



# Düzce Üniversitesi Bilim ve Teknoloji Dergisi

*Araştırma Makalesi*

## Tarım Arazilerinde Harcanan Su Miktarını Yapay Zekâ Teknikleri Kullanarak Belirlenmesi<sup>1</sup>

 Hilmi Cenk BAYRAKÇI<sup>a\*</sup>,  Rabia Sultan ÇİÇEKDEMİR<sup>b</sup>,  Merdan ÖZKAHRAMAN<sup>a</sup>

<sup>a</sup> Mekatronik Mühendisliği Bölümü, Teknoloji Fakültesi, Isparta Uygulamalı Bilimler Üniversitesi, Isparta, TÜRKİYE

<sup>b</sup> Mekatronik Mühendisliği Bölümü, Lisansüstü Eğitim Enstitüsü, Isparta Uygulamalı Bilimler Üniversitesi, Isparta, TÜRKİYE

\* Sorumlu yazarın e-posta adresi: cenkbayrakci@isparta.edu.tr

DOI:10.29130/dubited.1015690

### ÖZ

Ülkemizde ve dünyada en önemli problemlerden birisi de su problemidir. Su tüketiminin en çok olduğu alanlar arasında tarım sektörü gelmektedir. Tarımda sulamada kullanılan önemli yöntemlerden birisi de damla sulama yöntemidir. Damlama sulama sisteminde bitkilerin köküne su vererek büyümesi sağlanarak daha az su tüketilmesi sağlanmaktadır. Ayrıca önemli noktalardan birisi de her bitkinin su ihtiyacının farklı olmasıdır. Çalışmada örnek bölgelerden ve açık erişimli internet sitelerinden toplanan İl, ilçe, bitkinin türü gibi veriler ışığında tespit edilen su harcama miktarına göre veri seti oluşturulmuştur. Elde edilen veri seti üzerinde farklı yapay zekâ algoritmaları kullanılarak harcanan su miktarı tahmin edilmeye çalışılmıştır. En doğru sonucu veren yapay zekâ modeli için ara yüz programı hazırlanarak kullanıcının ara yüz programından girmiş olduğu değerlere göre en doğru su harcama miktarını veren ve sonucu ekranda görüntüleyen bir sistem oluşturulmuştur.

**Anahtar kelimeler:** Yapay zekâ, Makine öğrenmesi, Tarım

## Determining the Amount of Water Used in Agricultural Lands Using Artificial Intelligence Techniques

### ABSTRACT

One of the most important problems in our country and in the world is the water problem. The agricultural sector is among the areas with the highest water consumption. One of the important methods used in irrigation in agriculture is the drip irrigation method. In the drip irrigation system, less water is consumed by giving water to the roots of the plants and allowing them to grow. In addition, one of the important points is that the water needs of each plant are different. In the study, a data set was created according to the amount of water consumption determined in the light of data such as province, district, plant species collected from sample regions and open access websites. The amount of water consumed was tried to be estimated by using different artificial intelligence algorithms on the obtained data set. An interface program has been prepared for the artificial intelligence model that gives the most accurate result, and a system has been created that gives the most accurate water consumption amount according to the values entered by the user from the interface program and displays the result on the screen.

**Keywords:** Artificial intelligence, Machine learning, Agriculture

<sup>1</sup>ICAAME 2021 konferansında sunulmuş, özet metin Kabul Edilen Özetler E-Kitabı ss.23'te verilmiştir.

Geliş: 01/11/2021, Düzeltme: 22/12/2021, Kabul: 25/12/2021

# I. GİRİŞ

Tarım, bitkisel ürünlerin ve bu bitkileri tüketen canlılardan elde edilen ürünlerin üretilmesi, üretilen ürünlerin niteliklerinin üst düzeye getirilmesi, uygun koşullarda depolanması ve işlenmesiyle ilgilenen bir bilim dalıdır. Bitkisel ve hayvansal üretimler tarım biliminin en önemli ve en başta gelen kollarıdır ve bu sektör, insanların yaşamaları için en temel gereksinimlerini karşılamaktadır [1]. Dünyada tarıma gösterilen ilginin nedeni ise iklim değişiklikleri, kullanılabilir su kaynaklarının büyük ölçüde azalması ve tarıma elverişli ortamların zamanla yok olmasıdır. Gelecekteki tahminlerin bazıları gıda ve su savaşlarının olması yönündedir ve bu tahminler bile konunun ne kadar önemli olduğunu anlatmaktadır [2].

Tarımda bitkinin gereksinimi olan ve havanın neme doymasıyla yeryüzüne düşen damlacıkların yeterli olmamasından dolayı bitkinin kök bölgesine yeterli miktarda ve zamanda su verilmesine sulama denmektedir [3]. Tüketilen su miktarı nüfus oranının artışı ile doğru orantılı olarak artmıştır. Dünya nüfusu 1700'lü yıllarda 700 milyon dolayında iken, su tüketimi yaklaşık 110 m<sup>3</sup> olup, bunun da yaklaşık % 90'ının tarımda sulama amaçlı kullanıldığı, 1990'lı yıllara gelindiğinde ise tüketilen su miktarının 40 kat arttığı belirtilmektedir. Dünyada bulunan su kaynaklarının değişmemiş olması tarım alanında kısıtlı koşullarda sulamanın yapılmasını zorunlu kılmıştır [4]. Son senelerde hızla gelişme gösteren sanayi sektörünün yoğun su gereksinimi ile tarımsal sulama için ihtiyaç duyulan miktarın sağlanabilmesi, Ülke nüfusunun artmasından dolayı kaliteli içme ve kullanma suyuna olan gereksinimin karşılanabilmesi için ülkemizin su kaynaklarının daha doğru ve verimli kullanılması oldukça fazla önem taşımaktadır. Bitkisel üretim bakımından ülkemizin birçok yerinde yağış olayının yetersiz olması, kurak ve yarı kurak olan bölge sayımızın çok olması dolayısıyla kaliteli ürün ve yüksek oranda verim elde etmek için sulama yapılması gerekmektedir [3]. Ülkemizin ekonomik olarak tüketilebilir su potansiyeli 110 km<sup>3</sup> tür ve bunun 40 km<sup>3</sup>'ü tüketilmektedir. Bu miktarın %75'i sulamada, %15'i içme ve kullanma sektöründe %10'u ise endüstride kullanılmaktadır. Bizim ve gelecek nesillerin yaşamını sürdürebilmesi için toprak ve su gibi doğal kaynakların tükenmemesi için tarımsal üretimde suyu bilinçli bir şekilde kullanmalıyız [5]. Doğru su ve toprak yönetimi, tüm bu kullanımların sonucunda ortaya çıkan kaliteli su ihtiyacını karşılayabilmektedir [3].

Günümüzde ve yakın gelecekte tarımsal araştırma konularının başında yapay zekâ gelecektir bunun nedeni ise yapay zekâ uygulamalarının tarımsal işlemlerin kolaylaştırılması ve çözüm veya iyileştirme bekleyen sorunlara alternatif çözümler geliştirme potansiyeli olmasıdır. Tarımda birçok alanda yapay zekâ teknikleri kullanılarak; bitkisel üretim planlamaları, bitkilerin sınıflandırılması, verim tahmini, bitki hastalık, zararlı ve yabancı otlarının tespiti, tarım robotlarında rota belirlenmesi ve uygulama kararlarının alınması, serada uygun çevre şartlarının belirlenmesi, işletme kararlarının alınması, sulama yönetimi, ürün rotasyonunun belirlenmesi, en uygun gübre ve alet makine seçimi, hayvan hastalıklarının tespiti, uygun yem rasyonlarının hazırlanması, hayvan davranışlarının belirlenmesi gibi konularda araştırmacılar tarafından çok sayıda çalışma yapılmıştır [6]. Kompleks ve çok olan verilerin daha hızlı, daha iyi yorumlanması ve çözümlenmesi yapay zekâ yöntemi ile sağlanır [7].

Yapay zekâ, zeki makinelerin zeki bilgisayar programları yapma bilimi ve mühendisliği olarak John McCarthy tarafından tanımlanmıştır. Silage'ye göre ise sezgisel programlama temelinde olan yaklaşım olarak görülürken Popov'a göre ise insanların yapmış oldukları çalışmaları bilgisayarlara yaptırabilme olarak tanımlanmaktadır [8]. Axe'nin görüşü ise yapay zekâyı akıllı programları hedefleyen birim şeklindedir [9].

Akıllı uygulama yazılımlarında, yapay zekâ, makine öğrenmesi, derin öğrenme gibi kavramlar çoğunlukla karıştırılır ve birbiri yerine kullanılır. Oysaki aralarında ufak farklılıklar vardır. Makine öğrenmesi yapay zekânın bir alt kümesi olarak yer alırken derin öğrenme ise makine öğrenmesinin bir alt kümesi olarak yer alır. Yani tüm makine öğrenmesi uygulamaları yapay zekâ olarak sayılır [8].

Makine öğrenmesi, bilgisayarın yani makinenin önceden edindiği bilgileri öğrenerek hafızasında tutması ve gelecekte oluşabilecek, öğrenmiş olduğu bilgiye benzeyen olaylar hakkında karar

verebilmesini ve oluşacak herhangi bir problemde çözüm üretebilmesidir [10]. Aynı zamanda makine öğrenmesi, çalışılmakta olan sorunun çözümünü kendi kendine öğrenebilen, karmaşık görüntüleri algılama ve veriye dayalı karar verebilme özelliği olan bilgisayar algoritmalarıdır [11]. Makine öğrenmesi geçmişteki verileri baz alarak yeni veri için en uygun modeli bulmaya çalışır. Çok fazla verinin elle girilmesi çok zahmetlidir ve oldukça zaman alır. Oluşturulan veya hazır bir şekilde açık erişimli internet sitelerinde kullanıma sunulan veri setleri ile makine öğrenmesi yöntemleriyle oluşturulan model, en yüksek performansı elde etmek için kullanılmaktadır. Günümüzde makine öğrenmesinin popüler olması ve önem kazanmasının nedeni ise uygulama alanı ne olursa olsun, çok miktardaki verinin analiz edilerek gelecek ile ilgili tahminlerde bulunması ve bizim karar vermemize yardımcı olmasıdır [10].

## **II. MATERİYAL VE METOT**

Yapılan çalışmada T.C. Tarım ve Orman Bakanlığının açık erişimli internet sitesinde yayınlamış olduğu TÜRKİYE'DE SULANAN BİTKİLERİN BİTKİ SU TÜKETİMLERİ isimli çalışmasından bazı veriler alınarak düzenlenmiş ve veri seti oluşturulmuştur. Veri setinde toplam 3203 adet veri kullanılmıştır.

Veri seti, Doğrusal (Linear) Regresyon, Ridge Regresyon, Lasso Regresyon, Gradyan Artırma (Gradient Boosting) Regresyon, XGboost Regresyon yöntemleri ile eğitilmiş ardından Python ile bir ara yüz oluşturulmuştur. Çalışmada kullanılan veri seti, Doğrusal (Linear) Regresyon, Ridge Regresyon, Lasso Regresyon, Gradyan Artırma (Gradient Boosting) Regresyon, XGboost Regresyon yöntemleri ve oluşturulan arayüz aşağıda detaylı bir şekilde verilmiştir.

### **A. MATERİYAL**

#### **A. 1. Veri seti**

T.C. Tarım ve Orman Bakanlığının araştırma sonuçlarından İl-ilçe, Tarih, Bitkiler ve Yağış verileri alınarak 3203 adet veri oluşturulmuş ve çalışmada kullanılmıştır. Tablo 1' de kullanılan veri setinin bir örnekleme görülmektedir. Oluşturulan veri setinde Tarih, İl, Bitki, Etc, Yağış ve Gölge sınıfı mevcuttur.

*Tablo 1. Veri seti*

	<b>Tarih</b>	<b>İl</b>	<b>Bitki</b>	<b>Etc</b>	<b>Yağış</b>	<b>Gölge</b>
<b>0</b>	OcakI	Adana	Ayçiçeği	0.0	41.9	80
<b>1</b>	OcakII	Adana	Ayçiçeği	0.0	23.3	80
<b>2</b>	OcakIII	Adana	Ayçiçeği	0.0	39.9	80
<b>3</b>	ŞubatI	Adana	Ayçiçeği	0.0	31.4	80
<b>4</b>	ŞubatII	Adana	Ayçiçeği	0.0	33.9	80
...	...	...	...	...	...	...
<b>3201</b>	AralıkI	Karaisalı	Karpuz	0.0	45.9	80
<b>3202</b>	AralıkII	Karaisalı	Karpuz	0.0	58.6	80
<b>3203</b>	AralıkII	Karaisalı	Karpuz	0.0	52.7	80

$$dt = \frac{[Etc * (\frac{Ps}{85}) - Ps]}{0.8575} \quad (1)$$

Denklem 1 'de ETc bitki su tüketimini (mm), Ps bitki tarafından gölgelenen alan yüzdesini (%), Pe etkili yağışı (mm), dt toplam sulama suyu ihtiyacını (mm) ifade etmektedir.

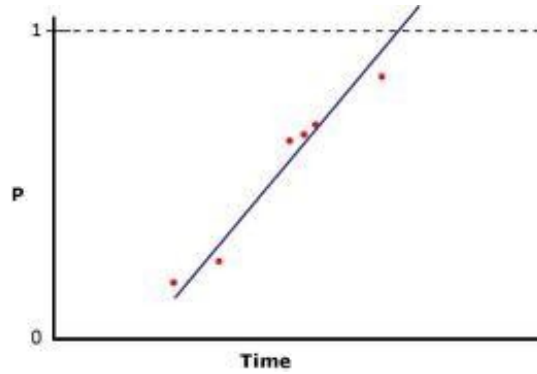
Bu bilgileri denklem 1'de yerine yazarak toplam sulama suyu ihtiyacını yani çıkış değerimizi hesaplattık ve tahmin etme işleminde kullandık. Tarih bilgisi ise 10'ar günlük periyotlar halinde her ay için verilmiştir. Çalışmada kullanılan veriler çok fazla olmadığından dolayı, random olarak eğitim ve test verisi oluşturan sklearn kütüphanesi içerisinde yer alan train\_test\_split komutu ile %75 eğitim, %25 test veri kümesi olarak ayrılır.

## B. METOT

Gerçekleştirilen çalışmanın ilk aşamasında T.C. Tarım ve Orman Bakanlığının internet sitesinde yayınlamış olduğu TÜRKİYE'DE SULANAN BİTKİLERİN BİTKİ SU TÜKETİMLERİ isimli çalışmasından il-ilçe, tarih, bitkiler ve yağış verileri alınarak veri seti oluşturulmuştur. Bu veri seti kullanılarak yapay zekânın bir alt dalı olan makine öğrenmesi ile eğitim gerçekleştirilmiştir. Gerçekleştirilen eğitimde makine öğrenmesinde kullanılan Doğrusal (Linear) Regresyon, Ridge Regresyon, Lasso Regresyon, Gradyan Artırma (Gradient Boosting) Regresyon, XGboost Regresyon algoritmaları kullanılmıştır. Python ile bir arayüz oluşturulmuştur.

### B. 1. Doğrusal (Linear) Regresyon

Şekil 1'de de görüldüğü gibi Doğrusal (Linear) Regresyon analizi basit doğrusal regresyon ve çoklu doğrusal regresyon olarak ikiye ayrılmaktadır. Doğrusal regresyon analizi, yanıt değişkeni ile tek bir açıklayıcı değişken arasındaki fonksiyonel ilişkinin doğrusallığıdır. Kullanılan değişkenin sayısına göre, basit doğrusal regresyon ve çoklu doğrusal regresyon olarak iki kategoriye ayrılır. Eğer bir yanıt değişkeni ve birden çok açıklayıcı değişken arasındaki ilişki tanımlanmak istenirse, ilişki çoklu doğrusal regresyon analizi ile incelenir [12].



Şekil 1. Doğrusal (Linear) Regresyon Grafiği

### **B.1.1. Basit Doğrusal Regresyon Modeli:**

Normal dağılım gösteren bilinen bir değişkenden, aralarında ilişki olan başka bir normal dağılım gösteren sayısal değişken tahmini basit doğrusal regresyon (simple linear regression) olarak tanımlanır. Bağımlı değişken (y) ile bağımsız değişken(ler) (x,..) arasındaki ilişkiyi inceler [13].

$$Y = a + bX + e \quad (2)$$

Denklem 2’de Y çıkış değerini yani bağımlı değişkeni, a sabit katsayısı, b regresyon katsayısını, X bağımsız değişkeni, e ise hata terimini ifade etmektedir.

Denklem 2’ de “a” olarak gösterilen değer, doğrunun y eksenini kestiği nokta analitik yöntemde sabit hata ölçüsüdür. “b” olarak gösterilen değer regresyon katsayısı (eğim) ise analitik yöntemde oransal hata ölçüsüdür. Bağımlı değişkende gerçekleşen ortalama değişim miktarı, bağımsız değişkende ki bir birimlik değişime yani artma veya azalmaya bağlıdır. Bir regresyon analizinin basit regresyon olarak tanımlanabilmesi için her iki değişkenin dağılımı da normal olmalıdır. İlişki katsayısı gibi regresyon katsayısı da pozitif veya negatif olabilir. İlişki katsayısı -1 ile +1 arasında değişirken, regresyon katsayısı her değeri alabilir. Regresyon katsayısı değişkenin ölçüldüğü birimden etkilenir [13].

### **B.1.2. Çoklu Doğrusal Regresyon Modeli:**

Açıklayıcı değişkenler olan bağımsız değişkenler ile cevap değişkeni olan bağımlı değişkendeki toplam değişim çoklu doğrusal regresyon modeli olarak tanımlanır. Bu regresyon analizinde, bağımsız değişkenler model oluşturur ve bu bağımsız değişkenlerden bazılarının modele katkısı çok önemsiz olabilir. Bu nedenle, bağımlı değişkeni "en uygun" şekilde açıklayacak bağımsız değişkenlerin belirlenmesi ve önemsiz değişkenlerin modelden çıkarılması gerekir. Bu sürece "değişken seçimi" denir [14].

$$Y = a_0 + b_1X_1 + b_2X_2 + b_3X_3 + e \quad (3)$$

Denklem 3’te Y çıkış değerini yani bağımlı değişkeni, a sabit katsayısı, b regresyon katsayısını, X bağımsız değişkeni, e ise hata terimini ifade etmektedir.

Formül (3)’te b değeri sabit değer ve modelde yer alan değişkenlere ait katsayıları gösterirken, e ise tesadüfi hata (rezidüel) göstermektedir. Rezidüel, örneklemdaki bireylere ait gözlenen değerle ile bu bireyler için modelden belirlenen değerler arasındaki farktır [13].

## **B. 2. Ridge Regresyon**

Ridge regresyon yöntemi Hoerl ve Kennard tarafından 1970 yılında geliştirilmiştir. Technometrics dergisinde yayınladıkları ilk makaleleri “Ridge Regresyon: Ortogonal Olmayan Problemler İçin Sapmalı Tahmin” makalesidir. Ridge regresyonun yöntemi, basit en küçük kareler yöntemine benzer bir yöntemdir. Ridge regresyon yönteminde,  $(X'X)$  matrisinin köşegen elemanlarına küçük ve pozitif bir sabitin eklenmesiyle standart formdaki değişken oluşturulur. Buna göre ridge regresyon çözümü denklem 4 şeklindedir.

$$(X'X) - 1 X'Y \quad (4)$$

Denklem 4’ te verilen  $(X'X)$  matrisinin köşegen elemanlarına küçük ve pozitif sabitlerin eklenmesindeki amaç, matris şartlı sayısının önemli oranda küçültülmesidir [15]. Çoklu bağlantı olduğunda bağımsız değişkenler arasındaki ilişkinin fazla olması nedeniyle  $X'X$  matrisinin varyansları önemli bir şekilde büyüyecektir. Böylece gerçekte önemli gözükken parametre değerleri varyansın artması neticesinde önemsiz çıkacaktır [16].

Çoklu bağlantı problemlerinin giderilmesine yönelik çeşitli öneriler vardır. Bunları aşağıdaki gibi özetlemek mümkündür.

- a) Bir veya daha fazla değişkenin modelden çıkartılması,
- b) Modele yeni değişkenler ilave edilmesi,
- c) Birbirleri ile ilişkili olan değişkenlerin toplanması,
- d) Temel Bileşenler Regresyon analizi ya da Ridge Regresyon analizi kullanarak değişkenlerin standardize edilmesi çözüm olarak verilebilir [16].

### B. 3. Lasso Regresyon

En küçük Mutlak Küçülme ve Seçim Operatörü olarak adlandırılan LASSO ilk olarak 1996 yılında Tibshirani tarafından ortaya atılmıştır. Bu yöntem, bazı değişken katsayılarını küçültürken, bazı değişken katsayılarını da sıfıra indirir [17]. Fakat değişkenler arasında fazla ilişki olduğunda ilişkili olan değişkenlerden bir tanesini modele alır ve diğer ilişkili değişkenleri modelden çıkarır. Bu özelliğe de değişken seçimi denir. Bu yöntemin en büyük avantajı, çok iyi tahminler yapabmesidir. Katsayıların sıfır olması veya sıfıra doğru daralması varyansı azaltabilmektedir. Özellikle diğer yöntemlerin çözmekte yetersiz kaldığı bir durum olan örnek genişliği küçük ve bağımsız değişken sayısı fazla olduğu durumda kullanışlı bir yöntemdir [18]. Lasso yöntemi, hem katsayıları tahmin etmek hem de değişken seçimi yapmak için kullanılır. Belirli sabit bir katsayıdan daha küçük olan katsayıların mutlak değerlerinden yola çıkarak artıkların kareleri toplamını minimize eden bir değişken seçme yöntemidir [17,18].

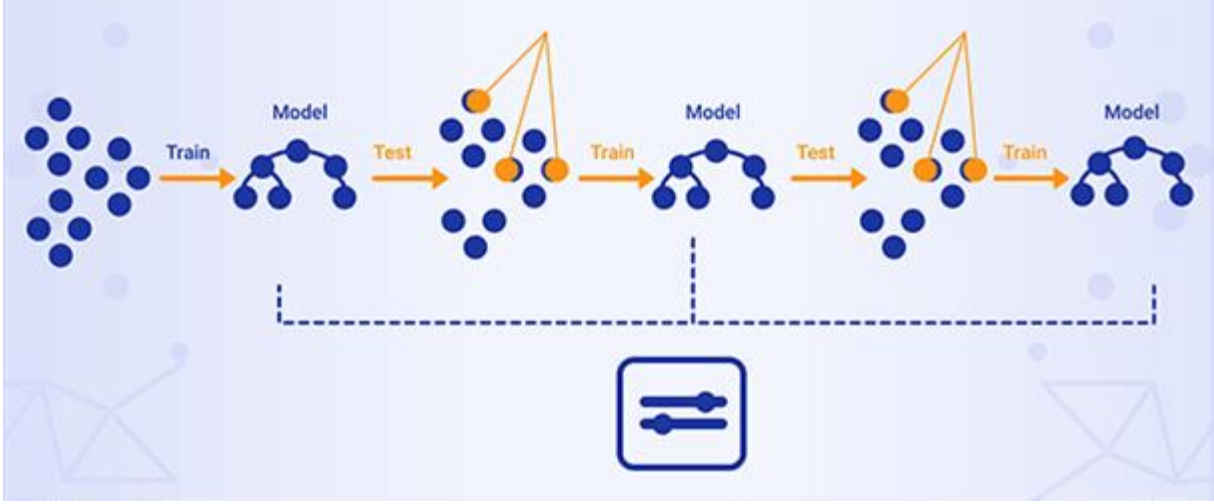
Denklem 5'te Lasso Regresyon fonksiyonu verilmiştir.

$$\beta_{Lasso} = \underset{\beta}{\operatorname{argmin}} \left( \frac{1}{2} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right) \quad (5)$$

### B. 4. Gradyan Artırma (Gradient Boosting) Regresyon

Gradyan Artırma, makine öğrenmesi algoritmasıdır ve karar ağaçlarının tahminini iyileştirmek için geliştirilmiş bir modeldir [19,20]. Friedman ve arkadaşları tarafından geliştirilmiş bir algoritmadır. Bu karar ağaçlarının en büyük dezavantajı basit ağaçlarda büyük bir "bias", karmaşık ağaçlarda ise büyük bir varyans oluşturmalarıdır. Torbalama bir veri kümesinden rasgele sayıda veri seçme yöntemidir ve ağacın varyansını azaltmak için kullanılır. Bu algoritmada diğer karar ağaçlarına benzer, zayıf halkaların birbirlerinin zayıf yönlerinden hata çıkararak daha güçlü bir halka ortaya çıkarır. Gradyan Artırma algoritmasına göre öncelikle ilk iterasyonda bir tahmin fonksiyonu oluşturulur ve bu fonksiyonlara ağaç adı verilir [19]. Bir sonraki ağaç oluşturulurken daha önce oluşturulan ağaçların hata oranları hafızada tutulur [20]. Tahminler ile gözlemler arasındaki fark hesaplanarak bu farklardan bir kayıp fonksiyonunu elde edilir. İkinci iterasyonda ise tahmin ve kayıp fonksiyonları birleştirilerek tekrar tahminler ile gözlemler arasındaki fark hesaplanır. Böylelikle sürekli üstüne ekleyerek tahmin fonksiyonunun başarısı arttırılmaya çalışılır ve hata oranının sıfıra yaklaşması sağlanır [19]. Gradyan Artırma Regresyon modeli için uygulama adımları aşağıda sıralanmıştır;

1. Regresyon ağacı oluşturulur.
2. Her bir ağacın dalı için tahminler ile gözlemler arasındaki fark yani hata oranı hesaplanır.
3. Hesaplanan hata oranları yeni gözlem verileri olarak kullanılır.
4. Yeni ağaç oluşturularak hata oranlarını en aza indirmeyi hedefler
5. Hata oranı sıfıra yaklaşıncaya kadar bu adımlar tekrarlanır.



*Şekil 2. Gradyan Yükseltme (Gradient Boosting) algoritmasının işleyiş süreci*

Şekil 2’de Gradyan Artırma algoritmasının işleyiş süreci verilmiştir.

## B. 5. XGboost Regresyon

XGBoost algoritması 2016 yılında Washington Üniversitesi’nde Tianqi Chen ve Carlos Guestrin tarafından bir araştırma projesi olarak geliştirildi [21]. XGBoost algoritması gradyan artırma temeli kullanan karar ağacı tabanlı ve karar ağacını güçlendiren makine öğrenme algoritmasıdır [21,22]. Sistem açık kaynak paketi olarak mevcut olup sistem makine öğreniminde kabul görmüş ve çok sayıda Kaggle yarışmasını kazanan algoritma olmasıyla ilgiyi üzerinde toplamıştır [21,22,24]. Ağaçları oluşturmak için CART(Classification and Regression Tree) algoritmasını kullanan XGBoost algoritması, Gradient Boosting Machines algoritmasının geliştirilmiş ve daha hızlı hale getirilmiş bir uygulamasıdır. Özellikle gradyan artırma algoritmaları için bellek ve donanım kaynaklarını verimli bir şekilde optimize etmektedir. XGBoost algoritması özellikle hesaplama hızına ve Gradient Boosting Machines algoritmasından farklı olarak düzenleme (regularization) gibi modelin performansına odaklanan pek çok avantajı vardır [23]. Avantajlarından bazıları ise; ölçeklenebilir, dağıtılmış veya paralel hesaplama ve algoritmik optimizasyon işlemlerini kullanarak milyarlarca örnek üretir ve tümü minimum kaynak kullanır. Bu nedenle, verilerin sınıflandırılması ve verilerin üst düzey ön işlenmesi gibi konularda oldukça etkilidir [22]. Bu gibi algoritmaların mantığında ise birden fazla küçük ve zayıf kişiyi bir araya getirerek daha güçlü bir kişi çıkarmaya dayanır [22,25]. Arttırma sıralı bir süreçtir; yani ağaçlar, daha önce yetiştirilmiş bir ağaçtan gelen bilgiler kullanılarak birbiri ardına büyütülür. Bu süreç verilerden yavaş yavaş öğrenir ve sonraki iterasyonlarda tahminini iyileştirmeye çalışır [25].

## B. 6. Arayüz Programı

Arayüz, bilgisayar ve kullanıcı arasında iletişimin başladığı ve kullanıcının ihtiyaç duyduğu bilgilere ulaşabildiği bir tasarımdır ve yapılan programların en önemli parçasıdır. Program ile kullanıcı arasındaki iletişim için her şeyden önce bilgisayar ekranı (arayüz) ve arayüzün nasıl tasarlandığı önem taşımaktadır [26].

Bu çalışmada tahmin etme işlemi gerçekleştirildikten sonra Şekil 3’te verilmiş olan arayüz tasarlanmıştır. Tasarlanan arayüzde kullanıcıdan, kullanılacak olan bilgilerin ComboBox içerisinden seçilmesi isteniyor ve hesaplama butonuna basıldığında tahmin edilen bitkinin su ihtiyacı Label içerisine yazdırılıyor.



*Şekil 3. Tasarlanan arayüz*

### **III. BULGULAR VE TARTIŞMA**

Çalışmada, veri setindeki Tarih, İl, Bitki, Etc, Yağış, Gölgeleme alanı giriş, bitki su tüketimi ise çıkış değeri olarak belirlenmiştir.

Doğrusal (Linear) Regresyon, Ridge Regresyon, Lasso Regresyon, Gradyan Artırma (Gradient Boosting) Regresyon, XGboost Regresyon algoritmaları ile tahmin etme işlemi gerçekleştirilmiştir. Bulunan değerler aşağıda verilmiştir.

#### **C. 1. Doğrusal (Linear) Regresyon Algoritması Bulguları**

Doğrusal (Linear) regresyon algoritmasında Tablo 2’de verilen değerler kullanılarak model eğitilmiştir.

*Tablo 2. Doğrusal (Linear) regresyon algoritmasında kullanılan değerler*

<b>Doğrusal Regresyon katsayıları (coefficients)</b>						
<b>Kullanılan Değerler</b>	0.00408814	-0.00158106	0.00304061	1.07188616	-0.98434975	0.30402362

Tablo 2’de ki değerler ile model eğitildiğinde model doğruluk oranı:

Linear Doğruluk : 0.9422738346364411 olarak bulunmuştur.

#### **C. 2. Ridge Regresyon Algoritması Bulguları**

Ridge regresyon algoritmasında Tablo 3 ve Tablo 4’te verilen değerler kullanılarak model eğitilmiştir.

*Tablo 3. Ridge Regresyon yönteminde kullanılan katsayılar (coefficients)*

<b>Ridge Regresyon katsayıları (coefficients)</b>						
<b>Kullanılan Değerler</b>	0.00408809	-0.00158097	0.00304043	1.07188584	-0.9843495	0.30401972



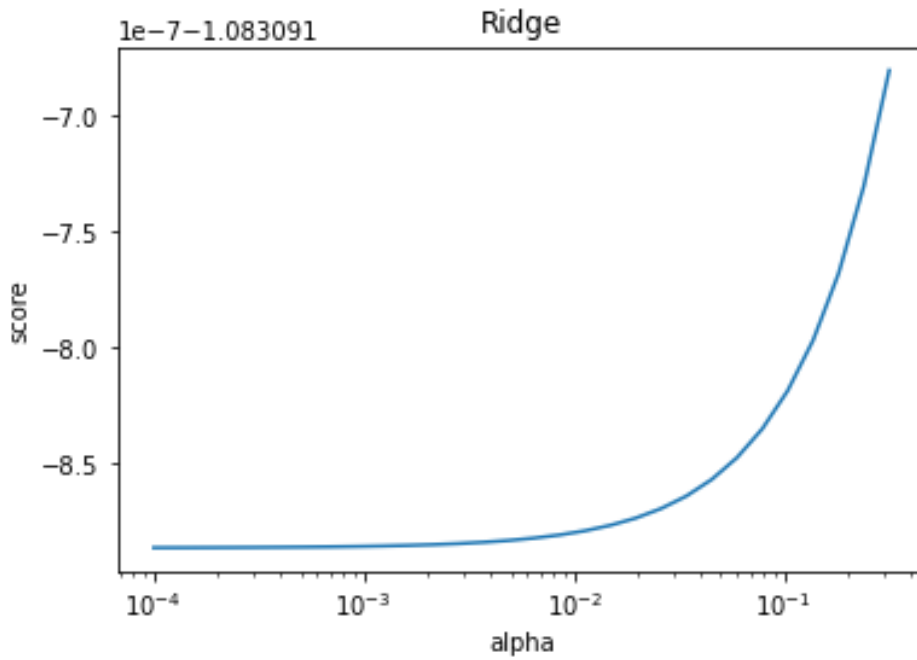
**Tablo 4.** Ridge Regresyonda En İyi Değerler

	<b>alpha</b>	<b>max_iter</b>	<b>random_state</b>
<b>Ridge Best Estimator</b>	0.31622776601683794	10000	42

Tablo 4’te verilen değerlerden alpha değeri bulunan en iyi alpha değeridir, max\_iter ve random\_state bizim belirlediğimiz parametrelerdir. max\_iter değeri maksimum iterasyon sayısını, random\_state ise belirlenen rasgele değerdir. random\_state belirlenmesinin amacı ise belirlenen sabit sayı girildiğinde sonucun hep aynı olmasıdır. Örneğin bu eğitimde random\_state yerine 42 yazıldığında eğitim farklı bir bilgisayarda da gerçekleşse bulunan sonuç aynıdır.

Tablo 3 ve Tablo 4’teki değerler kullanılarak model eğitildiğinde model doğruluk oranı:

Ridge Doğruluk: 0.9422747313642407 olarak bulunmuştur. Grafiği ise aşağıdaki gibidir.



**Şekil 4.** Ridge Regression Grafiği

Şekil 4’te alphanın score’a göre değişimi verilmiştir. Şekil 4’teki grafik incelendiğinde alpha değeri 10<sup>-2</sup> ye kadar sabit kalmış daha sonra ise alpha değeri arttıkça score değeri de artmıştır.

### C. 3. Lasso Regresyon Algoritması Bulguları

Ridge regresyon algoritmasında Tablo 5 ve Tablo 6’da verilen değerler kullanılarak model eğitilmiştir.

**Tablo 5.** Lasso Regresyon yönteminde kullanılan katsayılar (coefficients)

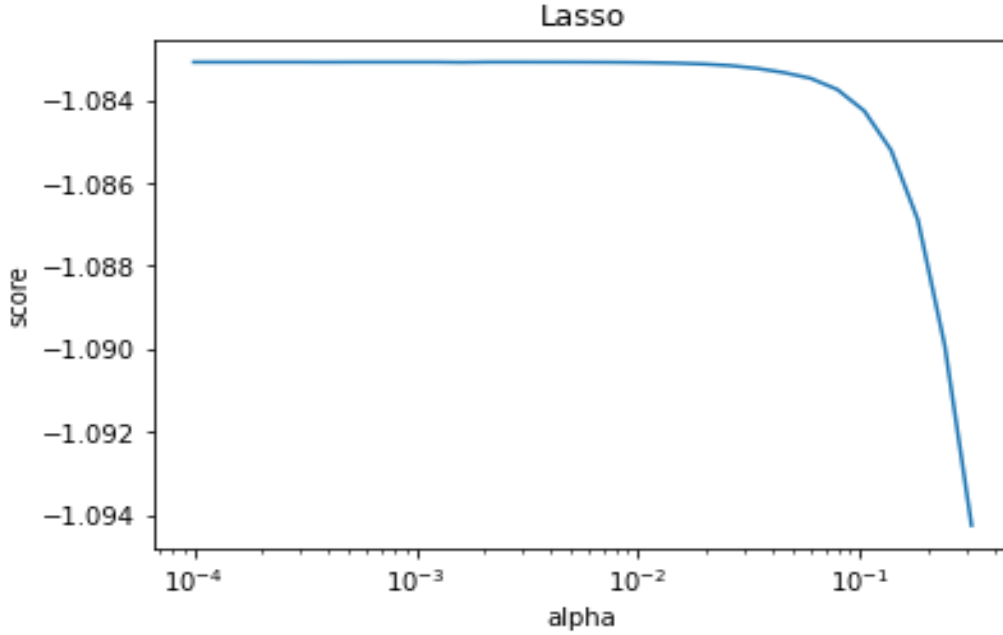
	<b>Lasso Regresyon katsayıları (coefficients)</b>					
<b>Kullanılan Değerler</b>	0.00407022	-0.00155213	0.00301901	1.07188815	-0.98434229	0.30382161

**Tablo 6.** Lasso regresyonda en iyi değerler

	<b>alpha</b>	<b>max_iter</b>	<b>random_state</b>
<b>Lasso Best Estimator</b>	0.0021261123338996556	10000	42

Tablo 5 ve Tablo 6'daki deęerler kullanılarak model eęitildięinde model doęruluk oranı:

Lasso Doęruluk: 0.9422382039109994olarak bulunmuştur. Grafięi ise aştadıki gibidir.



Şekil 5. Lasso Regresyon Grafięi

Şekil 5'te alphanın score'a göre deęişimi verilmiştir. Şekil 5'teki grafik incelendięinde alpha deęeri  $10^{-4}$  te en yüksek score deęerine sahip olup yaklaşık olarak  $10^{-1}$  alpha deęerine kadar bu deęeri korumuşt daha sonrasında ise score deęeri düşmüştür.

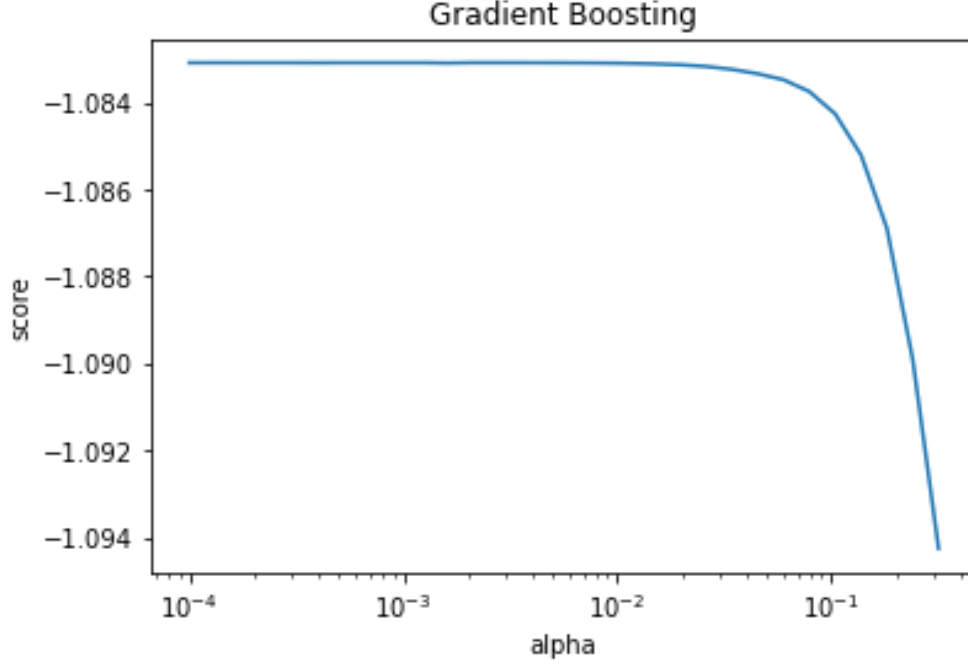
#### C. 4. Gradyan Artırma (Gradient Boosting) Regresyon Algoritması Bulguları

Gradyan Artırma (Gradient Boosting) regresyon algoritmasında Tablo 7'de verilen deęerler bulunmuştur.

Tablo 7. Gradyan Artırma (Gradient Boosting) Regresyon algoritması eęitim sonuçları

	Doęruluk Oranı	The mean squared error (MSE)
<b>Bulunan Deęerler</b>	0.998	2.0755

Grafięi ise aştadıki gibidir.



*Şekil 6. Gradyan Artırma (Gradient Boosting) regresyon grafiği*

Şekil 6’da alphanın score’a göre değişimi verilmiştir. Şekil 5’teki grafik incelendiğinde alpha değeri  $10^{-4}$  te en yüksek score değerine sahip olup yaklaşık olarak  $10^{-1}$  alpha değerine kadar bu değeri korumuş daha sonrasında ise score değeri düşmüştür. Şekil 6’da verilmiş olan Gradyan Artırma (Gradient Boosting) Regresyon Grafiği Şekil 5’te verilmiş olan Lasso Regresyon Grafiğine oldukça benzemektedir.

### C. 5. XGBoost Regresyon Algoritması Bulguları

XGBoost regresyon algoritmasında Tablo 8’de verilen değerler en iyi hiperparametre tahmini için kullanılmıştır.

*Tablo 8. XGBoost Regresyon algoritması için en iyi hiperparametre tahmininde kullanılan params değerleri*

	params değerleri				
colsample_bytree	0.4	0.5	0.6		
learning_rate	0.01	0.02	0.09		
max_depth	2	3	4	5	6
n_estimators	100	200	500	2000	

*Tablo 9. XGBoost Regresyon algoritması için en iyi değerleri elde eden hiperparametreler*

	colsample_bytree	learning_rate	max_depth	n_estimators
<b>En iyi hiperparametreler</b>	0.4	0.09	2	2000

Tablo 9’da verilmiş olan hiperparametreler, hiperparametre optimizasyonu yöntemlerinden biri olan GridSearchCV metodu kullanılarak 1800 farklı varyasyon arasından en iyi değerleri elde eden hiperparametreler olarak belirlenmiştir. GridSearchCV metodunda params değerleri kullanılmış ve 10 katlı cross validation uygulanmıştır.

Model doğruluk oranı ise %99.95 olarak ölçülmüştür.

### C. 6. Beş Algoritma İçin Elde Edilen Bulguların Karşılaştırılması

Veri setinde bulunan parametrelerin önem dereceleri aşağıda verilmiştir.

*Tablo 10. Veri setinde bulunan parametrelerin önem dereceleri*

	Tarih	İl	Bitki	Etc	Yağış	Gölge
<b>Doğrusal Regresyon</b>	0.073668	0.003452	0.004278	0.672296	0.237698	0.008607
<b>Ridge Regresyon</b>	0.073668	0.003452	0.004278	0.672296	0.237698	0.008607
<b>Lasso Regresyon</b>	0.073668	0.003452	0.004278	0.672296	0.237698	0.008607
<b>Gradyan Artırma</b>	0.073668	0.003452	0.004278	0.672296	0.237698	0.008607
<b>XGBoost Regresyon</b>	0.073668	0.003452	0.004278	0.672296	0.237698	0.008607

Tablo 10’da da görüldüğü gibi çalışmada kullanılan beş algoritma için her bir sınıfın önem dereceleri aynıdır ve Tarihin önemi %7.3, İlin önemi %0.345, Bitkinin önemi %0.427, Etc’nin önemi %67.22, Yağışın önemi %23.76, Gölgenin önemi %0.860 olarak belirlenmiştir. Çalışmada kullanılan regresyon algoritmaları içerisinde en yüksek doğruluk oranını %99.95 doğruluk oranı ile XGBoost başarılı model olarak belirlenmiştir.

## IV. SONUC

Tarım arazilerinde harcanan su miktarını yapay zekâ teknikleri kullanarak belirlenmesi gerçekleştirilmiştir. Çalışmada giriş parametresi olarak Tarih, İl, Bitki, Etc, Yağış, Gölge, çıkış parametresi olarak toplam sulama suyu ihtiyacı belirlenmiştir. Çalışmada makine öğrenme algoritmaları içerisinde yer alan Doğrusal (Linear) Regresyon, Ridge Regresyon, Lasso Regresyon, Gradyan Artırma (Gradient Boosting) Regresyon, XGboost Regresyon algoritmaları kullanılmıştır.

*Tablo 11. Gerçekleştirilen eğitim modellerinin doğruluk oranları*

	Doğrusal (Linear) Regresyon	Ridge Regresyon	Lasso Regresyon	Gradyan Artırma (Gradient Boosting) Regresyon	XGBoost Regresyon
<b>Doğruluk Oranı (%)</b>	94.23	94.23	94.22	99.8	99.95

Tablo 11’de verilmiş olan doğruluk oranları ışığında tahmin etme işleminde en doğru sonucu verecek olan algoritma %99.95 doğruluk oranı ile XGBoost Regresyon algoritması seçilmiştir ve seçilen bitkinin toplam sulama suyu ihtiyacı tahmin edilmiştir.

Bütün algoritmalarda veri setinde bulunan parametrelerin önem dereceleri aynıdır. En önemli parametre bitki su tüketimi olarak bilinen Etc olarak bulunmuştur. Etc, bir bitkinin ele alınan bir dönem için, her hangi bir hastalık zararlı etkisinin bulunmadığı, bitki besin elementi bakımından eksiklik çekmediği ve su stresi olmadığı standart koşullar altındaki su tüketimidir. İkinci en önemli parametre olarak Yağış parametresi bulunmuştur. Üçüncü en önemli parametre olarak Tarih parametresi bulunmuştur. Her bir bitkinin ihtiyaç duyduğu su miktarı büyüme mevsimi boyunca

farklılıklar göstermektedir bu nedenle bitkinin dikileceği tarihte oldukça önemlidir. Yapılacak olan sulama genel olarak Yağış miktarı, bitkinin Etc katsayısı ve tarih parametrelerine göre değişmektedir.

**TEŞEKKÜR:** Çalışmada kullanılan veri setinin oluşturulmasında açık erişimli internet sitesinde yayınlayan T.C. Tarım ve Orman Bakanlığına, verdikleri bilgilerden dolayı Dinar İlçe Tarım Müdürlüğüne teşekkürlerimi iletirim.

## **V. KAYNAKLAR**

- [1] B. Aksoy, H. D. Halis ve O. K. M. Salman, "Elma bitkisindeki hastalıkların yapay zekâ yöntemleri ile tespiti ve yapay zekâ yöntemlerinin performanslarının karşılaştırılması," *International Journal of Engineering and Innovative Research*, c. 2, s. 3, ss. 194-210.
- [2] E. Kılavuz ve İ. Erdem , "Dünyada tarım 4.0 uygulamaları ve Türk tarımının dönüşümü", *Social Sciences*, c. 14, s. 4, ss. 133-157, 2019.
- [3] S. Dorak , B. B. Aşık ve G. Özsoy , "Tarımda su kalitesi ve su kirliliğinin önemi: Bursa Nilüfer Çayı örneği," *Bursa Uludağ Üniversitesi Ziraat Fakültesi Dergisi*, c. 33, s. 1, ss. 155-166, 2019.
- [4] S. Örs ve M. Ekinçi , "Kuraklık stresi ve bitki fizyolojisi," *Derim*, c. 32, s. 2, ss. 237-250, 2016.
- [5] İ. Aras , "Damla sulama yöntemi", *Tarla Bitkileri Merkez Araştırma Enstitüsü Dergisi*, c. 15, s. 1-2, ss. 49-60, 2006.
- [6] İ. Terzi, M. M. Özgüven, Z. Altaş, & T. Uygun, "Tarımda yapay zeka kullanımı" *International Erciyes Agriculture, Animal & Food Sciences Conference Proceedings Book*, Kayseri, Türkiye, 2019, ss. 245-255.
- [7] A. Turgut , A. Temir , B. Aksoy ve K. Özsoy , "Yapay zekâ yöntemleri ile hava sıcaklığı tahmini için sistem tasarımı ve uygulaması," *International Journal of 3D Printing Technologies and Digital Industry*, c. 3, s. 3, ss. 244-253, 2020.
- [8] B. L. Aylak, O. Oral ve K. Yazıcı , "Yapay zeka ve makine öğrenmesi tekniklerinin lojistik sektöründe kullanımı," *El-Cezeri*, c. 8, s. 1, ss. 74-93, 2021.
- [9] K. Öztürk & M. E. Şahin , "Yapay sinir ağları ve Yapay Zekâ'ya genel bir bakış," *Takvim-i Vekayi*, c. 6, s. 2, ss. 25-36, 2018.
- [10] M. A. Kutlugün, "Gözetimli makine öğrenmesi yoluyla türe göre metinden ses sentezleme," Yüksek Lisans tezi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği, İstanbul Sabahattin Zaim Üniversitesi, İstanbul, Türkiye, 2017.
- [11] Y. Gültepe , "Makine öğrenmesi algoritmaları ile hava kirliliği tahmini üzerine karşılaştırmalı bir değerlendirme," *Avrupa Bilim ve Teknoloji Dergisi*, s. 16, ss. 8-15, 2019.
- [12] A. Arı and H. Onder , "farklı veri yapılarında kullanılacak regresyon yöntemleri," *Anadolu Tarım Bilimleri Dergisi*, c. 28, s. 3, ss. 168-174, 2013.
- [13] S. Kılıç , "Doğrusal regresyon analizi," *Journal of Mood Disorders*, c. 3, s. 2, ss. 90-92, 2013.

- [14] G. T. Kayaalp, M. Çelik Güney ve Z. Cebeci , "Çoklu doğrusal regresyon modelinde değişken seçiminin zootekniye uygulaması", *Çukurova Üniversitesi Ziraat Fakültesi Dergisi*, c. 30, s. 1, ss. 1-8, 2015.
- [15] M. R. Özkale, "Çoklu iç ilişki ile ilgili problemler," Doktora tezi, İstatistik Bölümü, Fen Bilimleri Enstitüsü, Çukurova Üniversitesi, Adana, Türkiye, 2007.
- [16] F. Üçkardeş, E. Efe, D. Narinç ve T. Aksoy "Japon bildircinlarında yumurta ak indeksinin ridge regresyon yöntemiyle tahmin edilmesi", *Akademik Ziraat Dergisi*, c. 1, s. 1, ss. 11-20, 2012.
- [17] H. K. Topal ve E. Ç. Akay, "Hanehalkı tüketim harcamalarının mikroekonometrik analizi LAD-LASSO yöntemi," *Journal of Econometrics and Statistics* , s. 33 , ss. 13-31, 2021.
- [18] D. Derici Yıldırım ve A. T. Çiftçi , "Etkili değişkenlerin cezalı regresyon yöntemleri ile belirlenmesi: Diyabet veri kümesi üzerine bir uygulama," *Mersin Üniversitesi Sağlık Bilimleri Dergisi*, c. 14, s. 1, ss. 105-112, 2021.
- [19] M. B. Keleş, A. Keleş ve A. Keleş, "Yapay zekâ teknolojisi ile uçuş fiyatı tahmin modeli geliştirme," *Turkish Studies*, c. 15, s. 4, ss. 511-520.
- [20] F. V. Şahinarslan, "Makine öğrenmesi algoritmaları ile nüfus tahmini: Türkiye örneği," Yüksek Lisans tezi, İşletme Lisansüstü Programı, Sosyal Bilimler Enstitüsü, İstanbul Teknik Üniversitesi, İstanbul, Türkiye, 2019.
- [21] A. Kurt, B. Buldu ve İ. H. Cedimoğlu, "Xgboost ve rastgele orman algoritmalarının ağ tabanlı saldırı tespitine yönelik performanslarının karşılaştırılması", *International Marmara Sciences Congress Proceedings Book*, c. 3, ss. 730-736, Kocaeli, Türkiye, 2020.
- [22] T. Chen, and C. Guestrin, "Xgboost: A scalable tree boosting system", *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, Seattle, USA, 2016, pp. 785-794.
- [23] S. S. Dhaliwal, A. A. Nahid and R. Abbas, "Effective intrusion detection system using XGBoost" , *Information*, vol. 9, no. 7, pp. 149, 2018.
- [24] H. Abar, "Xgboost ve Mars yöntemleriyle altın fiyatlarının kestirimi", *Ekev Akademi Dergisi*, c. 24, s. 83, ss. 427-446, 2020.
- [25] M. Saraswat, (2016, December 20). *Beginners Tutorial on XGBoost and Parameter Tuning in R* [Online]. Available: <https://www.hackerearth.com/blog/developers/beginners-tutorial-on-xgboost-parameter-tuning-r/>
- [26] E. K. Çakmak, "Arayüz tasarımında yeni bir yaklaşım: Paralel öğretim tasarımı", *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, c. 27, s. 1, ss. 1-22, 2007.