

## Veri Madenciliği ve Makine Öğrenimi ile Döviz Kuru Tahmini Uygulaması

Ahmet Esin ERBUDAK<sup>1</sup>, Oğuz ATA<sup>2\*</sup>

<sup>1</sup> Bilgisayar Mühendisliği, Lisansüstü Eğitim Enstitüsü, Altınbaş Üniversitesi, İstanbul, Türkiye

<sup>2</sup> Yazılım Mühendisliği, Mühendislik Mimarlık Fakültesi, Altınbaş Üniversitesi, İstanbul, Türkiye

<sup>1</sup> ahmet.erbudak@hotmail.com <sup>2\*</sup> oguzata@gmail.com

(Geliş/Received: 23/02/2022;

Kabul/Accepted: 24/08/2022)

**Öz:** Makine öğrenimi 1990'lı yıllardan, veri madenciliği kullanılmaya başlandığı döneme dek yaygın olamamıştır. Bu çalışmada Türkiye Cumhuriyet Merkez Bankası sayısal veriler toplanan bilgilerin, veri madenciliği ile veri gruplarına göre bölümlere ayrılması sağlanmıştır. Makine öğrenimi ile ilgili kullanılan algoritmalar karar destek ağaçları, doğrusal regresyon, destek vektör makineleri, gauss regresyondur. Ocak 2020-Mayıs 2021 dönem süresindeki veriler kullanılarak oluşturulan döviz kuru modelleri oluşturulmuştur Sayısal halde toplanan veriler 1352 döviz kuru verisi için 4 ayrı makine öğrenimi yöntemi kullanılmış olup birinci yöntemde %99,84, ikinci yöntemde %99,18, üçüncü yöntemde %93,72, dördüncü yöntemde ise %86,83 doğru olarak sınıflandırılmıştır. Doğruluk oranı en yüksek orana göre strateji ve plan geliştirmektedir Veri madenciliği ve makine öğrenimi metotlarının tahmin programlarının yapımındaki değerlendirmeler ileriki dönemlerde yapılacak uygulamalar için faydalı olacaktır. Çıkarılan sonuçlar literatürde kullanılan metotların doğruluk oranları ile karşılaştırılarak analiz edilmiştir.

**Anahtar kelimeler:** Veri madenciliği, makine öğrenimi, tahmin, doğruluk, karar ağaçları

### Exchange Rate Forecasting Application with Data Mining and Machine Learning

**Abstract:** Machine learning did not become widespread until the 1990s when data mining began to be used. In this study, the numerical data collected from the Central Bank of the Republic of Turkey has been divided into sections according to data groups by data mining. The algorithms used for machine learning are decision support trees, linear regression, support vector machines, gaussian regression. Exchange rate models were created by using the data for the period of January 2020 -May 2021. Four different machine learning methods were used for 1352 exchange rate data, and they were classified as 99.84% accurate in the first method, 99% in the second method, 93.72% in the third method, and 86.83% in the fourth method. It develops strategies and plans according to the highest accuracy rate Evaluations in the construction of prediction programs of data mining and machine learning methods will be useful for future applications. The results were analyzed by comparing the accuracy rates of the methods used in the literature.

**Key words:** Data mining, machine learning, prediction , accuracy, decision trees

#### 1. Giriş

Veri bilimi yapılandırılmış ve yapılandırılmamış verilerle bilginin kayıt altına alınması, saklanması, işlenmesi için bilimsel yöntemleri süreçleri, algoritmaları kullanarak bilgiye dönüşümü olarak tanımlayabiliriz. Veri madenciliği sistemlerinde kullanılan algoritma biçimine göre; verilerin içinde gömülü olan sınırlı bilgiler bulunmakta, başka ihtiyaçları karşılanmakta ve verilerin farklı yorumlamaları yapılmaktadır [1]. Veri madenciliği tanınmayan, saklı ve faydalı örüntülerin çok boyutlu veri tabanlarından otomatik şekilde elde edilmesini sağlayan veri tabanlarındaki bilginin araştırılması ve analiz sürecidir [2]. Veri madenciliği veri tabanı teknolojisi, istatistik, yapay zeka, makine öğrenimi, örüntü tanımlama ve veri görselleştirme gibi pek çok teknik alan arasında bağlantı sağlayan çok disiplin bir alandır [3]. Veri madenciliği, yapay zeka ve istatistik alanlarındaki gelişmelerden yararlanır. Her iki disiplin de örüntü tanıma ve sınıflandırma sorunları üzerinde çalışmaktadır. Her iki toplum da sinir ağlarının ve karar ağaçlarının anlaşılmasına ve uygulanmasına büyük katkılarda bulunmuştur [4]. Veri madenciliği kullanıldığı alanlara örnek vermek gerekirse; pazarlama, bankacı sektörü, uzay bilimleri, tıp , sosyal bilimler gibi alanlarda sıkça kullanılmaktadır. Veri madenciliği bu bağlamda ,yapay sinir ağları, karar ağaçları, regresyon analizi gibi tekniklerle içi içedir. [5]. Bu çalışma da insan-bilgisayar etkileşiminin en önemli yöntemlerinden olan 4 model makine öğrenimi hesaplamaları incelenmiştir.

Veri madenciliği sürecinde izlenen adımlar genellikle aşağıdaki şekildedir [6] :

1. Problemin tanımlanması
2. Verilerin hazırlanması,

\* Sorumlu Yazar [oguzata@gmail.com](mailto:oguzata@gmail.com) Yazarların ORCID Numarası: <sup>1</sup> 0000-0002-1327-4866 , <sup>2</sup> 0000-0003-4511-7694

3. Modelin kurulması ve değerlendirilmesi,
4. Modelin kullanılması,
5. Modelin izlenmesi.

Bu çalışma da insan-bilgisayar etkileşiminin en önemli yöntemlerinden olan 4 model makine öğrenimi hesaplamaları incelenmiştir. Bu çalışmada cevap aranan temel sorular şu şekilde sıralanmıştır :

- Veri madenciliği teknikleri nelerdir ?
- Tahmin ve strateji oluşturmada hangi modeller uygulanmalıdır?
- Veri Madenciliğinin makine öğrenimin etkisi nedir? Makine öğrenimin gelişimi ve kullanılması gereken alanlar için hangi argümanlardan faydalanılmaktadır?

## 2. Benzer Çalışmalar

2007 yılında, veri madenciliği kullanılarak yaşam çözümlenmesi yöntemleri kredi kartı sahipleri bilgileri ile ele alınmış ve daha sonra aynı bilgileri içeren veri kümesi yaşam olasılıkları, regresyon modelleri ve hazard olasılıkları modelleri Nihal Ata, Erençül Özkök ve Uğur Karabey tarafından incelenmiştir. Çalışma sonucunda müşterilerin kredi kartı kullanımını bırakma durumunu medeni durum, yaş, ve gelir bilgileri ile ilişkili olduğu görülmüştür[7].

2009 yılında, finansal tablolardaki hileleri tespit etmek için veri madenciliği kullanılarak H. Ali Ata ve İbrahim H. Seyrek tarafından yapılan çalışmada İMKB’de işlem gören imalat sektöründe faaliyet gösteren 100 firma bilgileri incelenmiştir. Bu çalışma sonucunda finansal tablodaki hilelerin aktif karlılık ve kaldıraç oranı ile ilişkili olduğu tespit edilmiştir[8].

Erhan Özüç (2016), tarafından uluslararası Fisher etkisini Türkiye üzerine bir uygulama ile çalışmıştır. Çalışmasının sonuç kısmında döviz piyasalarında tahmin için çeşitli modellerin kullanılabilceğini, bu modellerin döviz kuru tahminlerinde ilgililerine fikir sunabileceğini belirtmiştir [13].

Laurids Gert Nielsen (2018) tarafından döviz kuru için makine öğrenimi üzerine bir uygulama ile çalışma yapılmıştır. Yazar R-kare yöntemini kullanarak kısa ve uzun zamanlı tahmini değerler için çalışmada bulunmuştur. Rastgele orman algoritması, arttırılmış ağaç yöntemi ve uzun , kısa süreli hafıza yöntemi modellerine göre daha yakın sonuç tespit edilmiştir. Tahmin modellerindeki farkın nedeni verilerdeki çok fazla değeri yüksek miktarda model parametresiyle birleştirdiğimizde, verilerin kolayca fazla sığdırılacağını ve model karmaşıklığını önemli ölçüde azaltmak zorunda kalınacağını, bu da model yeteneklerinin tam olarak kullanılamayacağını belirtmiştir[19].

Hamis Miraji Ally Simba (2020) tarafından faiz ve döviz kuru ilişkisini içeren Türkiye üzerine yapılan bir değerlendirme çalışması yapılmıştır. Çalışmada Mundell-Fleming modeli kullanılmıştır. Bu modele göre, dışa açık ve esnek kur mekanizması benimseyen bir ekonomide yurtiçi faiz oranlarındaki bir yükselme ülkeye döviz girişini arttırmakta ve bu durum döviz kurlarının düşmesine neden olmaktadır. Çalışmada önce faiz ve döviz kuru ilişkisine ilişkin literatüre yer verilmiş devamında değişkenler arasındaki ilişki Johansen eş bütünleşme ve Granger nedensellik analizi çerçevesinde incelenmiştir[20].

Veri madenciliği ve makine öğrenimin döviz kuru tahmini alanında yazılan çalışmalar genellikle “Döviz kuru öngörü sonucu“, “Döviz kuru tahmininde yapay sinir ağları”, ”Makine öğrenimin finans uygulamalarına etkisi” , “ Döviz kurunun istatistiksel öğrenimi” gibi konular ele alınmıştır. Bu bağlamda yapılan çalışmalarda makine öğreniminin veri madenciliği ile ilgili etkileşim süreçlerine, dövizde uygulanan faiz oranlarının kur nasıl etkilediğine, doğrusal mı, kübik mi yoksa kümeleme fonksiyon olup olmadığına odaklanılmıştır.

## 3. Amaç ve Hedef

Makine öğrenimi modelleri kullanılarak gerçek verilerin makine öğrenimi teknikleri ile etkileşime girdiğinde oluşacak durumlar açısından analiz etme, doğruya en yakın hesaplamalar amaçlanmıştır. Araştırma özellikle COVID-19 döneminde olan verilerden sağlanmıştır. Bunun nedeni olağan dışı durumlarda makine öğrenim modellerinin ekonominin durağanlık döneminde doğruluk payının ne kadar olacağı amaçlanmıştır. Geliştirilen modellerin ekonomi ve istatistiksel çalışmalarda doğru stratejiler

belirlenebilmesi için sınıflandırma başarısı olarak %90 başarı oranının sağlanması hedeflenmiştir. Çalışmada yer alan makine öğrenimi modelleri ve fonksiyonlar ile hata payının gerçek veriler üzerindeki sapma oranının ne kadar olacağı hedeflenmiştir. Bu sapma oranına göre geleceğe yönelik tahmin hesaplanması amaçlanmıştır. Çalışmanın önemli amaçlarından biri de kur dalgalanmalarının ülke ekonomisine verdiği zararın önüne geçmek ve en aza indirgenmesi hedeflenmiştir. Bu çalışmanın cevap aradığı temel amaç gerçek veriler kullanılarak makine öğrenimi teknikleri aracılığıyla en yüksek doğruluk oranı yakalamaktır. Bu konuyu açıklama argümanı ise matematiksel varyasyonlar sonucu tahmin fonksiyonlarının dış etkenler dışında verilere göre değişkenlik gösterdiği ve bu sapmanın nedeninin piyasalarda gerçekleşen döviz alım satımının ortaya çıkardığı etkidir. Ülkemizde ve dünya gen elinde diğer çalışmalar incelendiğinde olağan dönemde yer alan veriler ile hesaplamalar yapılmıştır. Bu modeller hata payı en düşük olan tekniğin diğer teknik modeller ile karşılaştırılabilmesi amacıyla yapılmış olup, doğru sonuca ulaşılma hedeflenerek ülke ekonomisinin döviz kuru dalgalanmaları sonucu zarara uğramasını önleyebilmek amacıyla yapılmıştır.

#### 4. Yöntem

Bu çalışmada program için kullanılan veriler Türkiye Cumhuriyeti merkez bankasından alınan değerler kullanılmıştır. Araştırma materyali olarak MATLAB R2020 programı kullanılmıştır. Toplanan verilerle 4 model makine öğrenimi teknikleri kullanılarak eğitilmiştir. Kullanılan teknikler karar ağaçları, doğrusal regresyon, destek vektör ve gauss algoritmalarıdır. Çalışma içerisinde kullanılan 4 model sonuçları verilmiş ve karşılaştırmaları yapılmıştır. Verilerin eğitim aşaması döviz kuru gerçek değeri ve faiz oran ı olarak iki sütunlu ayarlanmıştır. Modellerde 4 adet bağımsız parametre olduğu için 4 adet girdi nöronu vardır. Kullanılan metrik fonksiyonlar kök ortalama kare hata, r-kare, ortalama kare hata ve ortalama mutlak hata'dır. Makine öğrenme modelinde, tahminleyicinin elde ettiği değerler ile doğru değerler arasındaki mesafenin bulunmasında çoğunlukla kullanılan ve hatanın büyüklüğüne ölçen kuadratik bir metriktir.

Karar ağacı indüksiyonu, sınıf etiketli eğitim verilerinden karar ağaç mekanizmasının öğrenilmesidir. Karar ağacı, her bir iç düğümün bir öznelik üzerinde bir testi gösterdiği, her dal testinin bir sonucunu temsil ettiği ve bir sınıf yapısının tuttuğu akış şemasına benzeyen bir ağaç yapısı şeklindedir [3]. Karar ağacı modeli ağaç görünümünde tanımlayıcı ve tahmin edici bir yöntemdir. Karar ağacı algoritması, veri miktarı, algoritmanın etkinliği ve hali hazırda hafızaya göre seri ve paralel süreçlerle yapılabilir. Seri ağaç algoritması, bir eğitim veri kümesi kullanılarak oluşturulan ikili ağaç olarak mantıksal bir model türüdür. Araştırmalarda, bağımsız değişkenlerden yararlanarak bir hedef değişkenin değerini tahmin etmede kullanılır[13].

Doğrusal regresyon algoritmaları sayısal bir sonuç ile bir sayısal değişken arasındaki bağıntıyı analiz ederken, doğrusal regresyon en yaygın olarak tercih edilen analiz yöntemidir. Doğrusal regresyonda, genel olarak değişkenin birçok değişik değerine sahiptir ve genellikle değişkenlerin gözlemlenen değerleri arasındaki olası değerleri varsayılmaktadır. Algoritma sonucunun ortalama değer ve değişkenin değeri arasında doğrusal bir ilişki olduğu varsayılmaktadır[15].

Bir destek vektör makinesi, gruplandırma, geçmişe ya da uygunsuz değerlerin saptanması gibi farklı görevler için kullanılabilen üstün veya ebedi boyutlu bir uzay alanında tanımlanan düzlem oluşturur [13]. Doğru bir fark, rastgele bir grubun en hazır eğitim veri konumuna , en çok uzaklık sahibi olan tanımlanan düzlem tarafından sağlanır, bu nedenle genel olarak sınır ne kadar çoksa hata oranı o kadar az olur[16].

Gauss yöntemini kapsayan makine öğrenimi algoritması, eğitilmiş verilerde gizli bir alanın değerini tahmin etmek için noktalar arasında paralellığın ölçüsü olarak (kernel fonksiyonu) kullanılmaktadır. Verilerin tahmini sadece o alan için tahmin edilmemektedir, aynı zamanda tek boyutlu Gauss dağılımıdır[17].

Regresyon, odak noktasının bir bağımlı parametre ile bir veya birden fazla bağımsız parametre arasındaki ilişki olduğunda, çeşitli değişkenleri modellemek ve analiz etmek için birçok teknik içerir. Regresyon analizi, bağımsız parametrelerden herhangi biri değiştiğinde, diğer bağımsız değişkenler sabit tutulduğunda bağımlı değişkenin değerinin nasıl değiştiğini kavramaya yardımcı olur. Regresyon analizi, bağımsız parametreler verildiğinde bağımlı değişkenin koşullu beklentisini tahmin eder yani, bağımsız parametreler belirlendiğinde bağımlı değişkenin yaklaşık değeridir[18].

KOKH(Kök ortalama kare hata) tahmin hata oranının klasik sapmasıdır. Hatalar, regresyon hattının veri noktalarından ne kadar uzakta olduğunun bir ölçüsüdür. KOKH(Kök ortalama kare hata) değerinin "0" olması modelin hatasız olduğunu gösterir [9].

$$KOKH = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad [9]$$

$n$ =Küme verisinin boyutu ,  $e$  = Gerçek değer- tahmini değer

R-kare, verilerin regresyon hattına ne kadar yakında olduğunu gösteren istatistiksel bir çözümdür. Benzer sürede belirleme katsayısı olarak da bilinir. R-kare , doğrusal bir model tarafından yorumlanan sonuç değişkeninin varyans yüzdesidir. R-kare her zaman 0 ile 1 değerleri arasındadır. “0” değeri model verilerinin hiçbirini açıklamadığını ifade eder. “1” değeri model verilerinin tamamının açıklandığını ifade eder [10].

AKT(Artıkların kareler toplamı)= Gerçek değer- Tahmini değer .Doğruya isabet edemeyen her nokta artıktır.

Her bir noktanın ortalamaya uzaklığının karesine, ortalamaya uzaklığın kareler toplamı (OUKT) denir.

OUKT =  $(\bar{X}I - \bar{X}J)$  şeklinde gösterilir. Hataların toplamının, ortalamaların toplamına olan oranı ne kadar küçük ise  $R^2$  o kadar yüksek olacaktır.

$R^2 = 1 - \text{Artıkların kare toplamı} / \text{Ortalama uzaklığın kare toplamı}$

$R^2 = 1 - \frac{\sum_i (y_i - \bar{y})^2}{\sum_i (y_i - \bar{y}_i)^2}$  [11]

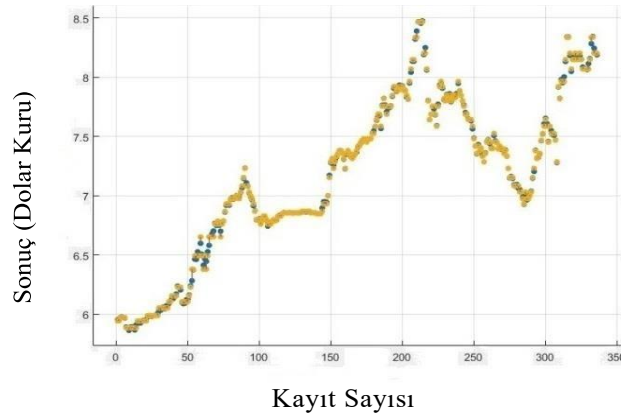
Ortalama kare hatası, regresyon kıvrımının bir dizi alana yakın olduğunu gösterir. Ortalama kare hatası, eğitim modelinin, tahmin edici verilerin verimini ölçer, her zaman artı değerdedir ortalama kare hatası değerinin sıfıra yakın olması tahmin edici veriler daha iyi verim sağlamaktadır. Ortalama kare hatası formülü;

$$OKH = 1 / \sum_{j=1}^n e^2 n_{j=1} \text{ [9]}$$

Ortalama Mutlak Hata, iki dayanaklı değişken veri arasında ayırımın miktarıdır. OMH(Ortalama mutlak hata), her doğru ölçüyle, veriye en uygun hat arasındaki ortalama dikey uzaklıktır. OMH(Ortalama mutlak hata), bir sıra tahmindeki hataların ortalama boyutun oranlayan, tüm hataların ortalamada denk olarak yoğunluğunun doğrusal bir sonucudur. OMH(Ortalama mutlak hata) değeri “0” ile sonsuz sayıda farklılık gösterebilir. Düşük değerlere sahip tahmin ediciler daha iyi verime sahiptirler.

$$OMH = 1 / \sum_{j=1}^n |e_j|$$

Bu çalışmadaki modellerde 341 döngü kullanılarak eğitilmiştir. Eğitim işlem sürecinden sonra gerçek dolar kuru verileri ile tahmin edilen değerler grafiklerde gösterilmiştir. Karar ağaçları makine öğrenimi, doğrusal regresyon makine öğrenimi, destek vektör makine öğrenimi, gauss regresyon makine öğrenimi yöntemleri grafik ve tablolardaki verilerle açıklanmıştır.



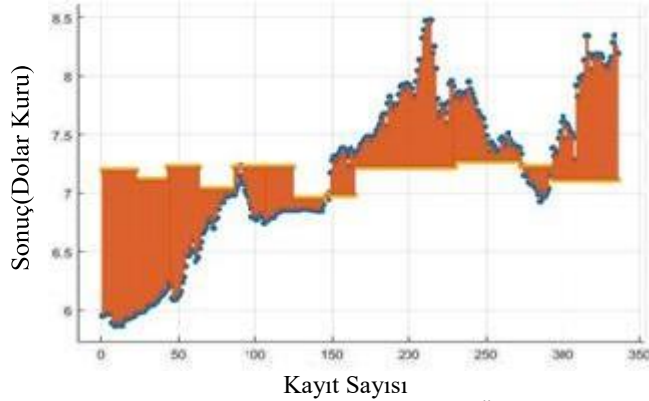
Şekil 1. Karar Ağaçları Makine Öğrenimi Grafiği.

Karar ağaçları makine öğreniminde şekil 1'deki grafikte yer alan mavi renkli noktalar gerçek verilerdir. Sarı ile gösterilen noktalar ise tahmin edilen değerdeki sayılardır. Kırmızı ile belirtilen çizgiler hata noktalarıdır. 30 Nisan 2021 tarihinde olan gerçek veri dolar kuru bazında 8,190'dur. Eğitilmiş ve tahmin edilen verinin değeri ise 8,203 olarak çok yakın bir değerde tahmin edilmiştir. Gerçek veri ile arasında 0,013 değerinde fark bulunmaktadır. Bu eğitim verisinde yer alan hata oranları ile ilgili veriler tablo 1'de verilmiştir.

**Tablo 1.** Karar Ağaçları Analizi.

Kök Ortalama Kare Hatası	R-Kare	Ortalama Kare Hatası	Ortalama Mutlak Hata
0,01522	1,00	0,00023164	0,0097679

Doğrusal regresyon makine öğreniminde grafikte yer alan mavi noktalar gerçek verilerdir. Sarı ile gösterilen noktalar tahmin edilen değerlerdir. Kırmızı ile belirtilen çizgiler gerçek veri ile tahmin edilen verinin hata noktalarıdır. 30 Nisan 2021 tarihinde olan dolar kuru 8,19'dur. Eğitimi yapılmış ve tahmin edilen veri 7,112'dir. Gerçek veri ile arasında 1,078 değer farkı bulunmaktadır.

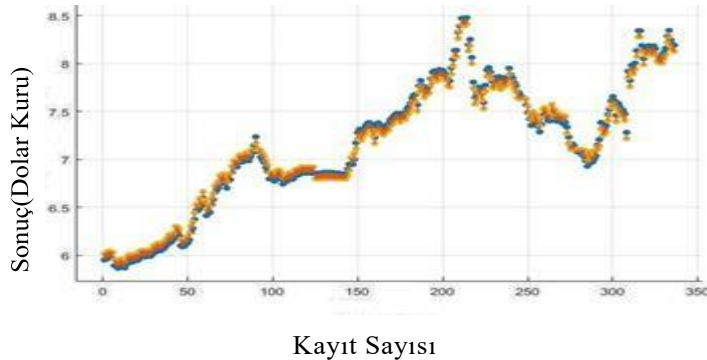
**Şekil 2.** Doğrusal Makine Öğrenim Grafığı.

Şekil 2'de yer alan grafikte doğrusal makine öğrenimi verilerinin eğitimi sonuçları ve hataları belirtilmiştir. Bu eğitim verisinde elde edilen veriler tablo 2 de yer almaktadır.

**Tablo 2.** Doğrusal Model Analizi.

Kök Ortalama Kare Hatası	R-Kare	Ortalama Kare Hata	Ortalama Mutlak Hata
0,66752	0,2	0,44558	0,5407

Destek vektör makine öğreniminde şekil 3'deki grafikte yer alan mavi noktalar gerçek verilerdir. Sarı ile gösterilen noktalar tahmin edilen değerlerdir. Kırmızı ile belirtilen çizgiler gerçek veri ile tahmin edilen verinin hata noktalarıdır. 30 Nisan 2021 tarihinde olan dolar kuru 8,19'dur. Eğitimi yapılmış ve tahmin edilen veri 8,123'dür. Gerçek veri ile arasında 0,67 değer farkı bulunmaktadır. Bu eğitim verisinde elde edilen veriler tablo 3'de yer almaktadır.

**Kayıt Sayısı**

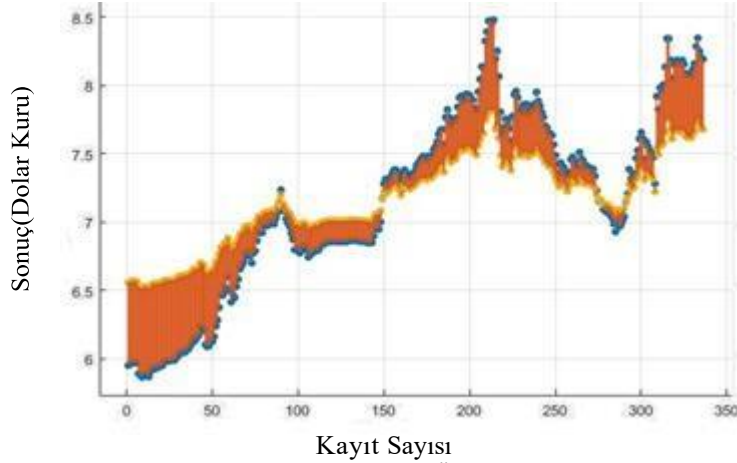
**Şekil 3.** Destek Vektör Makine Öğrenimi Grafiği

Şekil 3’de yer alan grafikte destek vektör makine öğrenimi verilerinin eğitimi sonuçları ve hataları belirtilmiştir. Bu eğitim verisinde elde edilen veriler tablo’ 3 de verilmiştir.

**Tablo 3.** Destek Vektör Makine Analizi.

Kök Ortalama Kare Hatası	R-Kare	Ortalama Kare Hatası	Ortalama Mutlak Hata
0,64769	0,99	0,41951	0,6418

Gauss makine öğreniminde şekil 4’teki grafikte yer alan mavi noktalar gerçek verilerdir. Sarı ile gösterilen noktalar tahmin edilen değerlerdir. Kırmızı ile belirtilen çizgiler gerçek veri ile tahmin edilen verinin hata noktalarıdır. 30 Nisan 2021 tarihinde olan dolar kuru 8,19’dur. Eğitimi yapılmış ve tahmin edilen veri 7,676’dır. Gerçek veri ile arasında 0,514 değer farkı bulunmaktadır. Bu eğitim verisinde elde edilen veriler tablo 4 de yer almaktadır.

**Şekil 4.** Gauss Makine Öğrenimi Grafiği.

Şekil 4’deki grafikte destek vektör makine öğrenimi verilerinin eğitimi sonuçları ve hataları belirtilmiştir. Bu eğitim verisinde elde edilen veriler tablo 4’ te yer almaktadır.

**Tablo 4.** Gauss Makine Öğrenimi Analizi.

Kök Ortalama Kare Hatası	R-Kare	Ortalama Kare Hatası	Ortalama Mutlak Hata
0,33726	0,75	0,1135	0,28072

Oluşturulan modellerde tahmin çalışması yaparken kullanılan yöntemlerde verilerin tahmini sonuçlarının ve hata oranlarının ölçümleri yapılmıştır. Bu analizler sonucunda çıkan değerlere göre kullanılan modellerde gerçek değerlerin ve tahmini değerlerin arasındaki sapma oranının ne kadar olacağını hesaplanması amaçlanmıştır. Çalışmada yer alan dört regresyon algoritması kullanılarak sınıflandırma işlemi yapılmıştır. Dört modelin karşılaştırma sonucu bulgular kısmında belirtilmiştir.

## 5. Veri Seti ve Bulgular

Oluşumu belirlenmiş ağ yapısı eğitilerek yapay zekâ uygulamalarında kullanılır. Sahip olunan veriler; eğitimi verisi ve test verisi şeklinde iki kısımdan oluşur. Yapılan çalışmada yıllık bazda veriler

kullanılmıştır. Yıllık veri seti oluşumunda dolar kuru olarak 341 ve dolar kurunu etkileyen döviz kuru faiz oranı 341 veri ile birlikte toplam 682 veri eğitim setinde kullanılmıştır. Çalışmamızda özellikle durağan dönem seçilmiştir. Zaman serilerinde durağan zaman ve durağan olmaya zaman serilerinin ayırımı önem taşımaktadır. Bu iki durum için farklı analiz yöntemleri kullanılmaktadır [12]. Çalışmamızdaki uygulama pandemi dönemi nedeni ile durağan dönem olarak nitelendirilmiş olup üzerinde durulan konu dönemlik dalgalanmaların sonucudur.

**Tablo 5.** Tahmin Modellerinin Karşılaştırılması.

Fonksiyon/Model	Karar Ağaçları Analizi	Doğrusal Adım Analizi	Destek Vektör Analizi	Gauss Analizi
<b>KOKH</b>	0,1522	0,66752	0,64769	0,33726
<b>R-Kare</b>	1,00	0,20	0,99	0,75
<b>OKH</b>	0,00023164	0,44558	0,41951	0,1135
<b>OMH</b>	0,0097679	0,5407	0,6418	0,28072

Çalışmamızda yer alan parite verileri ve tahmini değerlerin karşılaştırılmasında en uygun model karar ağaçları makine öğreniminde geri yayılım analizi modelidir. Gerçek veri değeri 8,190'dır. Tahmini değer ise 8,203'tür. Doğruluk oranı %99,84 ile en yakın model olduğu gözlemlenmiştir. En yakın ikinci model ise de destek vektör makine öğrenimidir. Tahmin edilen değer 8,123'tür. Gerçek değer ile arasında 0,514 oranında sapma değeri vardır. Doğruluk oranı %99,18'dir. Üçüncü model gauss makine öğrenimidir. Tahmin edilen değer 7,676'dır. Gerçek veri ile arasında 0,514 değer farkı bulunmaktadır. Bu eğitim verisinde elde edilen veriler tablo 6'da yer almaktadır. Doğruluk Oranı %93,72'dir. Dördüncü model ise doğrusal makine öğrenim algoritmasıdır. Tahmin edilen değer 7,112'dir. Gerçek veri ile arasında 1,078 değer farkı bulunmaktadır. Bu eğitim verisinde elde edilen veriler tablo 5'de yer almaktadır. Doğruluk oranı %86,83'tür. Doğruluk oranı bakımından en yüksek değere yaklaşan karar ağaçları tekniğidir.

**Tablo 6.** Fonksiyon Modellerinin Karşılaştırılması.

Fonksiyon/Değerler	Gerçek Değer	Tahmini Değer	Doğruluk Oranı
<b>Karar Ağaçları Analizi</b>	8,190	8,203	%99,84
<b>Destek Vektör Analizi</b>	8,190	8,123	%99,18
<b>Gauss Analizi</b>	8,190	7,676	%93,72
<b>Doğrusal Adım Analizi</b>	8,190	7,112	%86,83

## 6. Tartışma ve Sonuç

Nihal Ata, Erençül Özkök ve Uğur Karabey tarafından yapılan çalışmada çalışmada yaşam çözümlemesi yöntemleri veri madenciliği konusu çerçevesinde ele alındıktan sonra kredi kartı sahiplerine ait bir veri kümesi için yaşam olasılıkları, hazard olasılıkları ve regresyon modelleri incelenmiştir. Uygulamada öncelikle yaşam olasılıkları 5'er yıllık 3 dönem(5,10,15) için elde edilmiştir. Daha sonra, yaşam ve hazard olasılıklarına ait grafikler verilmiş ve müşteri kaybı açısından yorumlanmıştır. Müşterilerin kredi kartını kullanmayı bırakmasını etkileyen risk faktörleri ise regresyon modelleri ile belirlenmeye çalışılmıştır[7].

H. Ali Ata ve İbrahim H. Seyrek tarafından yapılan çalışmada İMKB'de işlem gören imalat sektöründe faaliyet gösteren firma bilgileri incelenmiştir. Karar ağacı modeli ile sınıflandırma işlemi olarak dolandırıcılığın yol açtığı zararları önlemek amacı ile mali tablolardaki sahtekarlığı tespit etmek için çalışma yapılmıştır. Yapılan çalışmanın sonuçları tablo 7'de yer almaktadır[8].

**Tablo 7.** Test Analizleri Sonucu.

Hisse Senetleri	Nano Farad	Farad	T-Değeri	Değerler
LGDT	7,69	7,71	-0,135	0,893
D/E	0,5	0,22	0,074	0,941
TIER	3,62	0,89	5,149	0,000
TD/TA	0,29	0,71	-5,084	0,000
I/S	0,13	0,24	-2,713	0,009
I/TA	0,12	0,18	-3,050	0,003
GP/NS	0,27	0,11	4,990	0,000
NP/TA	0,09	-0,06	6,755	0,000
LTD/TA	0,3	0,31	-0,212	0,833
NP/NS	0,11	-0,14	4,149	0,000
LGTA	8,28	7,94	2,728	0,008
WC/TA	0,47	0,46	0,355	0,723
CA/CL	3,31	1,22	5,754	0,000
C/CL	1,18	0,14	4,435	0,000
OP/NS	0,14	-0,11	5,901	0,000
QA/CL	2,50	0,75	6,110	0,000
TRE	1,99	3,67	-0,576	0,567
TRA	1,26	0,96	1,496	0,140
TRAR	158,62	13,40	1,103	0,276
TRI	15,50	6,89	2,738	0,008
AR/S	0,14	894,15	-1,000	0,322
T/S	-0,03	-0,03	0,016	0,987
GP/TA	0,25	0,12	5,351	0,000
IE/OE	0,09	0,33	-3,601	0,001

Yukarıda bahsedilen seçilmiş değişkenleri kullanarak, karar ağaçlarına ve sinir ağlarına dayalı iki sınıflandırma modeli geliştirilmiştir. Her iki yöntemde de rastgele seçilen vakaların neredeyse yarısı (100 üzerinden 47) model eğitimi için kullanılmış ve kalan vakalar (100 üzerinden 53) model doğrulaması için kullanılmıştır. Eğitim aşamasında, ağaç vakaların %95,74'ünü doğru, %4,26'sını yanlış sınıflandırmaktadır. Model, test verileri üzerinde iyi bir performans göstermektedir. Oluşturulan karar ağacı 53 vakadan 36'sını doğru sınıflandırıyor, bu da %67,92 doğru sınıflandırma anlamına geliyor[8].

Erhan Örüç tarafından yapılan çalışmanın temel amacı Türkiye'de faiz oranları ile enflasyon arasındaki ilişkiyi ifade eden Fisher hipotezinin geçerliliğini test etmektir. Fisher'e göre ekonomideki nominal faiz oranları iki unsura bağlıdır. Birincisi reel faiz oranları; diğeri ise enflasyon oranlarıdır [21]. Tablo 8 'de çalışma ilgili veriler yer almaktadır.

**Tablo 8.** Dikey Fuller Testi[21].

Lag	Y	S	S&T	Y	S	S&T
0	-1.03761	-1.79562	-3.15232	-2.15321	-3.59906	-1.60654
1	-0.94326	-1.58536	-3.03065	-2.03906	-3.37638	-1.60251
2	-0.79766	-1.20507	-2.76804	-1.87692	-3.08595	-1.56948
3	-0.78275	-1.17734	-2.84521	-2.0351	-3.51012	-1.60732
4	-0.74305	-0.87331	-2.57099	-1.14137	-2.33253	-1.19283
5	-0.70659	-0.5674	-2.42824	-1.18717	-2.39225	-1.22088



6	-0.68904	-0.46258	-2.50112	-1.1854	-2.1463	-1.34669
7	-0.91923	-0.40348	-1.93309	-1.36946	-2.49084	-1.37273
8	-0.91654	-0.42594	-1.96364	-0.90472	-2.23945	-1.08094
9	-0.91302	-0.44446	-1.99689	-0.78375	-2.12306	-1.05505
10	-0.89950	-0.46633	-2.05698	-0.77283	-1.92789	-1.14532
11	-0.90577	-0.4835	-2.07559	-0.35296	-1.58469	-1.13603
12	-0.90166	-0.58429	-2.1629	-0.36801	-1.73969	-1.041

Yapılan analizler çerçevesinde faiz oranları ve enflasyon serilerinin birim kök içerdiği bulunmuş ve eş bütünleşme testi konucunda iki değişken arasında uzun dönemli ilişkinin var olduğu sonucuna varılmıştır.

Enflasyonun nominal faiz oranlarını teorik olarak beklenildiği gibi pozitif ve oldukça yüksek etkilediği bulunmuştur. Bir başka deyişle Fisher etkisi Türkiye için geçerlidir. Fakat tam Fisher etkisi gözlenmemiştir[21].

Laurids Gert Nielsen (2018) tarafından döviz kuru için makine öğrenimi üzerine bir uygulama ile çalışma yapılmıştır[19]. Farklı modeller için birçok teorik özelliği tartıştıktan sonra, şimdi modellerin FX verileri üzerindeki performansı değerlendirilmiştir. Dört döviz çifti EUR/USD, GBP/USD, EUR/CHF ve EUR/SWE için tahmin değerinin performansı değerlendirilmiştir. Tüm modeller dikkate alınmış ancak doğrusal regresyon yöntemlerinden elde edilen sonuçlar sunulmayacaktır, çünkü her durumda önemsiz bir hatanın yenilenme olduğu bulunmuştur, bu da onların en küçük kareler yönteminden farklı olmayan tahminler yaptıkları anlamına gelmektedir[19]. Tablo 9 'da çalışma ile ilgili veriler yer almaktadır.

**Tablo 9.** R<sup>2</sup> Model Analizi.

	EUR/USD	GBP/USD	EURCHF	EUR/SWE
<b>OLS</b>	0.541	0.403	1.04	1.202
<b>XGB</b>	0.523	0.257	0.90	1.085
<b>RF</b>	0.542	0.289	1.09	1.576
<b>MARS</b>	0.503	0.113	0.97	1.243
<b>LSTM</b>	0.549	0.367	0.89	1.329

Piyasaların verimli olmaya yakın olduğuna inanırsak, herhangi bir yanlış fiyatlandırmanın hızla düzeltilmesini bekleyeceğimiz için, uzun bir ufuk öngörmenin kısa bir ufuktan daha zor olması şartı bir sonuç olmadığı belirtilmiştir[19].

Hamis Miraji Ally Simba tarafından Türkiye'de 2002-2020 dönemi için faiz oranı ve döviz kuru arasındaki ilişkiyi tespit etmek amacıyla; hem sabit hem de durağan olmayan zaman serisi verilerini kontrol etmek için Dickey Fuller (ADF) testi, eş bütünleşmenin varlığını kontrol etmek için Johansen eş bütünleşme testine başvurulmuştur. devamında hata düzeltme için vektörel hata düzeltme (VECM) testi ve Nedenselliği kontrol etmek için de klasik Granger nedensellik testi kullanılmıştır[20].

Granger ve Newbold (1974) durağan olmayan zaman serileri ile yapılan analizlerde sahte regresyon problemi ortaya çıkabileceğini kanıtlamıştır. Bu nedenle zaman serileri ile yapılan analizlerde kullanılan serilerin durağanlığı test edilmeye başlanmıştır[20]. Tablo 10 'da çalışma ile ilgili veriler yer almaktadır.

**Tablo 10.** Dickey Fuller (ADF) Birim Kök Testi [20].

Değişkenler	Seviye I (0) Trend ve Kesişme		İlk fark (1) Trend ve Kesişme	
	T- İstatistiği	Olasılık	T- İstatistiği	Olasılık
<b>Faiz Oranı</b>	-3.367	0.059	-9.709	0.0000
<b>Döviz Kuru</b>	4.156	1.000	-8.205	0.0000

Görüldüğü gibi geleceği tahmin eden modellemelerin kullanımında, yapay sinir ağları başarılı bir şekilde kullanılmaktadır. Bu yapı veriler arasındaki dikkat edilmesi zor bağlantıları gösterebilir. Yapılan bu çalışmada döviz kuru yapay zeka ağ yapısının tasarımı ile öngörülmüştür. Tahmin edilen modeller kendi içinde ve tahmin gücü diğer yüksek modeller ile karşılaştırılarak verim analizi yapılmıştır. Döviz kuru model tahmini sürecinde her veri noktasında geri yayılım algoritması ile oluşan hatanın asgari ölçüde olması sağlanmıştır. Problem aşamasına göre kurulmuş ağ yapısı iyi sonuçlar verir. Bu sebeple tahmin edilecek veriye göre uygun ağ yapıları ve modeller araştırılmalıdır. Döviz kurlarında tahmin çalışması yaparken kullanılan yöntemlerin sağlaması gereken varsayım serilerinin durağan olması esas alınmıştır. Literatürdeki çalışmalarda döviz kurlarının durağanlık analizi yapılırken genel olarak serilerin değerlerinin alınarak birim kök testlerinin uygulandığı klasik metodoloji bulunmaktadır. Bu yöntem döviz kuru değerlerinin tahminini kolaylaştırmaktadır. 2001 yılında serbest dalgalanan kur rejimine geçildikten sonra döviz kurlarından fiyatlara geçiş etkisinin azaldığı görülmüştür. T.C Merkez Bankası serbest dalgalanan kur düzenine geçtikten sonra döviz kurlarında aşırı değişim haricinde müdahale etmeyeceğini ve döviz kurlarında ilişkin hedefi olmadığını belirtmiştir. Fakat son dönemlerde enflasyon oranının yükselmesi nedeni ile döviz kuru oranlarını düşürmek için T.C Merkez Bankası müdahalelerde bulunarak kur oranlarına etki etmiştir. Çalışmamızda özellikle durağan dönem seçilmiştir. Durağan ve durağan olmayan şeklindeki ayırım zaman serisi analizinde büyük önem taşımaktadır.

Bunun nedeni, zaman serisi analizleri için geliştirilmiş olan olasılık teorilerinin sadece durağan zaman serileri için geçerli olmasıdır.

Zaman serilerinde durağan olmama durumu genelde zamanla değişen ortalama ve zamanla değişen varyans olmak üzere iki şekilde ortaya çıkmaktadır [12]. Çalışmamızdaki uygulama pandemi dönemi nedeni ile durağan dönem olarak nitelendirilmiş olup üzerinde durulan konu dönemlik dalgalanmaların sonucudur.

Çalışmadan elde edilen sonuçlara bakıldığında döviz kuru ve sapmaların görülmesinde yaşanan krizlerin tahmini değerlerin öngörülmesinde önemli bir gösterge olabileceği söylenebilir. Yapılan benzer çalışmalarda yer alan çalışmalar ile ilgili bilgilere çalışmamızda yer verilmiştir. Bu bilgiler yazarların çalışmalarında uyguladıkları yöntemleri ve sonuçları ifade etmektedir. Bu çalışmanın önemi, döviz kuru ile ilgilenen kişilerin ve kurumların kur tahmin yöntemlerini seçerken karşılaştırma yapmasını sağlayarak, hangi kur tahmin yöntemin kendisine uygun olduğunu belirlemeye çalışmasıdır.

## 7. Referanslar

- [1]. Y. Zhao, Y. Chen ve Y. Yao, «User-centered Interactive Data Mining,» International Conference on Cognitive Informatics, no. 6, pp. 2006, 457-466
- [2]. U. Fayyad, G. P. Shapıra ve P. Smyth, «Veri Madenciliğinden Veritabanlarında Bilgi Keşfine,» AI Magazine, pp.03-15, 1996.
- [3]. J. Han, K. Micheline ve P. Jian, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2012.
- [4]. M. B. a. G. Linoff, Introduction to Data Mining and Knowledge Discovery, Third Edition, USA: Two Crows Corporation, 2005.
- [5]. H. Akpınar, «Veri Tabanlarında Bilgi Keşfi Ve Veri Madenciliği,» İstanbul Üniversitesi İşletme Fakültesi Dergisi, pp. 1 22, 2000.
- [6]. C. Shearer, The Crisp-DM Model: The New Blueprint for Data Mining”, Seattle: Journal of data warehousing, 2000.
- [7]. N. Ö. E. v. K. U. Ata, «Survival Data Mining : An Application To Credit Card Holders,» Sigma Mühendislik ve Fen Bilimleri Dergisi, cilt 26, no. 1, pp. 33-42, 2008.
- [8]. A. v. S. İ. Ata, «The Use of Data Mining Techniques in,» Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, cilt 14, no. 2, pp. 157-170, 2009.
- [9]. «Veri Bilimcisi,» <https://veribilimcisi.com/>.
- [10]. «Corporate Finance Institute,» <https://corporatefinanceinstitute.com/>.
- [11]. P. J. Bickel ve K. A. Doksum, Mathematical Statistics, San Diego: Chapman and Hall/CRC, 2015.
- [12]. K. İlarıslan ve R. Aşıkođlu, «Birleşme Ve Satın Almaların Finansla Oranlar Yoluyla İşletmelerin Finansal Performansı Üzerindeki Etkisinin İncelenmesi,» Haziran 2011. [https://acikbilim.yok.gov.tr/bitstream/handle/20.500.12812/26971/yokAcikBilim\\_406171.pdf?sequence=-1](https://acikbilim.yok.gov.tr/bitstream/handle/20.500.12812/26971/yokAcikBilim_406171.pdf?sequence=-1).
- [13]. G. Gözğör, «Döviz Piyasalarında Etkinlik ve Beklentiler Kavramları Üzerine Bir İnceleme,» İktisadi Yenilik Dergisi, cilt 3, no. 2, 2016
- [14]. N. Rahpeymai, Data Mining with Decision Trees in the Gene Logic, Skövde-İsveç: Masters Dissertation in bioinformatics, 2002.
- [15]. G. Rodriguez, «Generalized Linear Models,» Princeton University, Princeton, 2007.
- [16]. T. Hastie, R. Tibshirani ve J. Friedman, The Elements of Statistical Learning, New York: Springer, 2008.
- [17]. L. Murray, «Introduction to Gaussian Processes,» Computer Science, University of Toronto, Toronto, 2008
- [18]. J. S. Armstrong, «Illusions in Regression Analysis,» International Journal of Forecasting, pp. 1-10, 2019]
- L.G. Nielsen, Machine Learning For Foreign, London: Department of Mathematics Imperial College London, 2018
- [20]. H. M. A. Simba, «Faiz ve Döviz Kuru İlişkisi : Türkiye Üzerine Bir Değerlendirme,» International Conference of

- Strategic, Viyana, 2020
- [21]. E.Örüş, “Fisher Etkisi : Türkiye üzerine bir uygulama”, Kastamonu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,cilt 13, no.3,pp.297-311,2016.