’lik doğru sınıflandırmaya sahip olduğu görülmektedir. Lojistik Ridge regresyon analizi ile elde edilen modellerin doğru sınıflandırma yüzdesi % 79,4’tür. Her iki yöntemde tüm parametreler için bir tahminde bulunmaktadır. Bir başka ifadeyle, değişken indirgemesi olmaksızın tüm model katsayılarını tahmin etmektedir. Lojistik-LASSO yönteminin doğru sınıflandırma yüzdesi ise % 87,5 olarak bulunmuştur.

Tablo 2. Akmeşe Deresi ve İnegöl Yenicedere Mansap İst. ait Analiz Sonuçları

Parametre	Lojistik Regresyon		Lojistik Ridge Regresyon $k = 0,26$		Lojistik LASSO Regresyon $\lambda = 0,245$	
	<i>b</i>	Odds	<i>b</i>	Odds	<i>b</i>	Odds
Sabit	26,148	-	26,48	-	1,045	2,84
BOD ₅	-0,398	0,67	-4,286	0,014	-0,188*	0,83
Ca ⁺⁺	0,115	1,12	0,393*	1,480	0,061*	1,06
Cl ⁻	0,355*	1,43	-2,545	0,078	-	-
DO	0,564*	1,76	0,356*	1,427	0,061*	1,06
EC	-0,039*	0,96	-2,784*	0,062	-0,001*	1
K ⁺	1,641*	5,16	-2,229	0,11	-	-
M-Al	0,076	1,08	-0,96	0,38	-	-
Mg ⁺⁺	-0,423	0,66	-4,121*	0,017	-0,140*	0,87
Na ⁺	-0,237	0,79	-2,948	0,052	-	-
NH ₄ N	-1,228	0,29	-1,986*	0,14	-	-
O-PO ₄	1,861*	6,43	1,419*	4,13	0,134*	1,14
pH	-3,592	0,028	0,834*	2,3	-0,281*	0,76
Pv	-1,893	0,15	-2,254	0,1	-	-
SO ₄	0,207*	1,23	-0,813	0,44	-	-
T	0,373*	1,45	1,357*	3,88	0,047	1,05
TDS	-0,042	0,96	-2,133*	0,12	-	-
TH	0,005	1,005	1,490*	4,44	0,002	1,002
DSO	61,2		79,4		87,5	

(•0.05 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.)

Tablo 2 incelendiğinde, biyolojik oksijen ihtiyacı (BOD₅) 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. Yapılan analizler sonucunda odd's oranları hesaplandığında lojistik regresyon analizine göre BOD₅ değerinin İnegöl Yenicedere Mansap istasyonuna göre Akmeşe istasyonunda 0,67 kat azaldığı, Lojistik Ridge yöntemine göre ise 0,014 kat azaldığı ve son olarak Lojistik-LASSO regresyon yöntemi için 0,83 kat azaldığı görülmektedir. Aynı zamanda BOD₅ 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 BOD₅ değerinin negatif işaretli çıkması İnegöl Yenicedere Mansap (K10) istasyonunda BOD₅ 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⁺⁺ ve Mg⁺⁺ miktarının kalsiyum karbonat (CaCO₃) eş değeri olarak tanımlanmıştır. Bunun sebebi, sudaki mevcut iyonlara göre daha fazla oranda Ca⁺⁺ ve Mg⁺⁺ iyonlarının bulunmasıdır [22]. Yapılan analizler sonucunda Ca⁺⁺ değerinin odd's oranları incelendiğinde lojistik regresyon yönteminde 1,12 kat arttığı, Lojistik Ridge yöntemi için 1,48 kat arttığı ve Lojistik-LASSO yöntemi için ise 1,06 kat arttığı görülmektedir. Tüm yöntemler için İnegöl Yenicedere Mansap istasyonuna göre Akmeşe Deresi'nde Ca⁺⁺ 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 Ridge yöntemi için 4,44 kat arttığı ve Lojistik-LASSO yöntemine göre ise 1,002 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⁺⁺ dur. Yapılan analiz sonuçları incelendiğinde, her bir yöntem için magnezyum değişkenine ait Odd's oranları hesaplanmıştır. Lojistik regresyon yöntemi için Mg⁺⁺ değerinin İnegöl Yenicedere Mansap istasyonuna göre Akmeşe istasyonunda 0,66 kat azaldığı, Lojistik Ridge yöntemine göre 0,017 kat azaldığı ve son olarak Lojistik-LASSO yöntemi için ise 0,87 kat azaldığı gözlenmektedir. Bu durumda Akmeşe Deresi istasyonundaki suyun daha sert olduğu şeklinde yorum yapılabilir.

Yüksek klorür konsatrasyonu bir kirlenme belirtisi olmakla birlikte önemli bir kalite parametresidir [23]. Yapılan analizler sonucunda klorürün odd's oranı hesaplandığında, Lojistik regresyon yöntemi için Cl⁻ değerinin İnegöl Yenicedere Mansap istasyonuna göre Akmeşe Deresi istasyonunda 1,43 kat arttığı, Lojistik Ridge yöntemine göre 0,78 kat azaldığı gözlemlenmiştir. Lojistik-LASSO yöntemi için ise iki istasyon arasındaki Cl⁻ değerinin değişiminin anlamlı olmadığı belirlenmiştir.

DO, sulardaki oksijen miktarını belirlemek için kullanılan bir parametredir. Canlı yaşamı için kritik öneme sahip olan çözülmüş 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 Ridge yöntemi için 1,427 kat arttığı ve son olarak 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 olduğunu bir kanıtıdır.

EC, suyun elektrik akımını iletibilmesinin 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 [24,25]. Yapılan analiz sonuçları incelendiğinde İnegöl Yenicedere Mansap istasyonuna göre Akmeşe Deresi istasyonunda, Lojistik regresyon sonuçlarına göre elektriksel iletkenliğin 0,96 kat azaldığı, Lojistik Ridge yöntemi için 0,062 kat azaldığı ve son olarak Lojistik-LASSO yöntemi için ise 1 kat azaldığı görülmektedir. Bu sonuçlar birlikte değerlendirildiğinde, İnegöl Yenicedere Mansap istasyonuna göre Akmeşe Deresi istasyonunda iyon miktarının daha az olduğunu göstermektedir.

K^+ iç sulara 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 istasyonuna göre Akmeşe Deresi istasyonunda K^+ değerinin 5,16 kat arttığı, Lojistik Ridge yöntemi için 0,11 kat azaldığı gözlemlenmiştir. Lojistik-LASSO yöntemi için ise K^+ değişkeninin değişiminin anlamlı olmadığı belirlenmiştir.

$O-PO_4$ 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 Ridge yöntemine göre 4,13 kat arttığı ve son olarak Lojistik-LASSO yöntemine göre 1,14 kat arttığı gözlemlenmiştir. 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 karıştığının bir göstergesi olabilir.

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 [26]. Yapılan analizler sonucunda Odd's oranları incelendiğinde, lojistik regresyon analizine göre NH_4N değerinin İnegöl Yenicedere Mansap istasyonuna göre Akmeşe deresi istasyonunda 0,29 kat azaldığı, Lojistik Ridge yöntemi için 0,14 kat azaldığı gözlemlenmiştir. Bu iki yöntem için Akmeşe Deresi istasyonunda NH_4N miktarının İnegöl Yenicedere Mansap istasyonuna göre daha az olduğu görülmektedir. Bu durum Akmeşe Deresi istasyonu çevresinde evsel veya endüstriyel kirlenme yaratacak yerleşim yerleri, fabrikaların bulunmaması bu sonucu desteklemektedir. Lojistik-LASSO yöntemine göre ise NH_4N değerinin istatistiksel açıdan anlamlı olmadığı gözlemlenmiştir.

Na^+ suda kolaylıkla çözünen ve suya tuzlu bir tat veren bir katyondur. Na^+ içeren sular tarımsal sulamalarda toprak ve bitkiler üzerinde zararlı etkiler yapmaktadır [23]. Yapılan analizler sonucunda odds oranları incelendiğinde İnegöl Yenicedere Mansap (K10) istasyonuna göre Akmeşe Deresi istasyonunda, Na^+ değerinin lojistik regresyon analizi yöntemine göre 0,79 kat azaldığı, Lojistik Ridge yöntemine göre de 0,052 kat azaldığı gözlemlenmiştir. Lojistik-LASSO yöntemine göre ise iki istasyon arasındaki Na^+ değerindeki değişiminin anlamlı olmadığı belirlenmiştir. Buradan hareketle lojistik ve

lojistik Ridge yöntemlerine ait sonuçlara dayanarak İnegöl Yenicedere Mansap istasyonunda daha fazla Na⁺ elementi bulunduğu ve suyun daha tuzlu olduğu söylenebilir.

SONUÇ

Bu çalışmada, Sakarya nehri üzerindeki iki farklı noktadan 2014-2018 yıllarına ait su kalite verilerinin istatistiksel analizi gerçekleştirilmiştir. Su kalite çalışmalarında sürdürülebilirlik ölçümlerin düzenli bir şekilde gerçekleştirilmesine bağlıdır. İstasyonlar arasında farklılığa neden olan su kalite karakteristiklerinin tespiti, daha sıklıkla ölçümlerin gerçekleştirilebilmesine ve önlemlerin alınmasına olanak sağlayacaktır. Bu çalışmada Sakarya nehri üzerindeki 2 istasyon arasında su kalitesinin değişimine neden olan parametreler lojistik regresyon, lojistik-LASSO ve lojistik-Ridge yöntemleri ile belirlenmiştir. Lojistik ve lojistik-Ridge regresyon analizlerinde tüm kalite karakteristikleri modelde yer alırken, modelde anlamsız bulunan birçok değişkenin olduğu görülmüştür. Bir başka ifadeyle, iki istasyon arasındaki su kalitesi değişiminde etkili olmayan değişkenlerde modelde yer almaktadır. Lojistik-LASSO yönteminde ise daha az sayıda değişkenin su kalitesi değişimi üzerinde etkili olduğu belirlenmiştir. Lojistik-LASSO regresyonda, lojistik regresyona göre daha katı bir değişken seçimi söz konusu olduğundan 9 su kalite değişkeninin iki istasyon arasındaki su kalite değişimden sorumlu olduğu belirlenmiştir. Lojistik-LASSO regresyona göre iki istasyon arasındaki su kalitesinin tespitinde bu değişkenlerin kullanılması %87,5 doğru sınıflandırma oranı ile doğru sonuç verecektir. Lojistik-LASSO regresyon yönteminde pozitif işaretli olan Ca⁺⁺, O-PO₄, T ve TH su karakteristiklerinin pozitif işaretli olduğu görülmektedir. Buna ek olarak Lojistik-LASSO yöntemiyle yapılan analiz sonucunda anlamlı çıkan diğer değişkenler BOD₅, DO, EC, Mg⁺⁺, pH'tır. Lojistik regresyon analizine göre tüm değişkenlerin kullanılması durumunda doğru sınıflandırma oranı %90.1'dir. Ancak istatistiksel olarak anlamlı bulunmayan değişkenlerinde modelde yer alması, modelden elde edilecek sonuçların güvenilirliğini azaltmaktadır. Su kalite karakteristiklerinin iki istasyon arasında değişim oranlarının yorumlanmasında ve sürdürülebilir bir su kalitesi çalışmasının yürütülmesinde lojistik-LASSO yönteminden elde edilen sonuçların kullanılması yararlı olacaktır.

KAYNAKLAR

- [1] Bulut, C., Kubilay, A. (2019): Eğirdir Gölü (Isparta/Türkiye) su kalitesinin mevsimsel değişimi. Su Ürünleri Dergisi 36(1): 13-23.
- [2] 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.
- [3] 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): 105-118.
- [4] 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.
- [5] Zeydan, Ö., Özdoğan, N., Taştepe, Ş. P., Demirtaş, D. (2019): Kozlu Deresinde (Zonguldak) Su Kalitesinin İncelenmesi. Artvin Çoruh Üniversitesi Doğal Afetler Uygulama ve Araştırma Merkezi, Doğal Afetler ve Çevre Dergisi 5(2): 187-197.

- [6] Özer, Ç., 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 5(2): 237-246.
- [7] 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.
- [8] Yu, S., Shang, J., Zhao, J., Guo, H. (2003): Factor analysis and dynamics of water quality of the Songhua River Northeast China. Water, Air, and Soil Pollution 144, 159-169. <https://doi.org/10.1023/A:1022960300693>
- [9] Zhang, Y., Guo, F., Meng, W., Wang, X. (2008): Water quality assesment and source identification of Daliao river basin using multivariate statistical methods. Environ. Monit. Assess. 152:105-121. DOI: 10.1007/s10661-008-0300-z
- [10] <http://www.sakarya.gov.tr/sakarya-nehri> (Erişim Tarihi: 04.02.2020).
- [11] Bircan, H. (2004): Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama. Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi 2: 185-208.
- [12] 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 42(23): 9326-9332.
- [13] Algamal, Z. Y., Lee, M. H. (2015): High dimensional logistic regression model using adjusted elastic net penalty. Pakistan Journal of Statistics and Operation Research 667-676.
- [14] Erar, A. (2013): Doğrusal Regresyon Analizine Giriş, Nobel Yayınevi, 292-320 s.
- [15] Montgomery, D. C., Peck, A. E., Vining, G. G. (2001): Introduction to linear regression analysis, third edition, John Wiley & Sons.Inc., 641 p.
- [16] Schaefer, R. L., Roi, L. D., Wolfe, R. A. (1984): A ridge logistic estimator. Communications in Statistics-Theory and Methods 13(1): 99-113.
- [17] Weissfeld, L. A., Sereika, S. M. (1991): A multicollinearity diagnostic for generalized linear models. Communications in Statistics-Theory and Methods, 20(4): 1183-1198.
- [18] Duffy, D. E., Santner, T. J. (1989): On the small sample properties of norm-restricted maximum likelihood estimators for logistic regression models. Communications in Statistics-Theory and Methods 18(3): 959-980.
- [19] Tibshirani, R. (1996): Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58(1): 267-288.
- [20] Hastie, T., Tibshirani, R., Wainwright, M. (2019): Statistical learning with sparsity: the lasso and generalizations. Chapman and Hall/CRC.
- [21] Fıdanoglu, I. (2009): İstatistiksel Daraltıcı (shrinkage) Model ve Uygulamaları, Çukurova Üniversitesi, Yüksek Lisans Tezi.
- [22] Şengün, E. (2013): Aksu Deresi Su Kalitesi ve Kirlilik Düzeyinin Belirlenmesi, Giresun Üniversitesi Fen Bilimleri Enstitüsü, s.66.
- [23] Tuncay, H. (1983): Su Kalitesi, Ders Notları, Ege Üniversitesi Ziraat Fakültesi Toprak Bölümü, İzmir.
- [24] Uslu, O., Türkman, A. (1987): Su Kirliliği ve Kontrolü., T.C. Başbakanlık Çevre Genel Müd. Yayınları Ankara, No: 1, 344-345.
- [25] Sawyer, C., McCarty P., Parkin G. (2013): Çevre Mühendisliği ve Bilimi İçin Kimya, Çeviri Editörü: İsmail TORÖZ.
- [26] Egemen, Ö., Sunlu, U. (1999): Su Kalitesi (Ders Kitabı) III. Baskı, Ege Üniversitesi Basımevi, İzmir.