

Optimization of the Use and Planning of Water Resources with Advanced Data Processing Methods

Ugur Akbulut^{1,*}, Zafer Aslan²

¹Istanbul Aydın University, Graduate School of Education Program of Computer Engineering, 34295, İstanbul, Türkiye

²Istanbul Aydın University, Department of Computer Engineering, 34295, İstanbul, Türkiye

•Received Date: Apr 08, 2022

•Revised Date: Jun 15, 2022

•Accepted Date: June 24, 2022

•Published Online: Sep 14, 2022

Abstract

One of the most basic substances necessary for the continuation of life is water. The necessity of protecting and developing clean water resources that can respond to population growth in the future and today is of great importance. Water is one of the most basic resources that meet the energy needs of human beings and ensure their survival. It is necessary to use the existing water potential economically and to make water and its use efficient. Forward estimation can be made using the time series of river flow rates. Measurements are made daily at flow observation stations. With the data we get from here, we can make predictions with mathematical models and today's machine learning systems. Regression analysis is one of the methods used. In this study, the daily and monthly flow rate, precipitation and air temperature values of the Çatalca-Istıranca River, which was selected as the study area we determined, were taken into account. Statistically, the daily average flow 2.97 m³/s, daily total precipitation 2.73 (mm=kg/m²) and average air temperature value for the 2004-2020 review period has been determined as 12.57 °C. The variation of stream flow over time was estimated by using the machine learning methods Linear Regression, Support Vector, Decision Tree, Random Forest, Extra Trees and Wavelet methods. The success performances of the applied models were compared, and the success ratio of the Extra Trees (90.48%) and Random Forest (88.96%) methods were found to be higher than the other methods in estimating the flow rate.

Keywords

River Flow Rate, Machine Learning Methods, Wavelet, Random Forest, Extra Trees

*Corresponding Author: Ugur Akbulut, ugurakbulut1@stu.aydin.edu.tr, [id 0000-0003-2867-523X](https://orcid.org/0000-0003-2867-523X)

İleri Veri İşlem Yöntemleri ile Su Kaynaklarının Kullanımı ve Planlanmasının Optimizasyonu

Uğur Akbulut^{1,*}, Zafer Aslan²

¹İstanbul Aydın Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği Programı, 34295, İstanbul, Türkiye

²İstanbul Aydın Üniversitesi, Bilgisayar Mühendisliği Programı, 34295, İstanbul, Türkiye

•Gönderi Tarihi: 08 Nis 2022

•Düzeltilme Tarihi: 15 Haz 2022

•Kabul Tarihi: 24 Haz 2022

•Çevrimiçi Yayın Tarihi: 14 Eyl 2022

Özet

Canlı yaşamının devam etmesi için gerekli olan en temel maddelerden biri sudur. Gelecekte ve günümüzde nüfus artışına yanıt verebilecek temiz su kaynaklarının korunması ve geliştirilmesi, gerekliliği büyük önem arz etmektedir. Su insanoğlunun enerji ihtiyacını karşılayan ve hayatta kalmasını sağlayan en temel kaynaklardan birisidir. Mevcut su potansiyelinin tasarruflu kullanılması su ve kullanımının verimli hale getirilmesi gerekmektedir. Nehir akım hızı zaman serisi kullanılarak ileriye dönük su potansiyeli tahmini yapılabilmektedir. Akım gözlem istasyonlarında günlük ölçümler yapılmaktadır. Bu verilere dayalı olarak, matematik modellerle ve makine öğrenmesi sistemleri ile tahmin çalışmaları yürütülmektedir. Bu çalışmada inceleme bölgesi olarak seçilen Çatalca İstranca nehrine ait günlük ve aylık ortalama akım, bölgeye ait günlük toplam yağış miktarı ve günlük ortalama hava sıcaklık değerleri göz önüne alınmıştır. İstatistiksel olarak 2004-2020 inceleme dönemine ait, günlük ortalama akım $2.97 \text{ m}^3/\text{s}$, günlük toplam yağış $2.73 \text{ (mm=kg/m}^2\text{)}$ ve ortalama hava sıcaklığı değeri $12.57 \text{ }^\circ\text{C}$ olarak saptanmıştır. Makine öğrenme yöntemlerinden Lineer Regresyon, Destek Vektör, Karar Ağacı, Rasgele Orman, Ekstra Ağaçlar, Dalgacık yöntemleri kullanarak akarsu akış miktarının zamanla değişimi tahmin edilmiştir. Uygulanan modellerin başarı performansları karşılaştırılmış, akış miktarı tahmininde Ekstra Ağaçlar (%90.48) ve Rasgele Orman (%88.96) diğer yöntemlere göre daha başarılı bulunmuştur.

Anahtar Kelimeler

Akış Tahmini, Sıcaklık, Yağış, Regresyon, Dalgacık

*Sorumlu Yazar: Uğur Akbulut, ugurakbulut1@stu.aydin.edu.tr, [ID 0000-0003-2867-523X](https://orcid.org/0000-0003-2867-523X)

1. GİRİŞ

Oluşumu sırasında dünyamız birçok aşamalardan geçmiştir. Canlıların ihtiyaçlarını karşılayan doğal kaynaklar bu aşamalarda meydana gelmiştir. Bu doğal kaynaklar arasında yaşam için gerekli olan ve farklı alanlarda kullanılan su, canlılar ve canlıların faaliyetleri için en önemli ve vazgeçilemez olanıdır. Su, yaşamın başlangıcından itibaren tüm canlılara sunulmuş; varoluşun birincil kaynağı olmuştur. Su canlıların yaşam kaynağıdır. Su olmadan hiçbir canlı yaşayamaz ve suyun yerini tutacak hiçbir şey yoktur. Kısacası canlılar için susuz bir yaşam düşünülemez. İlk uygarlıklar su kaynaklarına kolaylıkla ulaşabilecekleri alanlarda yaşam ortamları kurmuşlardır. Dünyamızın doğal süreci gereği doğa olayları belirli olmayan ama tekrarlayan ve tahmin edilebilir bir döngü içerisinde. Bu sebeple insanlar yaşadıkları konumlar için zaman içerisinde olan taşkın, deprem, heyelan vb. sıra dışı doğa olaylarını önleyebilmek, kontrol altına alma, suyun depolama ve etkili kullanımı için çözümler üretmeye çalışmışlardır. Faydası ve zararını deneyimleyerek değişik yöntemler kullanmışlardır. Yapılan araştırmalarda, yazının keşfinden bu yana dünya'nın su kaynaklarından (Tablo1 ve Tablo 2) daha iyi faydalanabilme ve büyük afet etkilerinden korunabilme yönünde çalışmalar yapıldığı görülmektedir [1].

Tablo 1. BM verilerine göre su kaynaklarının yeryüzünde dağılımı su kaynaklarının yeryüzünde dağılımı (BM verilerine göre) [2]

KİTALAR	NÜFUS (%)	SU KAYNAĞI (%)
Avrupa	13	8
Asya	60	36
Afrika	13	11
Kuzey Amerika	8	15
Güney Amerika	6	26
Avustralya ve Adalar	1	5

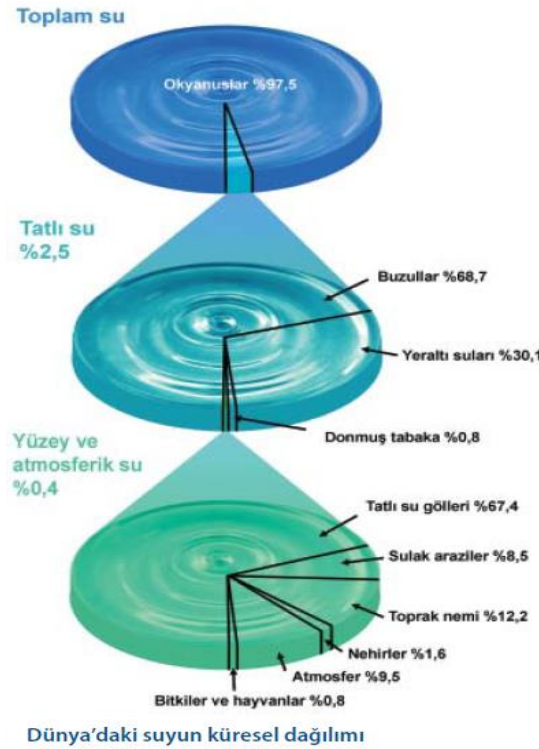
Tablo 2. Su kaynaklarının km^3 olarak dünya üzerindeki dağılımı [3]

SU KAYNAĞI	MİKTAR KM^3	%
Denizler	1,348,000,000	97.39
Kutuplar	27,820,000	2.01
Yeraltı Suları ve Akideler	8,062,000	0.58
Göl ve Nehirler	225,000	0.02
Atmosfer Buharı	13,000	0.0001
TOPLAM	1,384,120,000	100

Hidroloji, su bilimi anlamına gelmektedir. Hidrolojinin ana kavramlarından biri hidrolojik devirdir. Bu devir, suyun okyanuslardan atmosfere buharlaşmaya başlaması ve daha sonra yağış olarak yeryüzüne dönmesi, nehirlerin veya yeraltı sularının akışı yoluyla okyanus ve denizlere

ulaşması şeklinde tamamlanır [4]. Su devamlı bir döngü içerisinde. Güneşin ve doğanın etkisiyle su buharlaşarak bulutları meydana getirirken, rüzgârın etkisiyle bulutları harekete geçirerek su buharını yayarlar. Bulutlar belirli yoğunluğa ulaştıklarında yağış olayı gerçekleşir. Toprağa düşen yağmur suları sızarak yeraltı sularını oluşturur, akarsu ve nehirlerin devamlılığını sağlayarak yılda ortalama 400 milyar litre su bu döngüyü oluşturur. Suyun bu dolanımı, küresel su dengesini oluşturur ve bu çevrim, 'hidrolojik döngü' olarak tanımlanır [5].

Şekil 1'e göre dünyamızda bulunan suyun yaklaşık %97.5'i tuzlu su, %2.5 ise tatlı su olduğu bilinmektedir. Tatlı suyun büyük bölümü yani yaklaşık %2'lik kısmı kutuplarda bulunan buzullar ve buz kütlelerini oluşturmaktadır [6]. Buradan kutupların gelecekte daha büyük bir önem kazanacağını söyleyebiliriz. Donmamış tatlı suların bir miktarı, yeraltı suyu ile doymuş yeraltı kaya katmanlarında birikmektedir. Yüzeyle ve havada nem olarak bulunan su miktarı ise tüm tatlı suların yaklaşık %0.4'üne karşılık gelmektedir. Kullandığımız su, akarsu ve yeraltı sularından meydana gelmektedir.



Şekil 1. Dünyadaki suyun dağılımı (GreenFacts, 2020) [7]

Gelecek nesillere bırakmamız gereken su, küresel iklim değişikliğinin etkisiyle son yıllarda tehlikeli bir şekilde azalmaya başlamış olan bir doğal kaynaktır.

1.1. İklim Değişikliğinin Su Kaynakları Üzerindeki Spesifik Etkisi

İklim değişikliği su döngüsünü daha az tahmin edilebilir hale getirdi. Dünyada nüfus artışından kaynaklı plansız şehirleşme, arazi kullanım şeklinin değişmesi, akarsu yatağına yapılan

müdahaleler, üretim atıkları vs. çevresel etkenler su döngüsünü kökten değiştirmektedir [8]. Bundan dolayı daha yüksek sıcaklıklar su kütlelerinde çözünmüş besinlerin büyümesini kolaylaştırmakta ve zararlı alg patlamalarına, ölü bölgelere ve balık ölümlerine yol açmaktadır.

Artan kıtlık koşullarında yeterli ve sürekli su temininin sağlanması için aşağıdaki önlemlerin alınması gereklidir [7].

- Daha iyi planlama ve teşvikler yoluyla su kullanımını optimize etmek; Uygun olan yerlerde su teminini ve kullanılabilirliğini genişletmek;
- Daha iyi kentsel planlama, çiftçiyi korumak için mahsul sigortasını genişletme ve vatandaş katılımı ile aşırılıkların ve belirsizliklerin etkisini sınırlamak için su bütçesi ve ekonomisine önem vermek [7].
- Su kaynaklarımızı akıllıca yönetmezsek, küresel enerji ve tarım sistemlerimizin iklim nötrlüğüne yönelik önemli değişimlerini engelleyebilecek korkunç sonuçlar olabilir. Bu, özellikle büyük su eksikliği ile karşı karşıya olan düşük gelirli ülkelerde geçerlidir [9].

Ülkemiz topraklarına düşen yağış miktarı yıllık ortalama 501 milyar m³ civarındadır. 274 milyar m³ lük bölümü toprak ve su yüzeyinden buharlaşarak açığa çıkar ve hidrolojik döngüye geri kazandırılır. Akarsular aracılığıyla 158 milyar m³ lük bölümü denizlere ve göllere akmaktadır [10]. Ülkemize komşu ülkelere de 7 milyar m³ su girişi olmaktadır. Ekonomik ve teknik koşullara göre ülkemizdeki su potansiyeli verimli bir şekilde kullanılamamakta ve yeterli çalışmalar yapılamamaktadır. Komşu ülkelere gelen sulardan yaklaşık 3 milyar m³, akarsularımızdan ise 95 milyar m³ lük kısmı verimli bir şekilde kullanılmaktadır. Yani yerüstü su olarak yaklaşık 98 milyar m³, yeraltı suyu ise yaklaşık 14 milyar m³ olacak şekilde ülkemizde 112 milyar m³ yıllık kullanılabilir su miktarı mevcuttur [11].

Sonuç olarak bu verilerden anlaşıldığı gibi dünyada ve ülkemizde kullanılabilir su kaynaklarının az olmasından kaynaklı su kullanım miktarının kontrol altına alınması gerekmektedir. En yüksek seviyede verimlilik elde edilmesi, işletilmesi, depolanması ve su sebebiyle oluşabilecek afet ve zararlardan en az oranda etkilenilecek seviyelere getirilmelidir. Bu durum gelecek nesiller ve şu an yaşayan canlılar açısından büyük önem arz etmektedir. Bu nedenlerden dolayı suyun bilim çevrelerince incelenmesi ve farklı yöntemler geliştirilerek insanların bilinçlendirilmesi gerekir.

Su kaynaklarıyla ilgili projelerin oluşturulması ve bu projeler tasarlanırken su ve atmosfer verilerinin; diğer bir tanımla yeraltı ve yerüstü akım, rüzgâr, nem, sıcaklık, yağış vb. verilerin ayrı ayrı veya kombinasyonlar halinde incelenmesi büyük önem arz etmektedir. Akarsuların günlük akım verileri, minimum- maksimum ve ortalama akım değerleri ayrıca değerlendirilerek

faydalı sonuçlar üretilebileceği gibi hidrolojik ve meteorolojik veriler ile birlikte kullanılarak anlamlı sonuçlar çıkarılabilir.

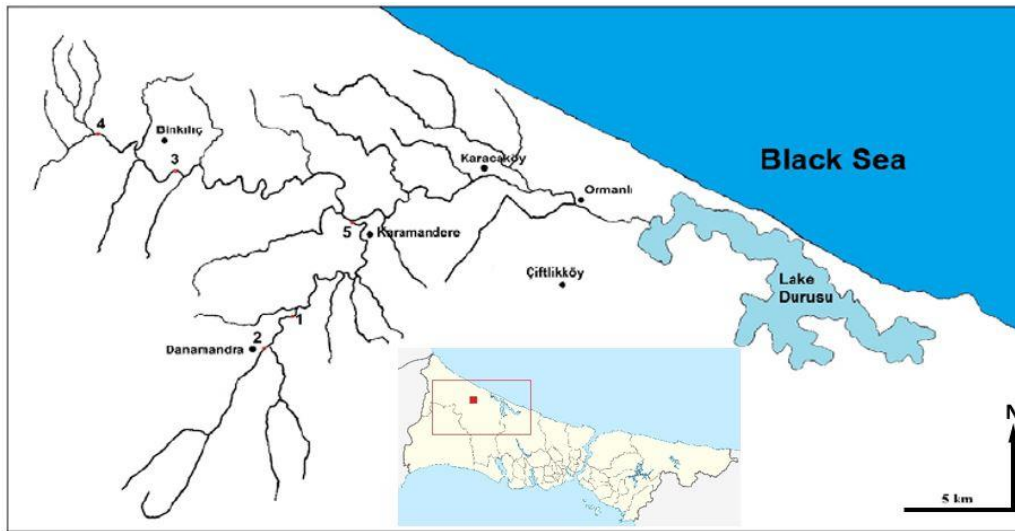
Bu nedenle aynı bölgelerdeki geçmiş akım verileri ile hava olaylarına ait veriler birleştirilerek regresyon ve wavelet dalgacık modeli kullanılarak geçmiş yıllardaki tekrar eden olayları saptayabiliriz. Belli aralıklarla tekrar eden olaylara bakarak ileri tarihlere dönük bir tahmin yapabiliriz. Yapılacak bu tahminlerin sonucunda suyun; optimum bir şekilde tüketilmesi, işletilmesi ve planlanması açısından etkili olacağını öngörebiliriz [12].

1.2. İnceleme Bölgesi ve Yeri

Bu çalışmada, Tablo 3'te konum bilgileri verilen Marmara Havzası D02A028 numaralı İstıranca D. Karamandere akım gözlem istasyonu inceleme bölgesi olarak göz önüne alınmış, bu istasyon verileri analiz edilmiştir. Tüm bölgeye ait akım verileri D02A028 Numaralı İstıranca Deresi Karamandere Köyü akım gözlem istasyonununun 1980-2020 yılları arası günlük Ortalama Akım verileri kullanılmıştır. Veriler Devlet Su İşlerinin internet sitesinden alınmıştır.

Tablo 3. İnceleme Bölgesi konum bilgileri

Havza/Şehir	İstasyon İsmi	Enlem Boylam	Veri Aralığı
Marmara/ İstanbul Çatalca	D02A028 İstıranca D. Karamandere	Boylam: 28°18'44" Doğu – Enlem: 41°22'45" Kuzey Rakım: 381	1980-2020



Şekil 2. İstıranca Deresi üzerindeki istasyonların konumu [13]

Yağış ve sıcaklık verileri Meteoroloji Genel Müdürlüğü 17047 Numaralı Çatalca Radar sahası istasyon 2004-2020 yılları arası günlük verileri kullanılmıştır.

2. YÖNTEM

Bu bölgeye ait Günlük Ortalama Akım, günlük Ortalama Yağış ve günlük Ortalama sıcaklık verileri;

- Linear Regresyon / Lineer İlişki
- Support Vektör / Destek Vektör
- Decision Tree /Kara Ağacı
- Random Forest /Rastgele Orman
- Extra Trees / Ekstra Ağaçlar
- Wavelet transform / Dalgacık Dönüşümü

yöntemleri kullanılarak akış tahmini yapılmıştır. Dalgacık dönüşümü kullanılarak yeni veriler elde edilmiş üstte belirtmiş olduğum yöntemler ile hybrid bir model oluşturulmaya çalışılmıştır. Çatalca (1) dışında Eskişehir Kütahya Porsuk Beşdeğirmen mevki (2), Tunceli Melekbahçe mevki (3), Isparta Eğirdir Göl girişi mevki (4) ve Isparta Gelendost mevki (5) incelenmiştir. Bu makalede sadece Çatalca Istanca ve civarı ile ilgili analizlere yer verilmiştir.

2.1 Korelasyon ve Regresyon Yöntemi

Korelasyon değişkenler arasındaki bağıntıyı belirlenmesini sağlar. En az iki farklı değişken değerleri arasındaki ilişkinin yönü ve şiddeti hakkında bilgiyi korelasyon analizi ile elde ederiz. Değişkenler arasındaki ilişki, bir neden-sonuç ilişkisi olmak zorunda değildir. Değişkenler arasındaki ilişkinin derecesi, korelasyon katsayısı olarak ifade edilir ve genel olarak “r” harfi ile gösterilir. Korelasyon katsayısının değeri -1 den küçük +1 den büyük olamaz, -1 ile +1 arasında bir değer alır.

- Bir değişkenin değerleri artarken diğer değişkenin değerleri artıyorsa, bu değişkenler arasında pozitif korelasyon vardır.
- Ancak değişkenler arasında ters orantı varsa biri artarken diğeri azalıyorsa, değişkenler arasında negatif korelasyon vardır.
- Korelasyon değerinin 0 (sıfır) olması durumunda değişkenler arasında ilişki yoktur [14].

Regresyon; değişkenler arasındaki vektörlerden birinin değerinin değişmesi ile diğer veri arasındaki ilişkiyi belirleyen terimdir.

Regresyon analizi değişkenlerden en az bir bağımlı, diğer değişkenler bağımsız olmak üzere aralarındaki ilişkinin, değişimlerden nasıl etkilendiğini inceleyen bir yöntemdir.

Regresyon ve korelasyon değişken değerleri arasındaki ilişkiyi araştıran yöntemlerdir. Regresyon analizinde ortaya bir fonksiyon elde edilir. Oluşan fonksiyon, bağımsız değişkenin alacağı değere göre bağımlı değişkenin alacağı değeri tahmin etmek için kullanılır. Bu yöntemle regresyon analizi denir.

Regresyon analizinde neden-sonuç ilişkisi vardır.

Bağımsız değişken neden sorusunun cevabını veren değişkenler veya değişiklik yapma olanağı olmayan değişkenlerdir.(X1)

Bağımlı değişken sonucunu tahmin etmek istediğimiz değişkenler olup (Y) ile gösterilmektedir.

Yani neden-sonuç ilişkisi vardır

$$Y=aX_1+bX_2+cX_3+\dots+k \quad (1)$$

Burada, X1, X2, X3 bağımsız değişkenleri, X: Bağımsız değişken, Y: Bağımlı değişken

a, b, c: değişken katsayısı, Regresyon doğrusunun eğimi,

k: lineer doğrunun Y-ekseni ile kesimini gösterir.

Bağımsız değişkenin bir adet olduğu regresyon analizine tek değişkenli regresyon analizi, bağımsız değişkenin birden fazla olduğu regresyon analizine çok değişkenli regresyon analizi denir [15].

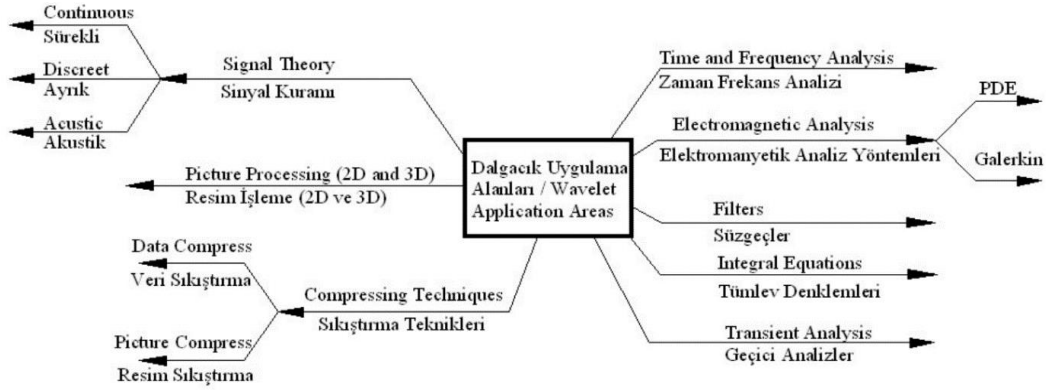
Determinasyon katsayısı (R2): Çıkan verilerin doğrusal bir eğriye uyduğunun en iyi ölçütü, regresyon analiziyle hesaplanan determinasyon katsayısıdır (R2).

Tanım aralığı 0-1 arasındadır. R2 = 1 olması, çıkan verilerin kusursuz bir doğrusal eğri sağlandığının kanıtıdır. Bu katsayı bağımlı değişken için gözlenen değerle modelde tahmin edilen değerlerin arasındaki korelasyon katsayısının karesidir. Veri noktası ne kadar fazla olursa, R2'nin güvenilirliği o kadar artacaktır. R2, bağımlı değişkendeki değişikliklerin, bağımsız değişkendeki farklılaşmalardan kaynaklanan yüzdesini vermektedir. Korelasyon katsayısı değeri (R2) 0'dan 1'e doğru yaklaştıkça çıkan tahminlerin doğruya yaklaştığı öngörülür [16].

2.2 Dalgacık (Wavelet) Dönüşümü

Dalgacık dönüşümleri, gibi sinyal işleme uygulamalarında (gürültü temizleme, sıkıştırma) başarılı sonuçlar veren, bir işaretin zaman-frekans optimizasyonunu sağlayan ve analizi için kullanılan bir sinyal dönüşüm türüdür [17].

Birçok alanda kullanılabilen dalgacık dönüşüm sismik verilerin incelenmesi ile başlamış "Sürekli Dalgacık Dönüşümünün" gelişmesini sağlamıştır. Fourier dönüşümünün hareketli sinyaller üzerindeki eksikliklerini ortadan kaldıran dalgacık dönüşümü, geliştirilmiş bir dönüşüm yöntemidir. Zaman frekans çözünürlüğü daha iyi sağlayan dalgacık dönüşüm yüksek frekanslı sinyaller için dar, düşük frekanslar için geniş boyutları aynı pencerede görebilme olanağı sunmaktadır. [17-18] Dalgacık dönüşümü çok çeşitli sinyal türlerine uygulanabilir (Bkz. Şekil 3) [19].



Şekil 3. Dalgacık dönüşümünün uygulama alanları [20].

Dalgacık dönüşümünde pencere görevini ana dalgacık fonksiyonu belirler. Bu ana dalgacık dönüşüm esnasında ölçeklenir ve ötelenir. Dalganın genişletilip daraltılması ile ölçekleme, zaman ekseninde dalganın kayması ile öteleme olayı gerçekleştirilir [21].

Ana dalgacığın fonksiyon kümesi $f(t)$ ile integrali olarak tanımlanabilir. Dalgacık fonksiyon kümesi $\psi_{a,b}(t)$, temel dalgacığın ölçeklenmesi ve ötelenmesiyle elde edilir.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad a > 0, a, b \in \mathbb{R} \quad (2)$$

a ölçekleme b ise ötelemeyi ifade eden değişkenlerdir. $f(t)$ işaretinin dalgacık dönüşümü

$$W_t(a, b) = \int_{-\infty}^{\infty} f(t) \psi_{a,b}(t) dt \quad (3)$$

Formüldeki gibidir [22].

2.3 Doğrusal (Linear) Regresyon

İki değişken arasındaki doğrusal ilişkiyi inceleyen istatistiksel yonteme doğrusal korelasyon ve basit doğrusal regresyon denilir.

Makine öğrenmesinde doğrusal (linear) model için denklemleri aşağıda belirtildiği gibidir.

Bağımsız değişken(ler) (x, \dots), Bağımlı değişken (y) ile ifade edilmektedir [23].

$$y' = b + w_1 x_1$$

→ y' öngörülen etikettir (istenen çıktı).

→ b bias (the y -intercept)

→ w_1 özellik 1 'in ağırlığıdır. Ağırlık “eğim” ile aynı kavramdır.

→ x_1 bir özelliktir (bilinen bir girdi)

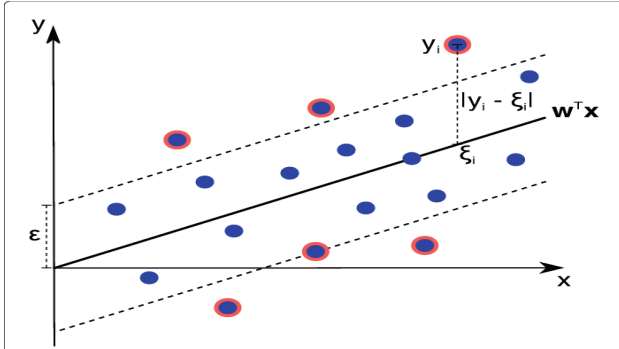
2.4 Destek Vektör Regresyonu (SVR)

SVR, birbirinden farklı değerleri tahmin etmek için kullanılan denetimli bir öğrenme algoritmasıdır. SVR' daki amaç, en uygun iki çizgi aralığını bulmaktır.

SVR aralığı oluşurken alabileceği en fazla noktayı içerisine almasını sağlamaktır. Çizilen maksimum aralıkların kestiği noktaları destek noktalarını belirler (Bkz. Şekil 4).

Diğer Regresyon modellerinden farklı olarak, SVR bir eşik değeri içindeki en iyi çizgiyi bulmayı amaçlar. Eşik değeri, hiper düzlem ile sınır çizgisi arasındaki mesafedir.

Doğrusal olmayan problemlerin çözümünde diğer yöntemlere göre problemleri çözmedeki yeteneği ve performansının çok daha iyi olduğu söylenebilir [24].



Şekil 4. Destek vektör regresyonu (SVR). $w^T x$ ile temsil edilen bir SVR regresyon fonksiyonunun resmi [25]

2.5 Karar Ağacı (Decision Tree)(DT)

Karar ağaçlarında anlaşılır ve mümkün olduğu kadar basit kural dizileri oluşturularak sınıflandırma işlemi yapılır. Yani sınıflandırmayı aşamalı olarak karar verme dizisi haline dönüştürmektedir.

- Kök, karar ağaçlarının ilk hücreleridir. Her bir aşamada kökteki koşula “Evet” veya “Hayır” veya daha basit cevaplara göre sınıflandırılır.
- Her bir gözlem düğümler aracılığıyla sınıflandırılır. Düğüm sayısının çokluğu modelin karmaşıklığının artmasını sağlar.
- Karar ağacının en altında bulunan bölümler bize sonucu verir [26].

2.6 Rassal Orman (Random Forest) (RF)

Karar ağacını çoklu halidir denilebilir. Birçok karar ağacı üreterek sınıflandırma yapan Rassal Orman algoritması, sınıflandırma değerini artırmayı amaçlayan bir algoritmadır. Birden fazla karar ağacı oluşturularak karar ormanı meydana çıkarılır. Rassal ormanda karar ağaçları bağlı olduğu veri setlerinden rastgele seçilerek alt kümeler oluştururlar.

Rassal orman iyi yanlarından birisi de sınıflandırma ve regresyon problemlerinde kullanılabilir olmasıdır [27-28].

2.7 Ekstra Rastgele Orman (Extra Tree Regressor) (ETR)

Rastgele Orman metodu, benzer bir yapıya sahiptir. Veri setlerinin kopyaları kullanılarak modeller eğitilir. Dalların ayrılması aşamasında düğümler için karar kriteri kullanarak en iyi ayrışmayı sağlamak yerine rastgele ayrışma yoluna gider.[28]

3. ANALİZ

Bu çalışmaya esas olan projenin alanı Marmara Havzası D02A028 Numaralı İstiranca Deresi. Karamandere akım gözlem istasyonudur. Veriler DSİ' kayıtlarından alınmıştır.

Yeri: Çatalca Karacaköy Bucağı Yolunun 23 Km'sindeki Karamandere Köyünün 1 Km akarsu ağzında yer almaktadır.

İstiranca Deresi-Çatalcaya ait verilerden 01.01.2005-30.09.2020 tarihleri arası Günlük Ortalama Akım, Günlük Ortalama yağış, Günlük Ortalama hava sıcaklığı verileri kullanılmıştır. 01.01.2005 -30.09.2017 yılları arasındaki 4656 adet günlük veriler eğitim için, 01.10.2017-30.09.2020 yılları arasındaki 1096 adet günlük veriler test için kullanılmıştır. Bu verilere ait istatistiksel değerler Tablo 4'te gösterilmiştir.

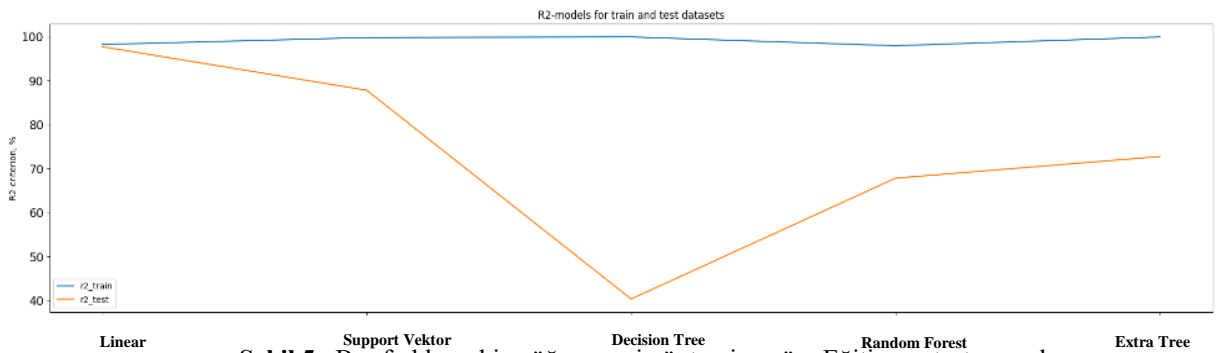
Tablo 4. Kullanılan veri kümesine ait istatistiksel değerler

D02A028 İstiranca D. Karamandere	Ortalama	Minimum	Maksimum	Standart Sapma	Çarpıklık	Değişinti	Mod (En Çok Tekrar)	Medyan (Ortanca)
Günlük Ortalama Akım(m ³ /sn)	2.97	0.06	206	8.19	9.18	67.04	0.2	0.76
Günlük Toplam Yağış(mm=kg/m ²)	2.73	0	184.2	8.92	8.61	79.54	0	0
Günlük Ortalama Sıcaklık(°C)	12.57	-8.8	31.1	7.47	-0.22	55.82	21.6	12.9

3.1 Makine Öğrenme yöntemleriyle Uygulamalar

Şekil 5'de beş ayrı makine öğrenme yöntemine göre akım değerlerinin tahmin sonuçları ve grafikleri gösterilmektedir.

Model	r ₂ _train	r ₂ _test	d_train	d_test	rmse_train	rmse_test
Linear Regression	98.29	97.77	67.73	70.81	102.75	159.62
Support Vector Machines Regression	99.86	87.87	8.93	33.9	29.63	372.07
Decision Tree Regressor	100.0	40.39	0.0	113.63	0.0	824.85
Random Forest	98.0	67.87	16.94	75.61	111.1	605.61
ExtraTreesRegressor	100.0	72.8	0.0	54.25	0.0	557.17



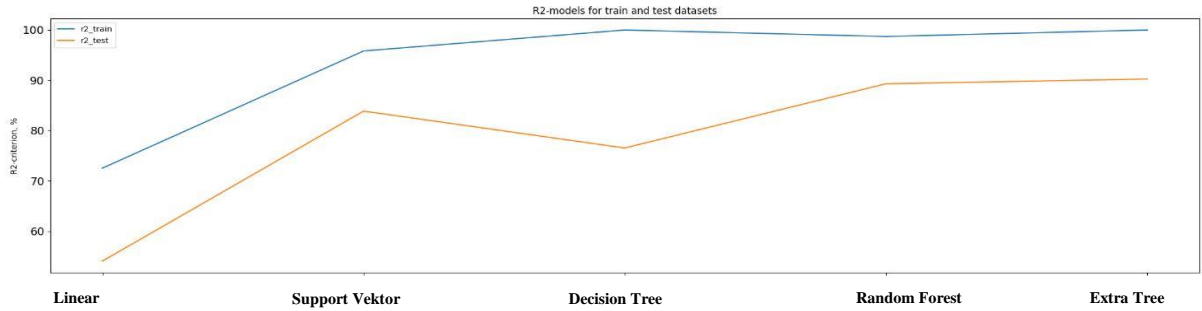
Şekil 5. Beş farklı makine öğrenmesi yöntemine göre Eğitim ve test sonuçları

Beş ayrı makine öğrenme yöntemine göre akım değerlerinin en başarılı tahmin sonucu Linear ve Support Vektör Regresyon modellerinde görülmektedir.

Şekil 6’da wavelet ayrıntı değerleri göz önüne alınarak oluşturulmuş hibrit uygulama sonuçları ve grafiği sunulmuştur.

Wavelet Ayrıntı değerleri göz önüne alınarak oluşturulmuş hibrit uygulama sonuçlarına göre Random Forest, Extra Trees ve Decision Tree yöntemlerinde Şekil 5’de hibrit uygulanmamış

Model	r2_train	r2_test	d_train	d_test	rmse_train	rmse_test
Linear Regression	72.52	54.04	243.76	236.49	440.7	422.4
Support Vector Machines Regression	95.83	83.87	72.76	92.53	171.66	250.26
Decision Tree Regressor	100.0	78.22	0.04	98.07	0.45	290.76
Random Forest	98.82	88.96	28.06	71.65	91.36	207.04
ExtraTreesRegressor	100.0	90.48	0.04	67.51	0.45	192.25



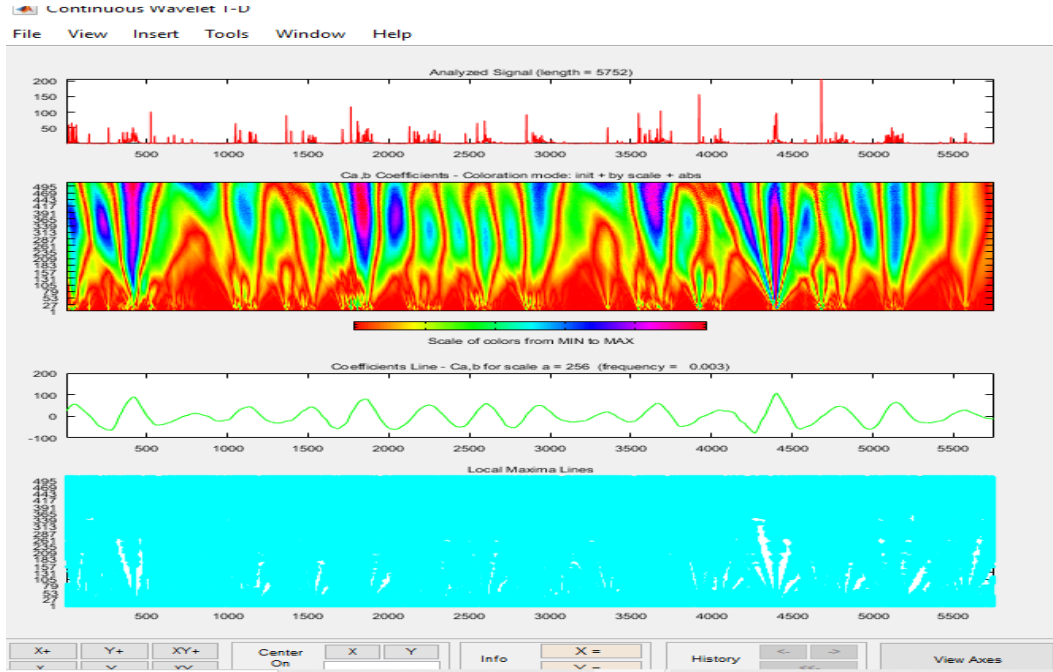
Şekil 6. Beş farklı makine öğrenmesi yöntemi ve Wavelet transformasyonu eğitim ve test sonuçları

modellere göre daha başarılı sonuçlar elde edilmiştir. Hibrit uygulanan modellerde belirgin başarı artış oranı gözlemlenmiştir.



Şekil 7. Dalgacık 1-D Db, Level 3

Şekil 7’de ortalama akım değerlerinin ayrıntıları sunulmaktadır. Son inceleme dönemi akım değerlerinde, büyük, orta ve küçük ölçekli etkilerin rolünde artış gözlenmiştir.



Şekil 8. Wavelet Continuous 1-D

Şekil 8’de Ortalama akım değişiminde küçük (kırmızı), orta(yeşil) ve büyük(mavi) ölçekli olayların rolü gösterilmektedir. Frekans ve ölçek ile ilgili ayrıntılar sunulmaktadır. Son inceleme dönemi akım değerlerindeki artışta küçük, orta ve büyük ölçekli olayların tümüyle etkili olduğu gözlenmiştir.

3.2 Ortalama Akım Tahmini Dağılımı

Şekil 9’da sunulan 10 farklı yöntem içerisinde sekiz modelde lineer ve SVR hariç diğer 6 modelin özellikle hibrid modele göre gözlem ve model sonuçları arasında $\alpha = 0.01$ güven seviyesinde anlamlı ilişkili olduğu saptanmıştır.

4. SONUÇLAR VE TARTIŞMA

Bu çalışmada günlük akım değerleri göz önüne alınmış, akarsu akım tahmini için ileri veri işlem yöntemlerinden yararlanılmıştır. Tahmin çalışmaları kapsamında farklı bilimsel alanlarda sıklıkla kullanılan beş farklı Makine öğrenmesi algoritmaları karşılaştırılmıştır. Makine öğrenmesi algoritmalarının sonucu karşılaştırılmış, Wavelet ayrıntıları da ilave edilerek, hibrid yöntemlerin hidrolojik verilere uygulanabilirliği denenmiştir. Makine öğrenmesi algoritmaları genel olarak wavelet modeli ile hibrid şeklinde göz önüne alınmış ve daha başarılı sonuçlar elde edilmiştir.



Şekil 9. A: Linear Regresyon B: Dalgacık Transformasyonu ve Linear Regresyon (Hibrid) C: Support Vector Regresyon D: Dalgacık Transformasyonu ve SVR (Hibrid) E: Decision Tree F: Dalgacık Transformasyonu ve Decision Tree (Hibrid) G: Random Forest I: Dalgacık Transformasyonu ve Random Forest (Hibrid) J: Extra Tree K: Dalgacık Transformasyonu ve Extra Trees (Hibrid)

Çatalca mevki D02A028 nolu İstıranca Deresi Karamandere Köyü Akım Gözlem İstasyonu “Günlük Ortalama Akım”, “Günlük Ortalama Hava Sıcaklığı”, “Günlük Toplam Yağış” verileri kullanılarak elde edilen sonuçlar wavelet Transformasyonu uygulanarak hibrid uygulama sonuçları ile karşılaştırılmıştır. Karar Ağacı, Rasgele Orman ve Ekstra Ağaçlar modellerinin Hibrid wavelet modeli ile birlikte uygulanması durumunda daha başarılı sonuçlar elde edilmiştir. Tahmin başarısını karşılaştıran grafikler (Şekil 6) verilmiştir. En başarılı tahmin sonucu W-Extra Ağaçlar ve W-Random Forest algoritmalarında görülmüştür. Doğa olaylarının düzensiz davranışları sergilemesi Wavelet (hibrid) modeli ile daha iyi açıklanmaktadır. W- Ekstra Ağaçlar modelinde belli periyotlarla salınımlarının tekrar ettiği saptanmıştır (Şekil 7- Şekil 8). Yüzde başarı oranları W- Ekstra Ağaçlar %91 ve W-Random Forest için %89 şeklinde görülmektedir. Bu yöntemler hidroloji alanında çalışma yapan araştırmacılar için geliştirilebilir yöntemler olarak yaygın bir şekilde kullanılması beklenmektedir. Burada belirlenen sonuçlar, bir sonraki araştırma konusunda ele alınarak, Yapay Sinir Ağı (ANN) ve Bulanık mantık (Fuzzy Logic) yöntemleri uygulanan sonuçlarla karşılaştırılabilir ve başarı oranları incelenebilir.

KAYNAKLAR

- [1] V. Batu, Yeraltı Suyu Hidrolojisi: Darcy ve Fick'ten Bu Yana Geçen Süre İçinde Yapılan çalışmalara Genel Bakış, Su Kaynakları, 1:1 (2008) 1-13.
- [2] Ankara Üniversitesi Açık Ders, “Dünyadaki Suyun Dağılımı Ve Su Tüketimi”, Erişim Tarihi 24.11.2021, <https://acikders.ankara.edu.tr/mod/resource/view.php?id=85361>
- [3] Ankara Üniversitesi Su Yönetimi, “Kentsel Ve Bireysel Su Tasarrufu”, Erişim Tarihi: 21.11.2021, <http://suyonetimi.ankara.edu.tr/wp-content/uploads/sites/88/2013/03/KENTSEL-VE-BIREYSEL-SU-TASARRUFU1.pdf>
- [4] N. Usul, Mühendislik Hidrolojisi. Ankara: ODTÜ Yayıncılık, 2017.
- [5] Bilim Ve Aydınlanma Akademisi, “Kapitalizmin Kısılcacında Su Kaynakları”, Erişim Tarihi: 11.10.2021, <https://bilimveaydinlanma.org/kapitalizmin-kiskacinda-su-kaynaklari>,
- [6] S. Özsoy, Su Ve Yaşam: Suyun Toplumsal Önemi, Ankara Üniversitesi, 2009.
- [7] GreenFacts, “Assessments of water resources and their management”, Erişim Tarihi 22.10.2021, <https://www.greenfacts.org/en/water-resources-assessments/index.htm>
- [8] V. Demir, A.Ü. Keskin, Yeterince akım ölçümü olmayan nehirlerde taşkın debisinin hesaplanması ve taşkın modellenmesi (Samsun, Mert Irmağı örneği), Geomatik, 7:2 (2022) 149-162.
- [9] SIWI, “COP26 has 4 goals. Water is central to all of them”, Erişim Tarihi: 27.10.2021 <https://siwi.org/latest/cop26-has-4-goals-water-is-central-to-all-of-them/>
- [10] İ. Çiçek, M. Ataoğlu, Türkiye'nin Su Potansiyelinin Belirlenmesinde Yeni Bir Yaklaşım, Coğrafi Bilimler Dergisi 7:1 (2009) 51-64.

- [11] K. Altan, A. Teksoy, A. Solmaz, Türkiye’de Yağış Ve Sıcaklığın Su Kaynakları, Tarımsal Ürün Verimi Ve Su Politikalarına Etkisi, *Uludağ Üniversitesi Mühendislik Fakültesi Dergisi*, 25:3 (2020) 1253-1270.
- [12] Ö. Terzi, M. Barak, Dalgacık-Sinir Ağı Yaklaşımı İle Yağış-Akış Tahmini: Kızılırmak Nehri Örneği, *Tarım Bilimleri Dergisi*, 21:4 (2015) 546-557.
- [13] Ç. Çelik, Istranca Deresi (İstanbul)’ndeki Tatlısu Kefalinin Yaş ve Büyüme Özellikleri, *Turkish Journal of Bioscience and Collections*, 3:1 (2019) 11-18.
- [14] H. Orhan, D. Kaşıkçı, Path, Korelasyon ve Kısmi Regresyon Katsayılarının Karşılaştırılması Olarak İncelenmesi, *Hayvansal Üretim* 43:2 (2002) 68-78.
- [15] Regresyon Analizi, Erişim Tarihi: 21.11.2021, <http://w3.balikesir.edu.tr/~bsentuna/wp-content/uploads/2013/03/Regresyon-Analizi.pdf>
- [16] Ankara Üniversitesi Açık Ders, “Korelasyon katsayısı”, Erişim Tarihi : 16.10.2021, https://acikders.ankara.edu.tr/pluginfile.php/62029/mod_resource/content/0/4.%20hafta.pdf
- [17] F. Aydın ve Z. Aslan, Yapay öğrenme yöntemleri ve dalgacık dönüşümü kullanılarak nöro dejeneratif hastalıkların teşhisi, *Journal of the Faculty of Engineering and Architecture of Gazi University* 32:3 (2017) 749-766.
- [18] J. Han, M. Kamber, J. Pei, *Data Mining Third Edition 3.5.2 Data Transformation by Normalization*, Elsevier, Waltham, 2012, pp. 113.
- [19] M. Küçük, N. Ağralıoğlu, Dalgacık Dönüşüm Tekniği Kullanılarak Hidrolojik Akım Serilerinin Modellenmesi, *İtüdergisi/d*, 5:2 (2006) 69-80.
- [20] G. Önbilgin, Ç. Kocaman, O. Özgönel, *Wavelets And Its Applications Of Power System Protection*, *G.U. J. Sci.*, 17:2 (2004) 77-90.
- [21] İ.V. Öner, M.K. Yeşilyurt, E.Ç. Yılmaz, *Wavelet Analiz Tekniği Ve Uygulama Alanları*, *Ordu Üniv. Bil. Tek. Derg.*, 7:1 (2017) 42-56.
- [22] H. K. Sevindir, S. Çetinkaya, C. Yazıcı, Makine öğrenmesi algoritmaları ve dalgacık dönüşümü ile EKG sinyalinden özellik çıkarımı, *BAUN Fen Bil. Enst. Dergisi*, 20:1 (2018) 94-109.
- [23] S. Kılıç, Doğrusal Regresyon Analizi, *Journal of Mood Disorders*, 3:2 (2013) 90-92.
- [24] M. Acı, M. Avcı, Ç. Acı, Destek Vektör Regresyonu Yöntemiyle Karbon Nanotüp Benzetim Süresinin Kısaltılması, *Journal of The Faculty Of Engineering and Architecture of Gazi University* 32:3 (2017) 901-907.
- [25] Destek Vektör Regresyonu (Svr), “Wtx İle Temsil Edilen Bir Svr Regresyon Fonksiyonunun Resmi.” Erişim Tarihi: 22.11.2021, www.Researchgate.Net/Uploaded By Frank M Boeckler
- [26] N. Beşli, M.E. Tenekeci, Uydu verilerinden karar ağaçları kullanarak orman yangını tahmini, *DÜMF Mühendislik Dergisi* 11:3 (2020) 899-906.
- [27] S. Demirezen, M. Çetin, Rassal Orman Regresyonu Ve Destek Vektör Regresyonu İle Piyasa Takas Fiyatının Tahmini, *Nicel Bilimler Dergisi*, 3:1 (2021) 1-15.
- [28] F.V. Şahinarslan, Makine Öğrenmesi Algoritmaları İle Nüfus Tahmini: Türkiye Örneği, İstanbul Teknik Üniversitesi Sosyal Bilimler Enstitüsü İşletme ABD, Yüksek Lisans Tezi, Haziran 2019.