

Araç Satın Alma Eğiliminin Yapay Zeka Yöntemleri ile Tahmin Edilmesi

M. Hanefi CALP^{1*}

¹*Yönetim Bilişim Sistemleri, İİBF, Ankara Hacı Bayram Veli Üniversitesi, Türkiye*

**hanefi.calp@hbv.edu.tr*

Özet – Bu çalışmada, müşterilerden alınan yaş, cinsiyet ve tahmini kazanç değerleri dikkate alınarak kişinin araç satın alma durumunun yapay zeka teknikleri ile tahmin edilmesi amaçlanmıştır. Bu bağlamda, yapay zeka tekniklerinden olan Lojistik Regresyon, Destek Vektör Makineleri ve Naive Bayes'ten yararlanılmıştır. Veri seti, Kaggle web sitesi üzerinden temin edilmiştir. Çalışma kapsamında veri seti farklı alternatiflerle eğitim ve test kümesi olarak ikiye ayrılmıştır. En doğru tahmini gerçekleştirmek adına test_size parametresi 30, 25, 20, 15 olarak belirlenmiş olup her bir parametre birçok kez test edilmiştir. Veri setini daha ayrıntılı bir şekilde açıklamak adına subplot ve scatter plot grafikleri kullanılmıştır. Yapılan analizler, grafik oluşturma ve tahminler google colab üzerinde gerçekleştirilmiştir. Sonuçlara göre en başarılı performansın Destek Vektör Makineleri yöntemi ile elde edildiği ortaya çıkmıştır.

Anahtar Kelimeler – *Yapay Zeka, Araç Satın Alma, Lojistik Regresyon, Destek Vektör Makineleri Ve Naive Bayes*

I. GİRİŞ

Yapay zeka (YZ), insanlar tarafından gerçekleştirilen davranışların makineler tarafından yapılabilmesidir. YZ, insan gibi düşünebilen ve insan beyninin çalışma mantığını icra edebilen bir yöntemdir. Bu yöntem, kendi kendine öğrenebilen özelliği sayesinde, kendisine girilen verileri kullanarak çıkarım yapar. Veri, YZ yöntemlerinde oldukça önemlidir. Veri ne kadar çoksa YZ o kadar doğru sonuçlar üretir. Sözkonusu veriler; gözlem, sensör, sosyal medya, araştırma, internet gibi birçok farklı kaynaktan elde edilebilmektedir. YZ ve verinin önemi engellenemez bir şekilde artmaktadır. YZ'nin, veri ile ilişkilendirilmesi durumunda çeşitli analiz ve tahminler yapılabilir. Günümüzde birçok çalışmada da çeşitli YZ yöntemleri kullanılarak gerçeğe yakın sonuçlar elde edilmiştir [1,2].

Bu çerçevede, literatürde birçok çalışma yapılmıştır. Calp (2020), orman yangınlarında yanmış alanların tahmin edilmesine yönelik yapay sinir ağına dayalı bir model geliştirmiştir. Yangınların ne zaman, nerede ve ne kadar çıkacağı bilinmediği için çalışma önemlidir. Çalışmada

tahmin etmek için yapay sinir ağı yöntemi kullanılmıştır. Bu noktada, ileri beslemeli bir geri yayılım sinir ağı kullanılmıştır. Önerilen model üzerinde Regresyon değerleri, Ortalama Mutlak Yüzde Hata (MAPE) ve Ortalama Kare Hata (MSE) dikkate alınarak bir performans değerlendirmesi gerçekleştirilmiştir. Sonuçlar, yapay sinir ağı modelinin düşük hata oranına ve yüksek tahmin doğruluğuna sahip olduğunu ortaya koymuştur [3].

Yürük (2019), BIST imalat sektörü uygulaması üzerinden işletmelerin finansal başarısızlığını tahmin eden bir model önermiştir. Çalışmada yapay zekâ tekniklerinden olan yapay sinir ağları ve destek vektör makinalarından yararlanmıştır. Ayrıca kullanılan modellerin performansı ROC eğrileri ile test edilmiştir. Çalışmada iki modelin performansları karşılaştırılmış, yapılan sınıflandırmanın doğruluk yüzdeleri hesaplanarak analiz edilmiştir [4].

Kılıç (2019), yapay sinir ağları ve regresyon yöntemlerini kullanarak geri esneme tahminini gerçekleştirmiştir. YZ yöntemi ile dışarıdan girilen veriler, sistemin durumunu tahmin etmede önemli

bir etkidir. Çalışmada, yöntem olarak Yapay Sinir Ağları (YSA) ve Regresyon yöntemleri kullanılmıştır. Farklı parametrelerin sonuçlara etkisini anlayabilmek amacıyla oluşturulan farklı YSA modelleri kıyaslanmış ve analiz edilmiştir. Sonuçlar, oluşturulan YSA modelleri için uygun parametre seçiminin çok önemli olduğunu ortaya koymuştur. Ayrıca, YSA yöntemi ile oluşturulan modelin, Regresyon modelden daha başarılı sonuçlar ürettiği görülmüştür [5].

Atalay (2017), büyük veri analizinde yapay zeka ve makine öğrenmesi uygulamaları konusu incelenmiştir. Öncelikle YZ ve makine öğrenmesi yöntemleri ile ilgili birtakım bilgiler verilmiştir. Daha sonra, sözkonusu yöntemlerin büyük veri kullanılarak gerçekleştirilen uygulamalarına ait bazı örnekler yer verilmiştir. Özetle; sınıflandırma, yapay sinir ağları, kümeleme, metin ve web madenciliği ve duygu analizi konularında büyük verilerle yapılan çalışmalar analiz edilmiş ve değerlendirilmiştir [6].

Ünsal (2018), YZ ve kümeleme tekniklerini kullanarak okul servisi rotalama probleminin optimizasyonu gerçekleştirmiştir. Çalışmada, Araç Rotalama Problemi (ARP)'nin alt başlığı olarak bilinen Okul Servisi Rotalama Problemini (OSRP) optimize etmeyi amaçlanmıştır. Bu çerçevede, ARP ve OSRP kavramları incelenmiş, problemler ve geliştirilen çözüm yöntemleri ile ilgili literatüre yer verilmiştir. OSRP'yi optimize etmek amacıyla kümeleme ve YZ yöntemleri kullanılmıştır. GPS, GIS araçları ve mobil uygulama tabanlı bir yazılım geliştirilerek sözkonusu yazılım projesi Ankara'da hizmet veren servis firmalarından elde edilen rota verileri üzerinde uygulanmıştır. Sonuç olarak, geliştirilen yöntemin, mesafe, zaman ve rakım değişimi parametreleri açısından başarılı sonuçlar ortaya çıkardığı görülmüştür [7].

Başar (2020), ülkelerin güvenli olup olmamaları üzerine tahmin yapan yapay zeka tabanlı bir model geliştirmiştir. Geliştirdikleri modelde tahmin yapmak amacıyla lojistik regresyon, yapay sinir ağları ve çok kriterli karar verme yöntemlerinden biri olan MOORA yöntemleri kullanılmıştır. Sözkonusu tahmin süreci için 106 ülkeye ait suç endeksi kullanılarak suç oranları ortaya çıkarılmıştır. Çalışma sonucunda yapay sinir ağları yöntemi ile oluşturulan modelin, güvenilirlik tahmini sürecinde yapılan tahminlerin başarı durumu diğer yöntemlere nazaran daha yüksek oranda olduğu tespit edilmiştir [8].

Yılmazel (2018), yapay sinir ağları yöntemi kullanılarak konut fiyat tahmini gerçekleştirmiştir. Çalışmada, satılık daire fiyatlarının tahmini için Eskişehir'de bulunan daireler dikkate alınmıştır. Daire büyüklüğü, oda sayısı, bulunduğu kat (1. Kat özelinde de dikkate alınmıştır.), binadaki toplam kat sayısı, ısıtma durumu (merkezi, kombi vb.), banyo sayısı, asansör durumu, otopark durumu, ankastre mutfak ve fiber internet bağlantısının bulunup bulunmadığı gibi birçok özelliikle birlikte, mahalle ve ulaşım araçlarına mesafesi gibi ölçütler de dikkate alınarak birçok model kurulmuştur. Geliştirilen modellerde gizli katmanda bulunan nöron sayısı farklı sayılarda denenerek 19 tane model üretilmiş ve sözkonusu modellerin performansları karşılaştırılarak analiz edilmiştir. Sonuç olarak yapay sinir ağları yönteminin konut fiyatı tahmininde etkin bir araç olduğu ortaya konulmuştur [9].

Sel (2020), alternatif yatırım araçları ile bitcoin fiyatları arasındaki ilişki yapay sinir ağı kullanılarak tahmin edilmiştir. Araştırmacı çalışmada, yaygın bilinen kripto para çeşitlerinden biri olan Bitcoin'in fiyatını tahmin etmeyi amaçlamıştır. Girdi değişkenler olarak; Nasdaq100, Dow-Jones, S&P500, İsviçre Frangı, İngiliz Sterlini, Eurostoxx Endeksleri, Altın, Euro, Gümüş yatırım araçlarını dikkate almıştır. Ayrıca, veriler 2013-2018 yılları arasındaki günlük kapanış değerleridir. Sonuç olarak, yatırım araçlarının fiyatlarının tahmini noktasında yapay sinir ağları yönteminin etkili olduğunu ortaya koymuştur [10].

II. MATERYAL VE YÖNTEM

A. Çalışmanın Amacı

Bu çalışmada, müşterilerden alınan yaş, cinsiyet ve tahmini kazanç değerleri dikkate alınarak kişinin araç satın alma durumunun YZ teknikleri ile tahmin edilmesi amaçlanmıştır.

B. Uygulama

Bu bölümde, çalışmanın uygulama kısmının tüm ayrıntılarına yer verilmiştir.

Veri seti

Veri seti Kaggle.com üzerinden elde edilmiş olup modelin oluşturulması ve sonuçların analizleri "google colab" uygulaması üzerinde gerçekleştirilmiştir. Çalışmada modellerin oluşturulmasında Lojistik Regresyon, Destek Vektör Makineleri ve Naive Bayes yöntemleri

kullanılmıştır. Tahmin modellerinin oluşturulması sürecinde, her yöntem ayrı ayrı birçok kez test edilmiş olup test kümesi %30, %25, %20 ve %15 olacak şekilde belirlenmiştir. Colab ve jupyterlab ile gerçekleştirilen analizlerde subplot ve scatterplot grafikleri kullanılmıştır. Öncelikle, veri setinde 4 adet integer ve 1 adet object türünde veri tipi bulunmaktadır. Ayrıca veri setinde boş değerler de bulunmaktadır. Kullanılan veri tipleri Şekil-1’ de gösterilmiştir.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   User ID         400 non-null   int64
1   Gender          400 non-null   object
2   Age             400 non-null   int64
3   EstimatedSalary 400 non-null   int64
4   Purchased       400 non-null   int64
dtypes: int64(4), object(1)
memory usage: 15.8+ KB
```

Şekil 1. Veri tipleri

Veri setini genel hatlarıyla inceleyecek olursak, 400 satırlık veri bulunmaktadır. Veri setinde bulunan müşterilerin ortalama yaş değeri 37.6, en küçük yaş değeri 18 ve en büyük yaş değeri 60 olarak görülmektedir. Ortalama yıllık kazanç değeri ise \$69.742, en düşük kazanç değeri \$15.000 ve en yüksek kazanç değeri ise \$150.000 olarak görülmektedir. Değerler detaylı bir şekilde Tablo 1’ de gösterilmiştir.

Tablo 1. Veri setinin genel özeti

	User ID	Age	EstimatedSalary	Purchased
count	400.000000	400.000000	400.000000	400.000000
mean	200.500000	37.655000	69742.500000	0.357500
std	115.614301	10.482877	34096.960282	0.479864
min	1.000000	18.000000	15000.000000	0.000000
25%	100.750000	29.750000	43000.000000	0.000000
50%	200.500000	37.000000	70000.000000	0.000000
75%	300.250000	46.000000	88000.000000	1.000000
max	400.000000	60.000000	150000.000000	1.000000

Bu değerler cinsiyete göre incelendiğinde ise erkek müşterilerin ortalama yaş değeri 36.8, en küçük yaş değeri 18 ve en büyük yaş değeri 60 olarak görülmektedir. Ortalama yıllık kazanç değeri ise \$67.642, en düşük kazanç değeri \$15.000 ve en yüksek kazanç değeri ise \$150.000 olarak görülmektedir. Değerler detaylı bir şekilde Tablo 2’de gösterilmiştir.

Tablo 2. Cinsiyete göre veri setinin özeti

	User ID	Age	EstimatedSalary	Purchased
count	196.000000	196.000000	196.000000	196.000000
mean	195.520408	36.867347	67642.857143	0.336735
std	117.178779	10.043636	32421.819181	0.473804
min	1.000000	18.000000	15000.000000	0.000000
25%	98.750000	29.000000	43000.000000	0.000000
50%	190.500000	36.500000	68000.000000	0.000000
75%	297.500000	42.000000	86000.000000	1.000000
max	399.000000	60.000000	150000.000000	1.000000

Kadın müşterilerde ise ortalama yaş değeri 38.4, en küçük yaş değeri 18 ve en büyük yaş değeri 60 olarak görülmektedir. Ortalama yıllık kazanç değeri ise \$71.759, en düşük kazanç değeri \$15.000 ve en yüksek kazanç değeri ise \$150.000 olarak görülmektedir. Değerler detaylı bir şekilde Tablo 3’te gösterilmiştir.

Tablo 3. Yaşa göre veri setinin özeti

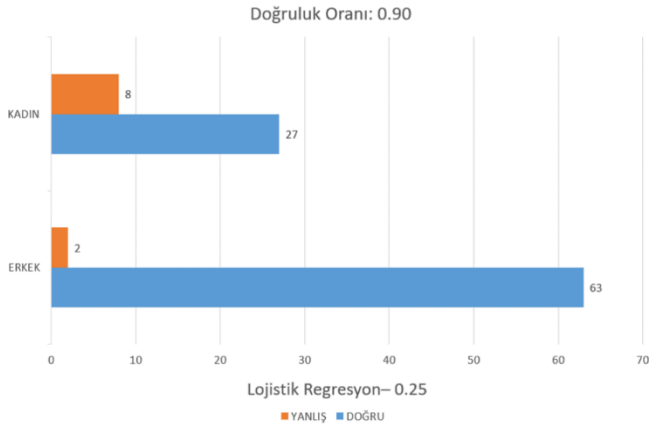
	User ID	Age	EstimatedSalary	Purchased
count	204.000000	204.000000	204.000000	204.000000
mean	205.284314	38.411765	71759.803922	0.377451
std	114.174085	10.859019	35595.236189	0.485942
min	3.000000	18.000000	15000.000000	0.000000
25%	103.750000	30.000000	44000.000000	0.000000
50%	209.500000	37.500000	70500.000000	0.000000
75%	303.500000	47.000000	90750.000000	1.000000
max	400.000000	60.000000	150000.000000	1.000000

C. Modellere Göre Test Sonuçları ve Analizleri

Lojistik Regresyon

Test aşamasında ilk olarak lojistik regresyon yöntemi uygulanmıştır. Test kümesi %30, eğitim kümesi ise %70 olarak ayrıldığında, Doğruluk oranı 0.85; test kümesi %25, eğitim kümesi ise %75 olarak ayrıldığında, Doğruluk oranı 0.90; test kümesi %20, eğitim kümesi ise %80 olarak ayrıldığında, Doğruluk oranı 0.862; test kümesi %15, eğitim kümesi ise %85 olarak ayrıldığında ise Doğruluk oranı 0.89 olarak saptanmıştır.

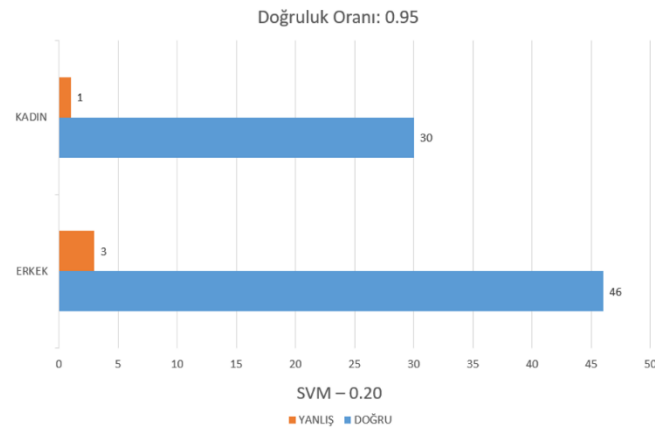
Sonuç olarak, en iyi başarı performansı, test kümesi %25, eğitim kümesi ise %75 olarak ayrıldığında elde edilmiştir (Şekil 2).



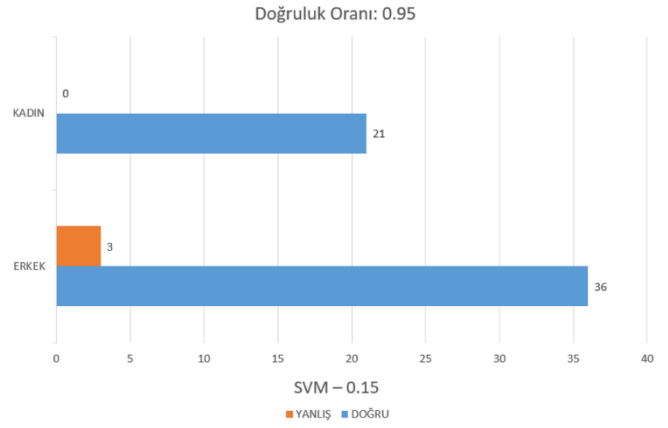
Şekil 2. Lojistik Regresyon modelinin test sonuçları

Destek Vektör Makineleri

Bu aşamada Destek Vektör Makineleri modeli oluşturulmuş ve test edilmiştir. Bu yöntemde, test kümesi %30, eğitim kümesi %70 olarak ayrıldığında Doğruluk oranı 0.941; test kümesi %25, eğitim kümesi ise %75 olarak ayrıldığında Doğruluk oranı 0.93; test kümesi %20, eğitim kümesi ise %80 olarak ayrıldığında Doğruluk oranı 0.95 ve son olarak test kümesi %15, eğitim kümesi ise %85 olarak ayrıldığında ise Doğruluk oranı 0.95 olarak elde edilmiştir. Sonuç olarak, en iyi başarı performansı, test kümesi %20, eğitim kümesi ise %80 ve test kümesi %15, eğitim kümesi ise %85 olarak ayrıldığında elde edilmiştir (Şekil 3, Şekil 4).



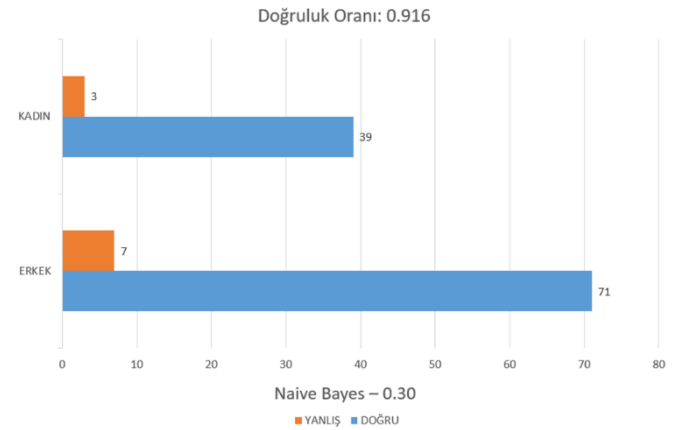
Şekil 3. Destek Vektör Makineleri modelinin test sonuçları-1



Şekil 4. Destek Vektör Makineleri modelinin test sonuçları-2

Naive Bayes

Bu aşamada ise Naive Bayes yöntemi ile model oluşturulmuş ve oluşturulan model test edilmiştir. Test sürecinde; test kümesi %30, eğitim kümesi ise %70 olarak ayrıldığında Doğruluk oranı 0.916; test kümesi %25, eğitim kümesi %75 olarak ayrıldığında Doğruluk oranı 0.89; test kümesi %20, eğitim kümesi ise %80 olarak ayrıldığında Doğruluk oranı ise 0.90; test kümesi %15, eğitim kümesi ise %85 olarak ayrıldığında ise Doğruluk oranı 0.88 olarak elde edilmiştir. Sonuç olarak, en iyi başarı performansı, test kümesi %30, eğitim kümesi ise %70 olarak ayrıldığında elde edilmiştir (Şekil 5).



Şekil 5. Naive Bayes modelinin test sonuçları

Elde edilen tüm sonuçların özeti niteliğindeki veriler karşılaştırmalı olarak Tablo 4'te, verilmiştir.

Tablo 4. Sonuçların karşılaştırılması

Sıra	Model	Başarı Performansı
1	Destek Vektör Makineleri	0,95
2	Lojistik Regresyon	0,90
3	Naive Bayes	0,91

III. SONUÇLAR

Bu çalışmada, Lojistik Regresyon, Destek Vektör Makineleri ve Naive Bayes olmak üzere yapay zeka teknikleri kullanılarak kişinin araç satın alma durumu tahmin edilmiştir. Tahmin sürecinde; müşterilerden alınan yaş, cinsiyet ve tahmini kazanç değerleri dikkate alınmıştır. Elde edilen sonuçlar, Destek Vektör Makineleri yönteminin diğer yöntemlere göre daha başarılı sonuçlar ürettiğini ortaya koymuştur.

Bununla birlikte, veriler ışığında tahmin sonuçlarında belirli noktalarda farklılıklar gözlemlenmiştir. Kadın müşterilerin ortalama yıllık geliri erkek müşterilere göre daha fazla olmasına rağmen, erkek müşterilerin araç satın alma eğilimi kadın müşterilere oranla daha fazla olduğu saptanmıştır. Bu noktada, yıllık gelirin araç satın alma eğilimine etkisinin düşük olduğu kanısına varılmıştır. Yaş ve cinsiyet gibi kriterler satın almada daha etkin rol oynamaktadır.

Veri setinden elde edilen sonuçlara göre, günümüz yaşamında erkeklerin arabalara olan ilgisinin kadınlara göre daha fazla olması tezinin doğruluğu kısmen teyit edilmiştir. Bu çalışmadan yola çıkılarak otomobil firmaları araç üretimlerinde tasarımlarını erkek müşterilerin zevk ve dönütlerine göre gerçekleştirmeleri araç satışlarında artış yaşanacağı söylenebilir.

KAYNAKLAR

- [1] Calp, M. H. (2019). Medical diagnosis with a novel SVM-CoDOA based hybrid approach. arXiv preprint arXiv:1902.00685.

- [2] Ertel, W. (2018). Introduction to Artificial Intelligence. Springer
- [3] Calp, M. H., & Kose, U. (2020). Estimation of burned areas in forest fires using artificial neural networks. *Ingeniería Solidaria*, 16(3), 1-22.
- [4] Yürük, M., Ekşi, İ. (2019). Yapay Zeka Yöntemleri ile İşletmelerin Finansal Başarısızlığının Tahmin Edilmesi: Bist İmalat Sektörü Uygulaması. *Mukaddime*, 10 (1), 393-422. DOI: 10.19059/mukaddime.533151
- [5] Kılıç, S. (2019). Yapay Sinir Ağları ve Regresyon Yöntemleri ile Geri Esneme Tahmini. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, 6 (2), 433-446. DOI: 10.35193/bseufbd.620010
- [6] Atalay, M, Çelik, E. (2017). Büyük Veri Analizinde Yapay Zekâ Ve Makine Öğrenmesi Uygulamaları - Artificial Intelligence and Machine Learning Applications in Big Data Analysis. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9 (22), 155-172. DOI: 10.20875/makusobed.309727
- [7] Ünsal, Ö, Yiğit, T. (2018). Yapay Zeka Ve Kümeleme Teknikleri Kullanılarak Geliştirilen Yöntem İle Okul Servisi Rotalama Probleminin Optimizasyonu. *Mühendislik Bilimleri ve Tasarım Dergisi*, 6 (1), 7-20. DOI: 10.21923/jesd.340220
- [8] Deniz Başar, Ö, Güneren Genç, E. (2020). Ülkelerin Güvenli Olmalarının Tahmininde Lojistik Regresyon, Yapay Sinir Ağları Ve Moora Yöntemlerinin Karşılaştırılması. *Journal of Life Economics*, 7 (2), 123-134. DOI: 10.15637/jlecon.7.008
- [9] Yılmazel, Ö, Afşar, A, Yılmazel, S. (2018). Konut Fiyat Tahmininde Yapay Sinir Ağları Yönteminin Kullanılması. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, (20), 285-300. DOI: 10.18092/ulikidince.341584
- [10] Sel, A., Zengin, N, Yıldız, Z. (2020). Alternatif Yatırım Araçları İle Bitcoin Fiyatları Arasındaki İlişkinin Yapay Sinir Ağı ile Tahmini. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 21 (2), 157-169. DOI: 10.37880/cumuiibf.616127.