

# SU KALİTESİ İNDEKSİNİN MAKİNE ÖĞRENİMİ ALGORİTMALARI YARDIMIYLA İNCELENMESİ

**Fatma Sena YÜKSEL**

## ÖZET

Su, yaşamın temel taşıdır ve insanların sağlıklı bir şekilde yaşayabilmesi, ekonomilerin büyümesi için gerekli bir unsurdur. su kaynakları sınırlıdır Dünya'nın yüzde 70'i su ile kaplı olmasına rağmen, kullanılabilir tatlı su kaynakları azalmakta ve su kirliliği sorunu artmaktadır. Dünya nüfusundaki hızlı artış, endüstrileşme ve kentleşme süreci, bu sorunu daha da karmaşık hale getirmektedir. Bu durum, canlıların sağlıklı yaşamını sürdürebilmesi ve ekonomilerin büyümesi için gereken temiz su kaynaklarının azalmasına yol açmaktadır. Su kirliliğinin azaltılması ve temiz su kaynaklarının artırılması için atıksuların etkin bir şekilde arıtılmasını gerekmektedir. Atıksuların arıtımında su miktarı ve kalitesinin etkin bir şekilde yönetilebilmesi için mevcut yüzey suyunun kirlilik düzeyinin belirlenmesi gerekmektedir. Ardından su kalite indeksleri yardımıyla suyun içilebilirliği test edilip sonuçlar kontrol altında tutulmalıdır. Süreç su kalite indeksleri yardımıyla daha kolay ve etkili bir şekilde gerçekleştirilebilmektedir. Bu çalışmada, kaggle.com sitesinden elde edilen “*Water Potability*” veri seti kullanılmıştır. Veri seti 21 farklı özelliği içermektedir. Su kalitesini, suyun içilebilirliğini anlamak için bu veri seti üzerinde çeşitli makine öğrenimi yöntemleri uygulanmıştır, bunlar arasında *K-NN*, *Karar Ağaçları*, *Yapay Sinir Ağları*, *Destek Vektör Makineleri* ve *Naive Bayes* yöntemleri bulunmaktadır. Bu yöntemlerin performansları değerlendirilmiş ve en uygun sonuca ulaşılmaya çalışılmıştır. En optimize sonuç % 95 doğruluk değeriyle yapay sinir ağlarından alınmıştır.

**Anahtar Kavramlar:** *Su Kalitesi, Makine Öğrenimi, Sınıflandırma, Destek Vektör Makineleri, Yapay Sinir Ağları, Karar Ağaçları, K-NN, Naive Bayes*

## 1. GİRİŞ

Doğal kaynaklardan biri olan su, insan yaşamının ve ekonomik ve çevresel kalkınmanın temel kaynağıdır. Giderek artan nüfus, hızlı kentleşme, sanayileşme ve bilinçsiz tarımsal sulama kısaca suyun bilinçsiz tüketimi kullanılabilir su kaynaklarının azalmasına ve su kalitesinin hızla düşmesine neden olmaktadır. Dünyanın yaklaşık %70'i sudan oluşmaktadır 1.4 milyar km<sup>3</sup> suya karşılık gelmektedir ancak içilebilir tatlı su kaynakları sınırlıdır Dünya üzerindeki yüzey sularının yalnızca %2.5'i tatlı su kaynaklarından oluşmaktadır. İçilebilir birincil tatlı su kaynaklarının sadece %0.3'ü göl ve akarsularda bulunmaktadır [1]. Nehir, dere, çay gibi içilebilir tatlı su kaynaklarının kalitesi, yağmur, erozyon gibi doğal olaylar, bilinçsiz kentsel, tarımsal ve endüstriyel faaliyetler, nüfus artışı, arıtılmayan kanalizasyon gibi temel unsurlar sonucu kirlilikle

karşı karşıyadır [2]. Bu kirliliğin sebep olduğu tatlı su kaynaklarındaki bozulma, canlıların yaşam kalitelerini koruyabilme ve sağlıklı bir yaşam sürdürebilmeleri için tükettikleri suyu etkilemektedir. Bu durum yakın gelecekte içme suyunun mevcudiyeti, toplumsal ve ekonomik kalkınma ve gelecekte yeterli ve kaliteli su tüketimi üzerinde ciddi problemlere sebep olacaktır.

Denizler, göller, nehirler gibi su yataklarının kalite seviyelerinin ölçülebilmesi için birçok teknolojik altyapı sistemi tasarlanmıştır. Bu sistemler suyun kalitesini biyolojik, fiziksel ve kimyasal pek çok etmene bağlı olarak ölçmektedir [3]. Su kalite standartlarına uymayan düşük kalitede su tüketimi insan sağlığını doğrudan etkileyip birçok hastalığın ortaya çıkmasına ve ölümcül sonuçlara yol açmaktadır [4]. Birleşmiş Milletler Gelişme Programı'na göre dünya üzerindeki hastalıkların %50'sinin ortaya çıkmasına bu durum sebep olmuştur [5]. Su kaynakları sonsuz olmadığından yüzey su kaynaklarının kalitesinin değerlendirilmesi ve takibinin yapılması su kaynaklarının yönetim planlarının oluşturulması gerekmektedir. Araştırmalara göre 2025 yılından sonra yaklaşık 3 milyar insanın su kıtlığıyla karşı karşıya kalacağı öngörülmektedir. Günümüzde ise her dokuz kişiden birinin güvenilir kaynaklardan içme suyuna ulaşamadığı belirtilmiştir. Su kalitesinin ve içilebilirliğinin doğru tahmini Dünya üzerindeki içilebilir suların yönetimi, su kirliliğinin kontrolü ve insan sağlığı için önem arz etmektedir.

Günümüzde birçok ülkede yüzey su kalitesinin takibi için su kalitesi gözlem istasyonlarından yararlanılmaktadır. Bu istasyonlarda su kalitesini etkileyen çeşitli faktörlerin ölçümü düzenli aralıklarla yapılmakta sonrasında elde edilen sonuçlara göre harekete geçilmektedir ama bu yöntem çok zaman alıcı ve maliyetli laboratuvar analizleri gerektirmektedir. Ayrıca belirli aralıklarla ölçüm yaptığı için su kalitesindeki değişimlerin hızını takip konusunda yetersiz kalmaktadır. Bu yüzden böyle geleneksel yöntemler yerine daha doğru ve güvenilir sonuçlar veren makine öğrenimi yöntemlerinden yararlanılmaktadır [6]. Makine öğrenimi yöntemlerinin kullanımı veri analitiği, akıllı sensörlerin maliyetinin daha uyguna gelmesi, elde edilen sonuçların pek çok işlem maliyetini azaltması yapay zekaya dayalı bu sistemlerin daha çok benimsenmesini sağlamıştır. Yeni nesil sürdürülebilir su yönetim sistemlerinde kısa ve uzun vadeli etkin tahminlerde bulunulabilmektedir bu nedenle gelişmiş ülkelerde sıklıkla kullanılmaya başlamıştır [7].

Literatürde suyun kalite ölçümünü makine öğrenimi yöntemleriyle yapan birçok çalışma bulunmaktadır. Bunlardan bazılarını örnek vermek gerekirse; Radhakrishnan ve ark. [8] su kalitesini sınıflandırma sürecinde makine öğrenme yöntemlerini kullandılar. Radhakrishnan ve ark. [8] destek karar makineleri (SVM), karar ağaçları (DT), naive bayes (NB) yöntemlerini kullanarak analizlerini gerçekleştirdiler. En iyi sınıflandırma sonucunu DT yöntemi ile sağladılar ve bu yöntem ile %98,50 oranında genel doğruluk başarıları elde ettiler. Khan ve ark. [9], su kalitesini sınıflandırma işleminde temel bileşenler analizi (PCA) yöntemi ile gradyan artırma yöntemini kullandılar. Khan ve ark. [9] çalışmalarında sırasıyla %95 ve %100 genel doğruluk başarıları elde ettiler. Venkata ve ark. [10] su kalitesini sınıflandırma sürecinde sinir ağı modellerinden yararlandılar.

Venkata ve ark. [10] tekrarlayan sinir ağı (RNN) ve LSTM modellerini deneysel analizlerde kullandılar. En iyi performans sonucunu LSTM yöntemi ile elde ettiler ve bu yöntem ile %94 oranında genel doğruluk başarıları sağladılar.

Bu çalışmada kaggle.com sitesinden alınan *Water Potability* su içilebilirliğinin belirleyici unsuru olan su kalite indeksine göz önünde bulundurularak makine öğrenimi yöntemleriyle analiz yapılacaktır. Kullanılacak yöntemler ise şunlardır: En Yakın Komşu Algoritması (K-NN), Karar Ağaçları (DT), Destek Vektör Makineleri (SVM), Yapay Sinir Ağları (YSA), Naive Bayes (NB). Bu yöntemlerin performansları çeşitli kriterlerde karşılaştırılarak en optimize sonuç bulunmaya çalışılmıştır.

## 2. METOD VE YÖNTEMLER

### 2.1. Veri Seti ve Ön İşleme

Kullanılan veri kümesi açık erişimlidir ve ‘.csv’ uzantılı dosyadan oluşmaktadır. Veri kümesinde toplam 20 öznitelik sütunu bir adet de etiket sütunu yer almaktadır. Etiket sütunu olan 21. Sütunda veriler, suyun öznitelikleri değerlendirilerek “1:güvenli, içilebilir” ve “0: güvenli değil, içilemez” olarak etiketlenmiştir. Toplam 7999 adet numune değeri yer almaktadır, veri kümesindeki her bir öznitelik parametresindeki değerler bir litre içerisindeki su seviyesi dikkate alınarak oluşturulmuştur. Veri setindeki numuneler incelendiğinde eksik veriler tespit edilip silinmiş ardından 1 ve 0 etiketine sahip verilerin sayısı eşit olacak şekilde veriler temizlenmiştir. Temizlenen verilere normalizasyon uygulanıp randomize edildikten sonra eğitim ve test verisi olarak belirli oranda ayrılmıştır.

No	Öznitelik	Tehlike Eşik Değeri	No	Öznitelik	Tehlike Eşik Değeri
1	Alüminyum	2,8	11	Virüsler	0
2	Amonyak	32,5	12	Kurşun	0,015
3	Arsenik	0,01	13	Nitratlar	10
4	Baryum	2	14	Nitritler	1
5	Kadmiyum	0,005	15	Cıva	0,002
6	Kloramin	4	16	Preklorat	56
7	Krom	0,1	17	Radyum	5
8	Bakır	1,3	18	Selenyum	0,5
9	Florür	1,5	19	Gümüş	0,1
10	Bakteri	0	20	Uranyum	0,3

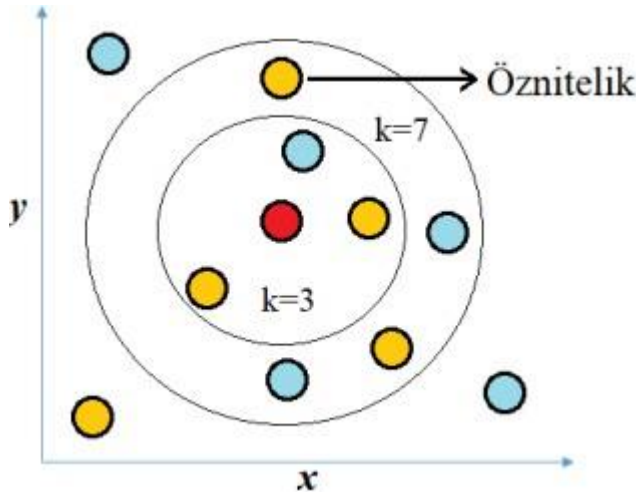
**Tablo 1.** Veri kümesinde yer alan öznitelikler ve açıklamaları

## 2.2. Makine Öğrenimi Yöntemleri

Sınıflandırmada temel amaç, nesnelerin sahip olduğu özelliklere bakılarak nesnelerin hangi sınıfa ait olduğunun belirlenme işlemidir. Çok farklı sınıflandırma türleri ve algoritmaları bulunmaktadır. Bu çalışmada kullanılan Karar Ağaçları, En Yakın Komşu, Naive Bayes, Yapay Sinir Ağları bunlardan bazılarıdır.

### 2.2.1. En Yakın Komşuluk Algoritması (K-NN)

KNN algoritması ya da diğer adıyla K-En yakın komşu algoritması makine öğrenim algoritmaları içerisinde en çok bilinen ve kullanılan algoritmalarından biridir. En yakın komşu (K-NN) yöntemi, girdi verisini kendisinden önceki girdi verilerinin yakınlık derecesine bağlı olarak sınıflandırma ve regresyon işlemini gerçekleştirebilen bir makine öğrenme yöntemidir. K-NN yöntemi mevcut girdi verilerinin durumunu saklar, yeni veri durumlarını da mevcut veri durumlarını dikkate alarak benzerlik ölçütlerini elde eder. Benzerlik olasılığı hesaplanırken özniteliklerin bulunduğu uzayda iki öznitelik arasındaki mesafeler hesaplanır. Ardından komşu öznitelikler ile mesafeler hesaplanır ve özniteliğin sınıflandırma sürecinde komşu öznitelikler arasında mesafe uzunluğu en yakın olan özniteliğin sınıf etiketine ataması gerçekleştirilir. Çalışma şekline baktığımızda, tanımlanan verilere göre yeni bir tanımlanması gereken nesne geldiğinde öncelikle K değerine bakılır. Burada eşitlik olmaması için genellikle  $k$  değeri  $\{1,3,5,10,\dots\}$  gibi tek sayı değerleri seçilir. Örneğin;  $k=3$  olarak tercih edildiğinde uzay kümesinde mevcut özniteliğin etrafındaki en yakın üç özniteliğe göre uzaklık mesafeleri ölçülür ve mevcut öznitelik kendisine en yakın özniteliğin sınıf türüyle etiketlenir [11,12].  $k$  değerinin ikili sınıflandırma işleminde seçimi şekil 1’de gösterilmiştir. Bu çalışmada da sırasıyla 1, 3, 5, 10 değerleri denenerek bu veri setinde en optimize sonucu hangisinin verdiği incelendi. Veriler arasındaki mesafeler hesaplanırken ise genellikle Kosinüs, Öklid ya da Manhattan uzaklığı gibi yöntemler kullanılır.



Şekil 1. K-NN yönteminde  $k$  değerine göre özniteliklerin seçimi

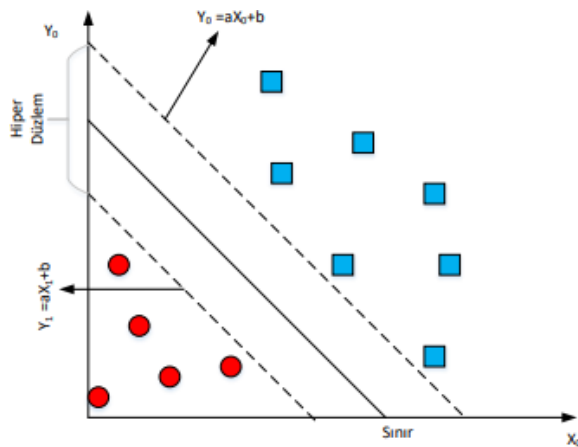
### 2.2.2. Karar Ağaçları (DT)

Karar ağacı (DT) yöntemi, karar ve yaprak düğümlerinden oluşan sınıflandırma ve regresyon işlemi için entropi değerlerini hesaplayan grafiksel tabanlı şematik bir makine öğrenme yöntemidir. Bu yöntemde sınıflandırma ve regresyon işlemleri için (ID3, C4.5, CART, vb.) algoritmalar kullanılır. Algoritma şemasında sırasıyla kök, düğüm ve yapraklar yer alır. Verilerin belirsizlikleri hesaplanırken entropi ( $E$ ) değerleri dikkate alınır. Entropi ( $E$ ) değeri olasılık değerlerinin hesaplanmasında kullanılan önemli bir parametredir. Entropi parametresinin hesaplanmasında aşağıdaki denklemde belirtilen formül kullanılır. Denklem incelendiğinde,  $N$  değişkeni toplam öznitelik sayısını temsil eder ve  $P$  değişkeni ise  $i$ . özneliliğin olasılık değerini temsil eder [14,15]. Sınıflandırma sürecinde entropi değeri minimum olan yaprak tercih edilir ve etiketlenir.

$$E = - \sum_{i=1}^N P_i \log_2 P_i$$

### 2.2.3. Destek Vektör Makineleri (SVM)

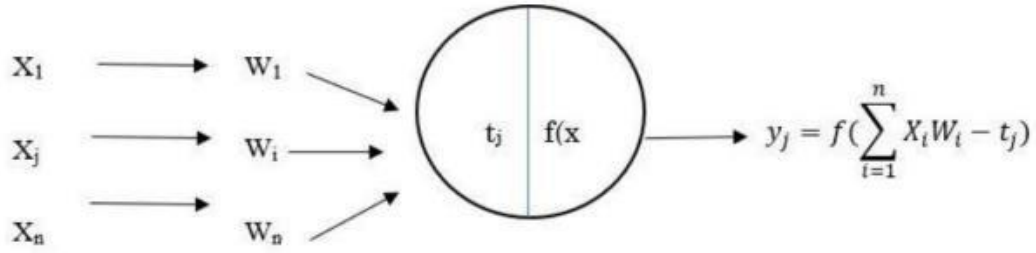
Destek Vektör Makineleri bir makine öğrenmesi tekniğidir. Yaygın olarak biyomedikal uygulamalarda ikili sınıflandırma problemleri için kullanılmaktadır. SVM’de amaç, iki sınıf arasındaki uzaklığın maksimum olduğu durumun bulunmasıdır. Bu verileri iki sınıfa ayıracak çok fazla sayıda ayırıcı düzlem bulunmaktadır. Optimum ayırıcı düzlem, her iki sınıfın en uç verileri arasındaki mesafenin (iki sınıfın destek vektörleri arasındaki mesafe) maksimum olduğu durumu sağlayan ayırıcı düzlemdir. Yapılacak işlemler olarak verileri lineer olarak ayrılabilirliği yüksek bir boyuta aktarmaktadır. Sonrasında da bunlar arasındaki maksimum sınırın bulunmasını sağlamaktadır [15]. Genel olarak SVM’ler problemleri, lineer olan ve olmayan olmak üzere ikiye ayırmaktadır. Lineer problemlerde SVM’nin kullanım amacı özellikler arasından geçen bir hiper düzlemi bulmaktır. SVM’nin temel amacı, +1 etiketli vektöre ait bir düzlem ve -1 etiketli vektöre ait düzlem yaratarak, gözlemleri bu iki düzlemden birisine atamaktır. Düzleme en yakın eğitim verileri, destek vektör olarak adlandırılır, [16, 17].



Şekil2. Hiper düzlem ve destek vektörleri

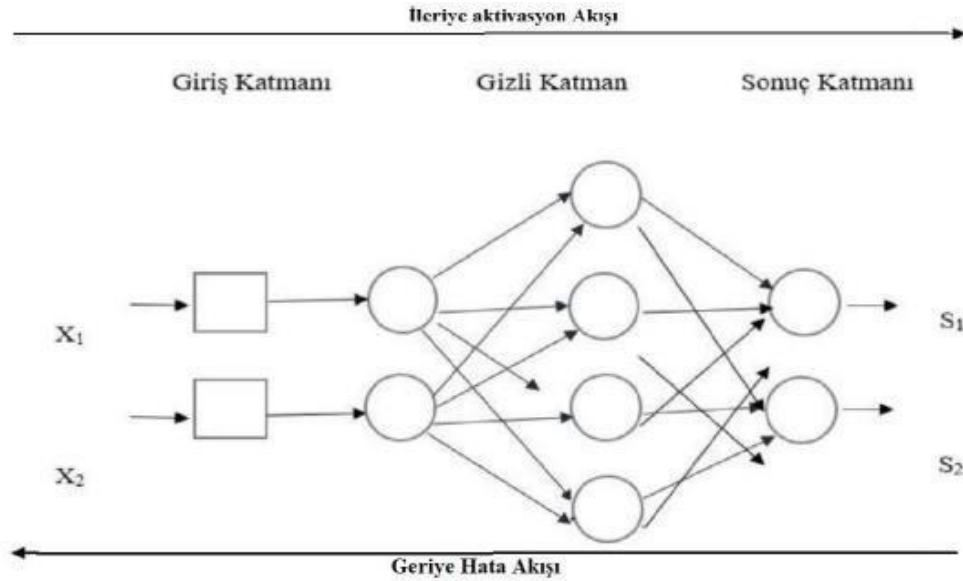
#### 2.2.4. Yapay Sinir Ağları (YSA)

İnsan beynindeki sinir hücrelerinin yapısını temel alan bir işleyişe sahip olan yapay sinir ağları (YSA) güçlü bir makine öğrenme yöntemidir. Sinir hücrelerinin oluşturduğu katmanlara benzeyen bir algoritma ile neden-sonuç ilişkisini açıklamayı amaçlar. Şekil 3’de de belirtildiği üzere YSA, genel olarak 5 temel unsurdan oluşmaktadır. Bunlar girdiler, ağırlık değerleri, toplam ve aktivasyon fonksiyonu ve çıktılarıdır.



Şekil3. Yapay Sinir Ağlarının Bileşenleri

Burada,  $W_i$  ağırlıklar,  $X_i$ ’ler girdiler,  $t_j$ ’ler sınır değerleri,  $f(x)$  aktivasyon fonksiyonunu göstermektedir [18]. Genel olarak bir YSA mimarisi Şekil 4’te verilmiştir.



Şekil4. YSA Mimarisi

YSA’nın çok katmanlı olması ve bu katmanlar arasında bağlantılar bulunması, ileri ve geri beslemeli olarak öğrenebilen bir sistem olması en önemli özelliğidir. Öğrenme işlevinde tahmin hatasını minimize etmek amacıyla ağırlık değerlerinden yararlanır.

### 2.2.5. Naive Bayes (NB)

Naive Bayes algoritması, belirli bir veri kümesindeki değerlerin frekansını ve kombinasyonlarını sayarak bir olasılık kümesini hesaplayan basit bir olasılık sınıflandırıcısıdır. Algoritma Bayes teoremini kullanır ve sınıf değişkeninin değeri dikkate alınarak tüm değişkenlerin bağımsız olduğunu varsayar. Bu koşullu bağımsızlık varsayımı gerçek dünya uygulamalarında nadiren geçerlidir, dolayısıyla Naive olarak nitelendirilir, ancak algoritma çeşitli kontrollü sınıflandırma problemlerinde hızlı bir şekilde öğrenme eğilimindedir. [19].

## 3. DENEY ANALİZİ VE SONUÇLAR

Karmaşıklık Matrisi (Confusion Matrix): bir sınıflandırıcının performansını ölçmek için kullanılan bir ölçüttür. Sınıflandırma yönteminin farklı sınıftaki bireyleri hangi ölçüde sınıflandırdığı gösteren bir analiz aracıdır. İki düzeyli bir sınıflandırma probleminde 2x2'lik bir matris şeklindedir.

True Positive (TP): Sınıflandırıcı tarafından doğru olarak tahmin edilmiş pozitif sınıf

True Negative (TN): Sınıflandırıcı tarafından doğru olarak tahmin edilmiş negatif sınıf

False Negative (FN): Sınıflandırıcı tarafından yanlış olarak tahmin edilmiş negatif sınıf

False Positive (FP): Sınıflandırıcı tarafından yanlış olarak tahmin edilmiş pozitif sınıf

Karmaşıklık matrisi yardımıyla çeşitli parametreler hesaplanmıştır. Bu parametreler arasında doğruluk (accuracy), hassasiyet (precision), duyarlılık (recall), özgünlük (specificity), F-ölçümü (F1- score), kappa katsayısı ve ROC eğrisi altındaki alan (AUC-ROC) gibi performans metrikleri bulunmaktadır. Performans ölçütleri kullanılırken veri setindeki etiketli durumların dengeli olup olmadığı gözönünde bulundurulmalıdır. İki sınıf sınıflandırma için bu durum her iki sınıftaki gözlemlerin yaklaşık eşit olması anlamına gelmektedir [20].

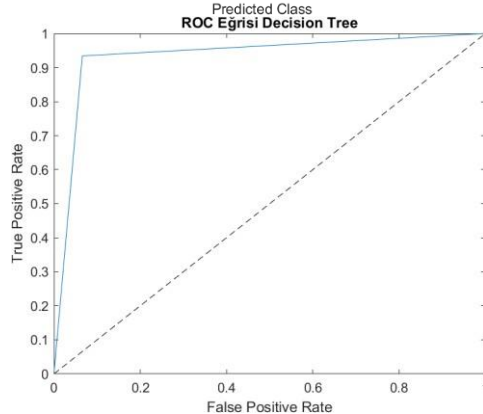
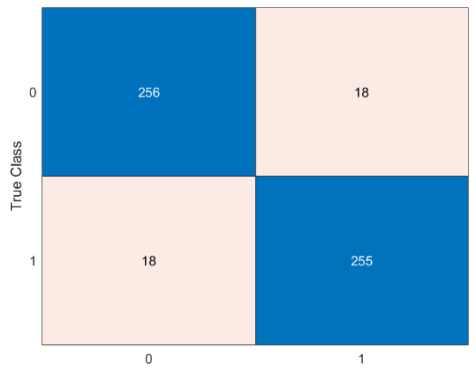
Veri setini bölerken 0.2 ve 0.3 yüzdeleri kullanılmıştır. K-NN algoritmasında ise tek tek 1 – 3 – 5 – 10 komşulukları denenerek en optimize sonucu veren yöntem incelenmiştir. Deneyler sonucu oluşan değerler aşağıdaki tabloda yer almaktadır.

Yöntem	Holdout	Accuracy	Kappa değeri	Sensitivity	Specificity	Precision	Recall	F-score	AUC
DT	0.2	0.93	0.87	0.92	0.96	0.96	0.92	0.94	0.94
DT	0.3	0.93	0.86	0.93	0.93	0.93	0.93	0.93	0.93
KNN – 10	0.2	0.89	0.78	0.89	0.88	0.89	0.89	0.89	0.88
KNN - 5	0.2	0.84	0.69	0.81	0.88	0.87	0.81	0.84	0.84
KNN - 3	0.2	0.82	0.64	0.79	0.84	0.84	0.79	0.82	0.82
KNN - 1	0.2	0.76	0.52	0.72	0.79	0.79	0.72	0.75	0.76
KNN - 10	0.3	0.87	0.74	0.90	0.84	0.85	0.90	0.87	0.87
KNN - 5	0.3	0.85	0.71	0.84	0.86	0.86	0.84	0.85	0.85
KNN - 3	0.3	0.82	0.64	0.81	0.83	0.83	0.81	0.82	0.82
KNN – 1	0.3	0.75	0.50	0.71	0.78	0.76	0.71	0.74	0.75
NB	0.2	0.83	0.67	0.77	0.90	0.89	0.77	0.83	0.84
NB	0.2	0.84	0.69	0.78	0.90	0.89	0.78	0.83	0.84
SVM	0.2	0.90	0.81	0.87	0.94	0.94	0.87	0.90	0.90
SVM	0.3	0.89	0.79	0.88	0.91	0.91	0.88	0.89	0.89
YSA	0.2	0.93	0.87	0.90	0.96	0.96	0.90	0.93	0.93
YSA	0.3	0.95	0.90	0.95	0.94	0.94	0.95	0.95	0.95

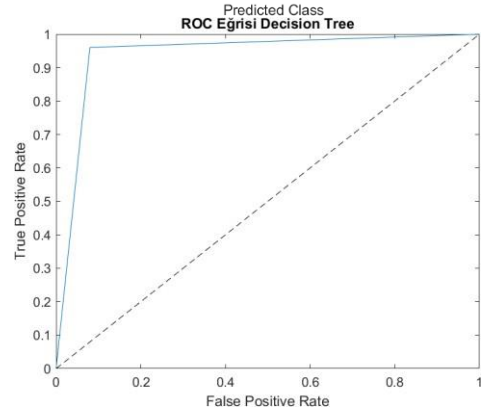
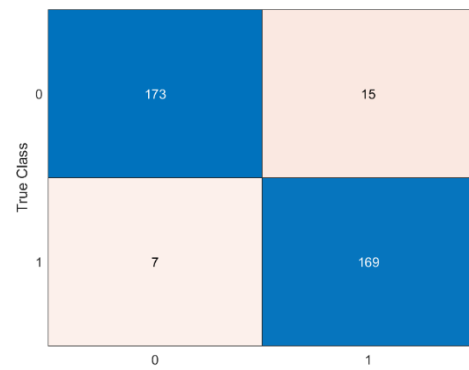
**Tablo1.** Algoritmalarından elde edilen sonuçları

Algoritmalar sonucu elde edilen karmaşıklık matrisleri ve ROC eğrisi grafikleri de aşağıda yer almaktadır.

**DT 0.3 HOLDOUT**

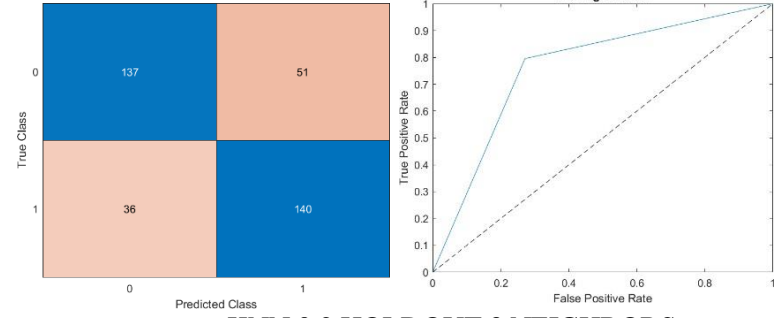


**DT 0.2 HOLDOUT**

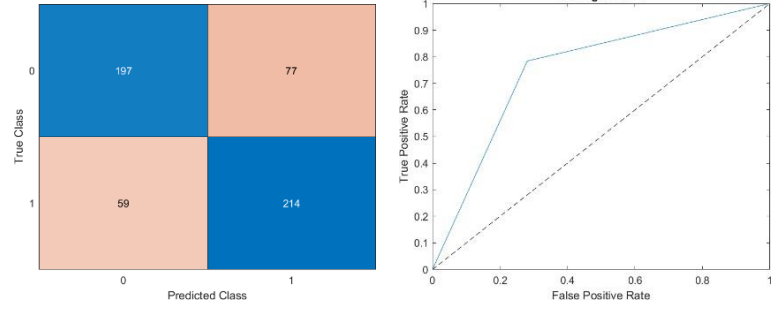




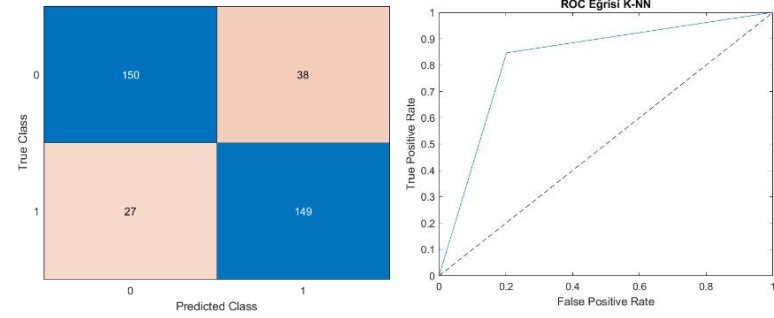
KNN 0.2 HOLDOUT 1 NEIGHBORS



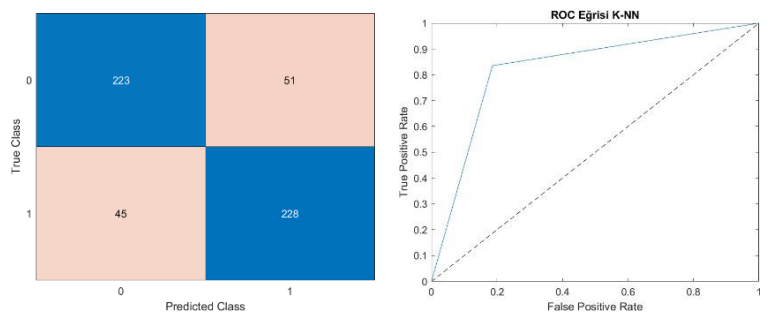
KNN 0.3 HOLDOUT 1 NEIGHBORS



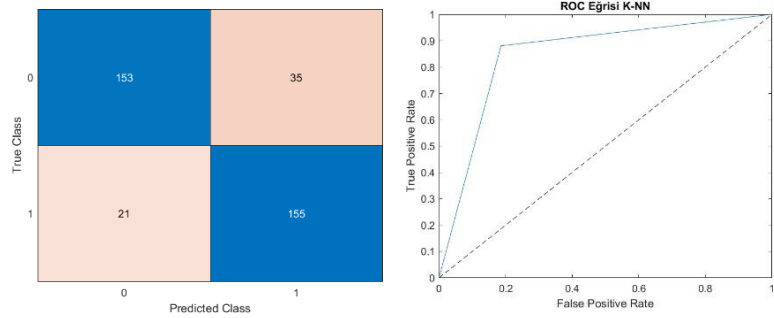
KNN 0.2 HOLDOUT 3 NEIGHBORS



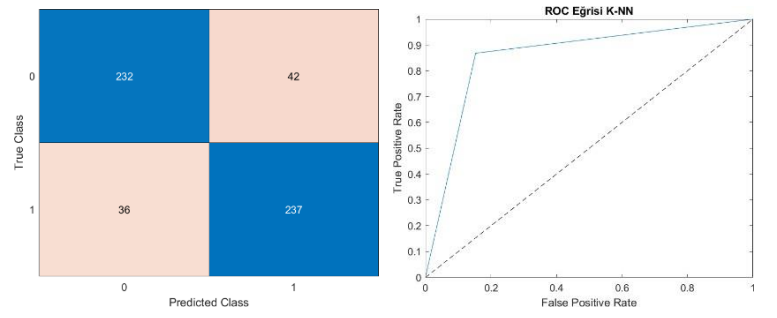
KNN 0.3 HOLDOUT 3 NEIGHBORS



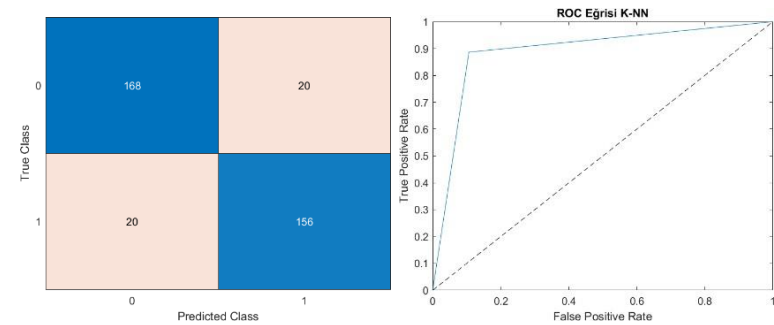
KNN 0.2 HOLDOUT 5 NEIGHBORS



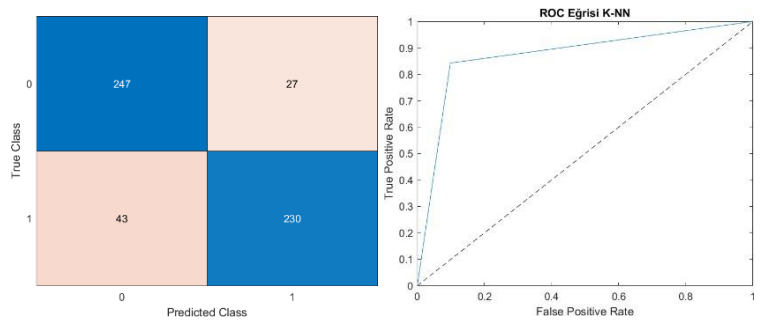
KNN 0.3 HOLDOUT 5 NEIGHBORS



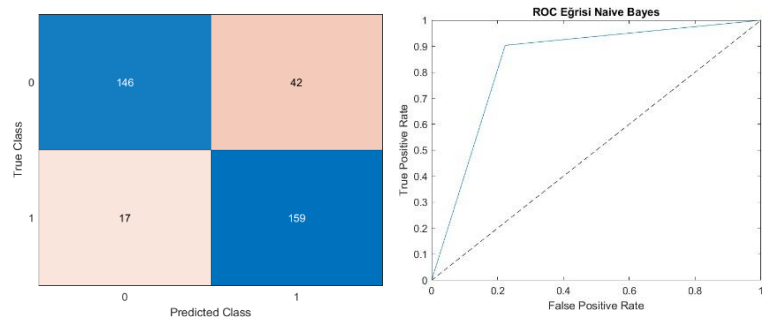
KNN 0.2 HOLDOUT 10 NEIGHBORS



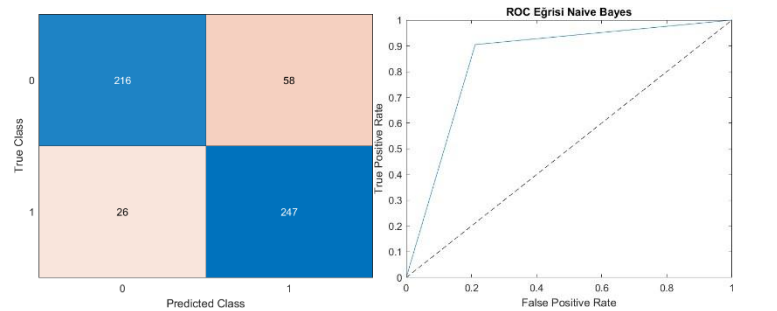
KNN 0.3 HOLDOUT 10 NEIGHBORS



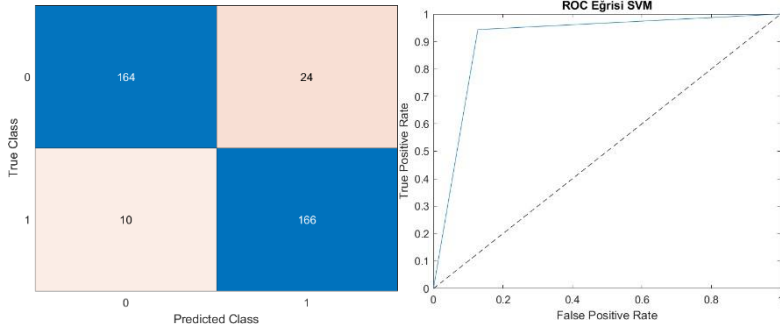
NAIVE BAYES 0.2 HOLDOUT



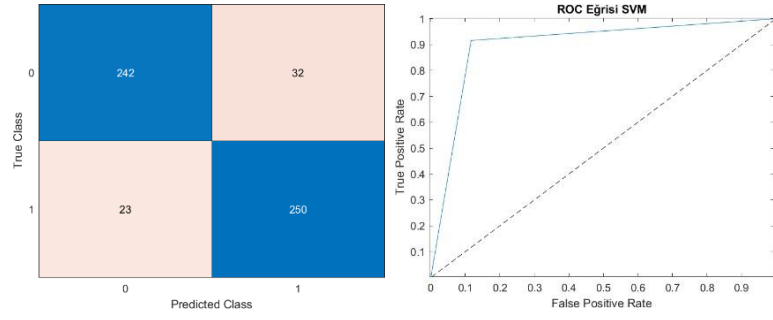
NAIVE BAYES 0.3 HOLDOUT



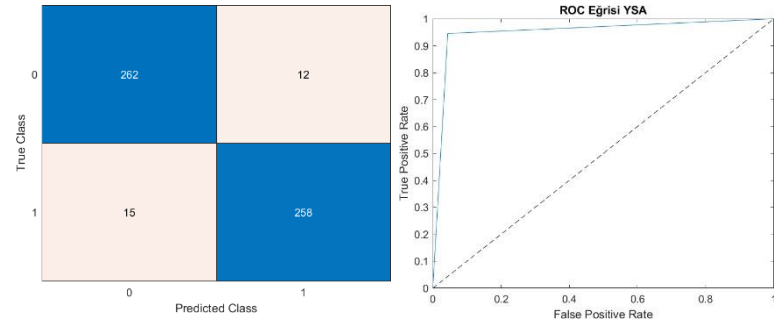
SVM 0.2 HOLDOUT



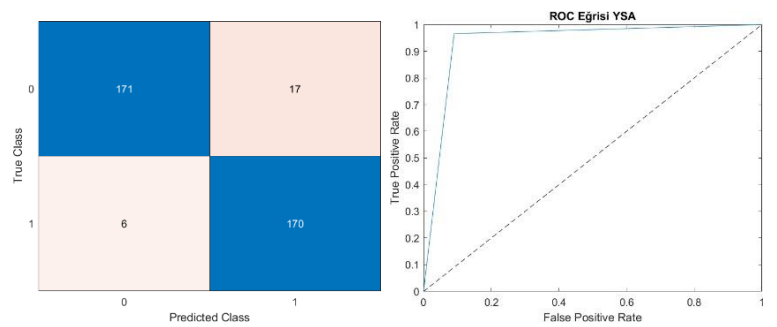
SVM 0.3 HOLDOUT



YSA 0.2 HOLDOUT



YSA 0.3 HOLDOUT



#### 4. TARTIŞMA VE SONUÇ

Bu çalışma, veri setine uygulanan beş farklı sınıflandırma yönteminin doğruluklarını kıyaslamaktadır. Belirli parametreler göz önünde bulundurularak uygulanan yöntemler sonucunda içilebilir su kalitesini belirlemede en yüksek doğruluğu veren yöntem Yapay Sinir Ağlarıdır. Yapay sinir ağları yöntemi diğer yöntemlerde de olduğu gibi hem 0.2 hem 0.3 yüzdesiyle bölünerek iki şekilde uygulanmıştır. Sonuçlar karşılaştırıldığında 0.3 yüzdeyle bölünmüş YSA'nın accuracy değeri, AUC değeri, kappa katsayısı gibi parametrelerde az bir farkla daha optimize sonuçlar verdiği görülmüştür. K-NN ile yapılan deneyler sonucunda ise en düşük doğruluk değerleri elde edilmiştir. Sonuç olarak uygulanan beş yöntem % 75 ve %95 aralığında doğruluk değerleri vermektedir.

Ayrıca veri setine ANOVA testi uygulanarak gereksiz özellikler çıkarıldığında nasıl sonuçlar vereceği test edilmiştir. Sonuç olarak, özellik çıkarılmadan önceki halinden daha iyi sonuçlar elde edilmemiş bazıları daha düşük sonuçlarla karşılaştırılmıştır.

## KAYNAKÇA

- [1] Koranga, M., Pant, P., Pant, D., Bhatt, A. K., Pant, R. P., Ram, M. & Kumar, T. (2021): DVM Model to Predict the Water Quality Based on Physicochemical Parameters. *International Journal of Mathematical, Engineering and Management Sciences*, 6(2): 645- 659.
- [2] Asadollah, S.B.H.S., Sharafati, A., Motta, D., Yaseen, Z.M. (2021): River water quality index prediction and uncertainty analysis: A comparative study of machine learning models, *Journal of Environmental Chemical Engineering*, 9, 1-14
- [3] T.H.H. Aldhyani, M. Al-Yaari, H. Alkahtani, M. Maashi, Water Quality Prediction Using Artificial Intelligence Algorithms, *Appl. Bionics Biomech.* 2020 (2020) 6659314. doi:10.1155/2020/6659314.
- [4] M. Allaire, H. Wu, U. Lall, National trends in drinking water quality violations, *Proc. Natl. Acad. Sci.* 115 (2018) 2078– 2083. doi:10.1073/pnas.1719805115.
- [5] Eren, B. & Çelebi, A. (2018): Sakarya Nehri Su Kalitesinin İstatistiksel Veri Değerleme Yöntemleri Kullanılarak Değerlendirilmesi, n 6th International Symposium on Innovative Technologies in Engineering and Science 09-11 November 2018 (ISITES2018 Alanya - Antalya - Turkey)
- [6] Bilali, A., Taleb, A. (2020): Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment, *Journal of the Saudi Society of Agricultural Sciences*, 19, 439-451.
- [7] Tousi, E.G., Duan, J., Gundyb, P.M., Bright, K., Gerba, C.P. (2021): Evaluation of E. coli in sediment for assessing irrigation water quality using machine learning, *Science of the Total Environment*, 799, 1-13.
- [8] N. Radhakrishnan, A.S. Pillai, Comparison of Water Quality Classification Models using Machine Learning, in: 2020 5th Int. Conf. Commun. Electron. Syst., 2020: pp. 1183–1188. doi:10.1109/ICCES48766.2020.9137903.
- [9] M.S. Islam Khan, N. Islam, J. Uddin, S. Islam, M.K. Nasir, Water quality prediction and classification based on principal component regression and gradient boosting classifier approach, *J. King Saud Univ. - Comput. Inf. Sci.* (2021).
- [10] D. Venkata Vara Prasad, L. Y Venkataramana, P.S. Kumar, G. Prasannamedha, K. Soumya, P. A.J., Water quality analysis in a lake using deep learning methodology: prediction and validation, *Int. J. Environ. Anal. Chem.* (2020) 1–16.
- [11] M. Jawthari, V. Stoffová, Predicting students' academic performance using a modified kNN algorithm, *Pollack Period.* 16 (2021) 20–26.
- [12] M.E. Sertkaya, B. Ergen, M. Togacar, Diagnosis of Eye Retinal Diseases Based on Convolutional Neural Networks Using Optical Coherence Images, in: 2019 23rd Int. Conf. Electron., 2019: pp. 1–5.
- [13] A. Topîrceanu, G. Grossec, Decision tree learning used for the classification of student archetypes in online courses, *Procedia Comput. Sci.* 112 (2017) 51–60.

- [14] H. Polat, M. Turkoglu, O. Polat, Deep network approach with stacked sparse autoencoders in detection of DDoS attacks on SDN-based VANET, IET Commun. 14 (2020) 4089–4100.
- [15] Demirci, D. A. (2007). “Vektör makineleri ile karakter tanıma” ( Yüksek lisans tezi, Bilgisayar Mühendisliği Bölümü, Yıldız Teknik Üniversitesi, İstanbul, Türkiye).
- [16] Sedat METLEK , Kıyas KAYAALP “Derin Öğrenme ve Destek Vektör Makineleri ile Görüntüden Cinsiyet Tahmini” Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 8 (2020) 2208-2228
- [17] Koranga, M., Pant, P., Pant, D., Bhatt, A.K., Pant, R.P., Ram, M., Kumar, T. (2021): SVM Model to Predict the Water Quality Based on Physicochemical Parameters, International Journal of Mathematical, Engineering and Management Sciences, Vol. 6, No. 2, 645-659.
- [18] Haykin, S. (1999): Neural Networks: A Comprehensive Foundation. Prentice Hall, USA.
- [19] Mücahid Mustafa Saritas, Ali Yasar (2019) Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification, International Journal of Intelligent Systems and Applications in Engineering
- [20] Uğuz, S. (2021): Makine Öğrenmesi Teorik Yönleri ve Python Uygulamaları ile Bir Yapay Zekâ Ekolü, Nobel Yayıncılık, 2. Baskı.