



T.C.

**ÇANAKKALE ONSEKİZ MART ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**MAKİNE ÖĞRENME TEMELLİ KONUT FİYAT TAHMİNLEMEDE META
MODEL UYGULAMASI: ÇANAKKALE İLİ ÖRNEĞİ**

YÜKSEK LİSANS TEZİ

CENGİZHAN DUMLU

Tez Danışmanı

DR. ÖĞR. ÜYESİ SAİT CAN YÜCEBAŞ

ÇANAKKALE – 2022



T.C.

ÇANAKKALE ONSEKİZ MART ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**MAKİNE ÖĞRENME TEMELLİ KONUT FİYAT TAHMİNLEMEDE META
MODEL UYGULAMASI: ÇANAKKALE İLİ ÖRNEĞİ**

YÜKSEK LİSANS TEZİ

CENGİZHAN DUMLU

Tez Danışmanı

DR. ÖĞR. ÜYESİ SAİT CAN YÜCEBAŞ

ÇANAKKALE – 2022

ETİK BEYAN

Çanakkale Onsekiz Mart Üniversitesi Lisansüstü Eğitim Enstitüsü Tez Yazım Kuralları'na uygun olarak hazırladığım bu tez çalışmada; tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi, tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu, tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi, kullanılan verilerde herhangi bir değişiklik yapmadığımı, bu tezde sunduğum çalışmanın özgün olduğunu, bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi taahhüt ve beyan ederim.

Cengizhan DUMLU

18/07/2022

TEŐEKKÜR

Bu tezin gerekleŐtirilmesinde, üstün bilgi birikimiyle alıŐmama farklı aılardan bakmamı saėlayan ve alıŐmam boyunca her türlü yol gösterici olan, pozitif tavrıyla beni cesaretlendiren, gece gündüz farketmeksizin benden bir an olsun yardımlarını esirgemeyen, beraber alıŐmaktan ve her zaman öėrencisi olmaktan gurur ve mutluluk duyduğum saygı deėer danışman hocam Dr. Öğr. Üyesi Sait Can YÜCEBAŐ'a sonsuz teŐekkür ederim.

Deėerli görüşleri ile araŐtırmamın Őekillenmesini saėlayıp tezime farklı bakıŐ aılları kazandıran, alıŐmam boyunca kıymetli vaktini ayırarak desteėini hiçbir Őekilde esirgemeyen deėerli bilim insanı Prof. Dr. Levent GEN'e teŐekkürü bir bor bilirim.

Bütün hayatım boyunca yanımda olan, sadece bu alıŐma için deėil hayatımın her evresinde beni cesaretlendiren ve moral veren kıymetli annem Őehrinaz DUMLU ve babam Sait DUMLU'ya sonsuz teŐekkür ederim.

Cengizhan DUMLU
anakkale, Temmuz 2022

ÖZET

MAKİNE ÖĞRENME TEMELLİ KONUT FİYAT TAHMİNLEMEDE META MODEL UYGULAMASI: ÇANAKKALE İLİ ÖRNEĞİ

Cengizhan DURLU

Çanakkale Onsekiz Mart Üniversitesi

Lisansüstü Eğitim Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı Yüksek Lisans Tezi

Danışman: Dr. Öğr. Üyesi Sait Can YÜCEBAŞ

18/07/2022, 40

Makine Öğrenmesi yöntemleri günümüzde birçok alanda olduğu gibi emlak fiyatlarının tahmin edilmesi için de yaygın olarak kullanılmaktadır. Gerçekleştirilen çalışmada satılık konut fiyatlarının tahmin edilmesi için bir meta model geliştirilmiştir. RapidMiner aracı kullanılarak gerçekleştirilen meta modelde Karar Ağacı, Rassal Orman, Destek Vektör Makinesi ve Yapay Sinir Ağı yöntemleri kullanılmıştır. Çalışmada Çanakkale iline ait satılık konut verileri kullanılmıştır. Bu veriler web kazıyıcı oluşturularak aktif olarak kullanılan bir emlak sitesi üzerinden elde edilmiştir. Modelin tahmin başarısının artırılması için elde edilen bu ham veriler belirli veri ön işleme adımlarından geçirilmiştir. Karar Ağacı, Rassal Orman, Destek Vektör Makinesi ve Yapay Sinir Ağı yöntemleri bir meta model altında birleştirilmiş ve nihai sonuç için Çoğunluk Oylaması kullanılmıştır. Hem meta model hem de meta model altındaki tüm alt modellerin eğitimi ve testi için 10 katmanlı çapraz geçerlilik yöntemi kullanılmıştır. Modellerin en iyi performansı verebilmesi için evrimsel optimizasyon temelli hiper parametre optimizasyonu uygulanmıştır. İlgili optimizasyonun performansa katkısını gösterebilmek adına tüm modeller varsayılan hiper parametre değerleriyle de kurulmuş, optimize edilmiş modeller ile performans karşılaştırması yapılmıştır. Hem alt modeller hem de meta model için optimize edilmiş hiper parametrelerle kurulan modellerin daha yüksek tahmin performansı gösterdiği görülmüştür. Alt modeller içinde en yüksek performansı 0,825 korelasyon sayısı ile Rassal Orman göstermiştir. Kurulan model 0,836'lık korelasyon ile alt modellerin tahmin performansının oldukça üstüne çıkmıştır.

Anahtar Kelimeler: Makine Öğrenmesi, Meta Model, Torbalama, Çoğunluk Oylaması, Optimizasyon, Emlak Fiyat Tahmini



ABSTRACT

META MODEL APPLICATION IN MACHINE LEARNING BASED HOUSE PRICE ESTIMATION: AN EXAMPLE OF CANAKKALE PROVINCE

Cengizhan DUMLU

Canakkale Onsekiz Mart University

Graduate Education Institute

Master of Science Thesis in Computer Engineering

Advisor: Assistant Professor Sait Can YÜCEBAŞ

18/07/2022, 40

Machine Learning methods are widely used for estimating real estate prices as it is used in many other domains. In this study, a meta-model has been developed to predict the prices of houses on sale. Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network (ANN) were used in the meta-model developed by using the RapidMiner tool. The data set consists of the houses on sale in Canakkale province. A web scraper was developed to retrieve the sale data from an actively used real estate web site. In order to increase the prediction performance a data preprocessing step was conducted. A ten-fold cross validation was used to train and test both the base learners and the meta model. Evolutionary based hyper parameter optimization was used to increase the performance of the models. In order to show the contribution of the optimization to the performance, all models were constructed with the default hyper parameter values, and performance comparisons were made with the optimized models. Base learners and meta model with optimized hyper parameters showed higher predictive performance. Random Forest showed the highest performance with a correlation number of 0,825 among the base learners. The established meta model was well above the prediction performance of the sub-models with a correlation of 0,836.

Keywords: Machine Learning, Meta Model, Ensemble, Majority Voting, Optimization, Real Estate Price Prediction

İÇİNDEKİLER

	Sayfa No
ETİK BEYAN	i
TEŞEKKÜR	ii
ÖZET	iii
ABSTRACT	v
İÇİNDEKİLER	vi
SİMGELER ve KISALTMALAR.....	viii
TABLolar DİZİNİ	ix
ŞEKİLLER DİZİNİ	x
BİRİNCİ BÖLÜM	
GİRİŞ	
İKİNCİ BÖLÜM	
KURAMSAL ÇERÇEVE/ÖNCEKİ ÇALIŞMALAR	
ÜÇÜNCÜ BÖLÜM	
MATERYAL VE METOT	
3.1. Materyal	9
3.2. Metot	13
3.2.1. Web Kazıyıcı	14
3.2.2. Veri Ön İşleme	14
3.2.3. Karar Ağacı	18
3.2.4. Rassal Orman	19
3.2.5. Destek Vektör Makinesi	20
3.2.6. Yapay Sinir Ağı	22
3.2.7. Hiper Parametre Optimizasyonu	25
3.2.8. Çoğunluk Oylaması	26
3.3. Meta Model	26
3.4. Validasyon	28
3.4.1. R Kare	28

3.4.2. Korelasyon Katsayısı	28
DÖRDÜNCÜ BÖLÜM	
ARAŞTIRMA BULGULARI	
4.1. Hiper Parametre Optimizasyon Değerleri	30
4.2. Korelasyon Sonuçları	32
4.3. R Kare Sonuçları	33
BEŞİNCİ BÖLÜM	
SONUÇ ve ÖNERİLER	
KAYNAKÇA	36

SİMGELER VE KISALTMALAR

HY	Hedonik Yöntem
RA	Regresyon Analizi
HFM	Hedonik Fiyat Modeli
LTT	Lancaster Tercih Teorisi
HM	Hedonik Model
MÖ	Makine Öğrenmesi
RM	Regresyon Modelleri
DVM	Destek Vektör Makinesi
KA	Karar Ağacı
RO	Rassal Orman
YSA	Yapay Sinir Ağı
RMSE	Kök Ortalama Kare Hatası
MAE	Ortalama Mutlak Hata
DR	Doğrusal Regresyon
MSE	Ortalama Kare Hata
MAPE	Ortalama Mutlak Yüzde Hata
KNN	K-En Yakın Komşu
RA	Regresyon Ağaçları
COD	Dağılım Katsayısı
PK	Pearson Korelasyonu
GAM	Gradyan Arttırma Makinesi
LSSVR	En Küçük Kareler Destek Vektör Regresyonu
GA	Genetik Algoritma
R^2	R Kare
EO	Evrimsel Optimizasyon
IQR	Çeyrekler Arası Mesafe
EA	Evrimsel Algoritmalar
TO	Topluluk Öğrenmesi
ÇO	Çoğunluk Oylaması

TABLULAR DİZİNİ

Tablo No	Tablo Adı	Sayfa No
Tablo 1	Öznitelikler ve özniteliklere ait veri türleri	9
Tablo 2	Özniteliklere ait kategoriler	10
Tablo 3	Öznitelik ve özniteliğe ait eksik veri sayısı	12
Tablo 4	Özniteliklere ait minimum, maksimum, ortalama ve çeyreklik değerleri	15
Tablo 5	Öznitelik ve her özniteliğe ait aykırı veri adedi	18
Tablo 6	Kernel fonksiyonları ve matematiksel eşitliği	22
Tablo 7	Toplam fonksiyonlarına ait matematiksel eşitlik	24
Tablo 8	Optimizasyon sonucu elde edilen hiper parametre değerleri ve varsayılan değerler	30
Tablo 9	Korelasyon katsayısına ait değerler	32
Tablo 10	Çalışma sonucu oluşan R^2 değerleri	33

ŞEKİLLER DİZİNİ

Şekil No	Şekil Adı	Sayfa No
Şekil 1	Satılık konut fiyat tahmini için geliştirilen meta model akış şeması	13
Şekil 2	Örnek box-plot grafiği	16
Şekil 3	Özniteliklerin değer aralığını gösteren box-plot grafiği	17
Şekil 4	Yapay sinir hücresi	23
Şekil 5	İki giriş ve üç katmana sahip sinir ağı	25



BİRİNCİ BÖLÜM

GİRİŞ

İnsanlar için en temel gereksinimlerden biri barınmadır. Barınma ihtiyacının karşılanabilmesi adına sağlıklı, güvenli, çevresel faktörlerden izole, konforlu ve kaliteli bir yaşam için konutlara ihtiyaç duyulmaktadır. Öyle ki yaşam standartları yükseldikçe konut bir yatırıma, lüks ihtiyaca ya da bireysel tahmin aracında dönüşmektedir.

Bir yatırım aracı olarak görülmeye başlaması ile konutlar, hanehalkları, bankalar, yatırımcılar ve emlak geliştiricileri için ilgi odağı haline gelmiştir (Schulz ve Werwatz, 2004). Hanehalkları bütçelerinin önemli bir kısmını konutlara ayırmaktadır. Konutlara yapılan yatırımlar sonucunda yüksek kâr beklentisi konut fiyatlarında artışa sebebiyet vermektedir. Konutlara olan talebin kısa zamanda karşılanamadığı durumda bu artış giderek yükselmektedir. Konut fiyatlarında meydana gelen fiyat dalgalanmaları ülke ekonomisi ve sosyoekonomik koşullar üzerinde büyük bir etkiye sahiptir. Bu etki hem hükümetleri hem de bireyleri yakından ilgilendirmektedir (Selim, 2009). Bu nedenle alıcılar, satıcılar, yatırımcılar, maliyeti azaltmak ve riski daha iyi yönetmek isteyen finansal kuruluşlar, konut kredisi sağlayan bankalar için konutun sahip olduğu gerçek değeri bilmek elzemdir.

Türkiye’de piyasaların hareketli olması, tüketici tarafından tercih edilen özelliklerin çeşitliliği ve bu özelliklerin fiyat üzerindeki etkisinin tam olarak bilinmemesi nedeniyle satılık konutlar için reel ölçüde fiyat tahmini yapmak oldukça zordur.

İlgili fiyat tahminini otomatik olarak yapabilmek için literatürdeki çalışmalar Hedonik Yöntem (HY) ve Makine Öğrenmesi (MÖ) olmak üzere iki temel yaklaşım kullanmaktadır. İlk olarak otomobil sektörünün fiyat endeksi için 1939 yılında Andrew Court tarafından kullanılan (Goodman, 1998) HY, Regresyon Analizi (RA) yardımı ile mülke ait özelliklerin ve niteliklerin belirli değerlerinin toplamını ele alarak her bir özelliğin değerini tahmin etmeye çalışmaktadır (Shimizu vd., 2010). Fiyata etki eden faktörlerin ağırlıklandırılması için gerekli uzman görüşü, değişen varyans durumu, aykırı veri noktaları ve spesifik model prosedürleri HY’nin dezavantajları olarak görülmektedir (Yazdani, 2021).

MÖ kendi kendine öğrenebilmesi sayesinde ihtiyaç duyulan uzman görüşünü ortadan kaldırmıştır. Konut fiyat tahmini için MÖ ile gerçekleştirilen çalışmalarda tek bir yöntemin kullanılmasından ziyade birden fazla yöntemin performansı karşılaştırılmaktadır. Çalışmalarda Regresyon Modelleri (RM), Destek Vektör Makinesi (DVM), Karar Ağacı (KA) temelli yöntemler, Rassal Orman (RO) ve Yapay Sinir Ağı (YSA) yöntemleri kullanım sıklığı ve tahmin performansı ile ön plana çıkmaktadır. RM zaman ve hesaplama maliyeti açısından öne çıkarken doğrusal ilişkiye sahip olmayan parametrelerin arasındaki ilişkiyi öğrenmede zayıf kaldığı görülmektedir. KA yöntemi tahmin sonucunun arkasında bulunan mantığı görsel olarak sunabilmesi ile tercih edilirken tek bir ağaç yapısına sahip olması dezavantajıdır. Bu dezavantajı RO birden fazla ağaç yapısı ile ortadan kaldırmaktadır. Tahminin arkasında bulunan faktörler göz ardı edildiğinde doğrusal olmayan problemler için DVM ve YSA yöntemleri başarılı sonuçları ile öne çıkmaktadır.

Bu tez çalışmasında Türkiye’de sürekli değişiklik gösteren satılık konut fiyatlarının standartlaştırılmasını sağlamak ve satılık emlağın gerçek fiyatını tahmin edebilecek bir meta model geliştirilmesi amaçlanmıştır. Tek yöntem temelli modellerin dezavantajlarını ortadan kaldırmak ve tahmin performansını artırabilmek için Ağaç Temelli Yöntemler, RM, YSA ve DVM yöntemleri torbalama mantığı ile birleştirilmiştir. Meta model içerisindeki alt modellerin tahmin performansını artırmak adına hiper parametre optimizasyonu yapılmıştır. Tahmin sonuçlarının seçiminde Çoğunluk Oylaması (ÇO) yaklaşımına başvurulmuştur. En yüksek oya sahip tahmin meta model sonucunu oluşturmuştur. Böylece model seçiminde yaşanan belirsizlikler ortadan kaldırılmıştır. İlgili meta modelin eğitimi, testi ve validasyonu için Çanakkale ilindeki satılık konut ilanları tasarlanan bir web kazıyıcı ile çekilmiştir. Model sonucunun genellenebilmesi için veriler tek bir tarihte değil, belirli bir tarih aralığında çekilmiştir.

İKİNCİ BÖLÜM

KURAMSAL ÇERÇEVE/ÖNCEKİ ÇALIŞMALAR

Çalışmanın bu bölümünde satılık konut fiyat tahmini için MÖ yöntemlerini kullanan çalışmalar incelenmiş ve bu çalışmaların bir özeti sunulmuştur.

Ecer (2014) konut fiyat tahmini için gerçekleştirdiği çalışmada 2013 yılında İzmir ilinde satılan 610 konut verisi ve 83 öznitelikten yararlanılmıştır. Gerçekleştirilen çalışmada YSA ve HM karşılaştırılmıştır. HM'ye göre konut büyüklüğü, duşakabin, giyinme odası, ankastre mutfak, panjur, şofben, asansör, konutun bulunduğu site içerisinde kapalı garaj bulunması, konum olarak okula ve iskeleye yakınlığı, deniz manzarası bulunması gibi etkenlerin konut fiyatını etkilediği görülmüştür. Model performansları Kök Ortalama Kare Hatası (RMSE), Ortalama Mutlak Hata (MAE), Medyanın Mutlak Sapması ve Theil U performans ölçütleri ile değerlendirilmiş ve YSA HM'e göre daha başarılı tahmin sunmuştur.

Gülağız ve Ekinci (2017) konut fiyatının tahmini için gerçekleştirdikleri çalışmada Doğrusal Regresyon (DR), Kara Ağacı Regresyonu ve Rassal Orman Regresyonu karşılaştırmıştır. Model oluşturmak için veri kümesi Kaggle sitesinden alınmış ve 1460 farklı konut ve 79 öznitelik kullanılmıştır. Yöntemlerin karşılaştırılmasında MAE ve Görelî Mutlak Hata'dan yararlanılmıştır. Veri kümesine 10 kat çapraz doğrulama uygulanmış ve hata metriklerine göre en başarılı tahmin DR ile sağlanmıştır.

Phan (2018) çalışmasında Avustralya'nın Melbourne şehri için 2016-2018 yıl aralıklarını ele almış ve konutun fiziksel niteliklerinin yanında farklı özellikler de kullanmıştır. Bu yıllar arasında satılmış olan 34857 konut ve her konut için yirmi bir öznitelik kullanılarak veri kümesi oluşturulmuştur. Veri kümesinde bulunan özniteliklerin yanı sıra konutun bulunduğu konum için enlem ve boylam bilgisi kullanılmıştır. Veri kümesi adimsal ve Temel Bileşen Analizi yardımıyla veri ön işlemeden geçirilmiştir. Ön işleme sonrasında konut sayısı yirmi bin ve öznitelik sayısı on bire düşürülmüştür. Çalışmada doğrusal, ağaç ve polinom RM, YSA ve DVM ile modeller oluşturularak tahmin performansları karşılaştırılmıştır. Karşılaştırma işlemi için Ortalama Kare Hata (MSE) hata

metriği kullanılmıştır. Her modelin değerlendirilmesi MSE DR MSE'ye bölünmesiyle en iyi sonucu veren model 0.56 değerlendirme oranı ile Stepwise & tuned DVM olmuş ve DVM modelinin ezberlemeye açık olduğu ortaya konulmuştur. Zaman maliyeti dikkate alındığında en etkili modelin Polinom Regresyon olduğu gösterilmiştir.

Huang (2019) çalışmasında ABD'nin Los Angeles şehrinin üç ilçesinde yer alan taşınmaz gayrimenkullerin fiyatını analiz etmiştir. Eğitim ve test aşamasında kullanılan veri kümesinde 24156 gayrimenkul verisi ve 18 öznitelik yer almıştır. Çalışmada hem doğrusal hem de doğrusal olmayan RM, Karar Ağaçları, Boosting, RO ve DVM metotları kullanılmıştır. Kullanılan veri kümesi için her model %50 eğitim ve %50 test olarak ayrılmıştır. Çalışmanın ana katkısı doğrusal modellerin karmaşık gayrimenkul verilerinde öngörücü olmadığı ve ağaç tabanlı doğrusal olmayan modellerin en doğru konut fiyat tahmini yaptığını göstermesidir. Ağaç tabanlı modeller öznitelik sayısı fazla olan modellerde yüksek performans sergilediği ve öznitelik seçiminde sağduyu ile eksiltme yapılması zararlı bir durum ortaya koymuştur.

Walthert ve Sigrist (2019) gerçekleştirdiği kapsamlı çalışmada İsviçre'nin 3471 farklı bölgesi için konut fiyat tahmini gerçekleştirmiştir. Çalışmada RM, derin öğrenme ve gradyan destekli regresyon ağaçları modelleri karşılaştırmıştır. Veri kümesi, 2011-2015 yıllarını arasında bulunan 88247 konut verisi ve 14 farklı öznitelikten oluşmuştur. Bu veriler eğitim ve test aşamasında kullanılırken model validasyonunun sağlanması için 2017 yılı verilerinden yararlanılmıştır. MSE ve Ortalama Mutlak Yüzde Hata (MAPE) hata ölçütleri kullanılarak model karşılaştırılması yapılmış ve karşılaştırma sonucunda MSE=0,0332 ve RMSE=0,1385 değerlerine sahip Gradyan Arttırma, Derin Öğrenme ve DR tahminlerinin ortalaması meta modelin diğer bütün modellerine göre çok daha iyi performans sergilediğini göstermiştir.

Baldominos vd. (2019) MÖ yöntemlerini kullanarak konut piyasasının çok altında olan konut fiyatlarını tanımlayan bir çalışma yapmıştır. Çalışmada İspanya'nın Madrid şehrinde yer alan tek bir bölge için 92 adet villa ve 2174 adet apartman dairesi kullanılmıştır. Veri kümesiyle YSA, K-En Yakın Komşu (KNN), DVM ve Regresyon Ağaçları (RA)

kullanılarak modeller oluşturulmuştur. Oluşturulan modeller için ezberlenme ve taraf tutmanın önlenmesi adına beş katmanlı çapraz geçerlilik testi uygulanmıştır. Yapılan çalışma sonucunda RA diğer modellere göre daha iyi performans göstermiştir.

Navvaro vd. (2019) İspanya'daki 433 belediyenin bulunduğu bölgeleri kapsayan 790631 konut için çalışma gerçekleştirmiştir. Literatürde yer alan diğer çalışmalar yalnızca şehir veya bir bölge için yoğunlaşmıştır. Veri kümesinde yer alan aykırı değerler Mahalanobis Uzaklığı ile temizlenmiştir. Bagging, Boosting ve RO modelleri oluşturulmuş ve model performansları Ortalama Oran, MAPE, Dağılım Katsayısı (COD) ve Medyan Mutlak Yüzde Hata kullanılarak değerlendirilmiştir. Modeller arası anlamlı bir fark görünmese de Bagging ve RO modeli COD ve MAPE değeri açısından daha iyi performans sergilemiştir.

Sevgen ve Tanrıvermiş (2020) çalışmalarında konut fiyat tahmini için RO yöntemi kullanmıştır. Veri kümesi için Ankara ilinin Yenimahalle ilçesinde bulunan 189 konut ve bu konutlara ait on üç öznitelik kullanılmıştır. Çalışmada önemi vurgulanan her daireye ait konum bilgisi için hastane, okul ve metroya olan uzaklıklar hesaplanmış ve veri kümesine eklenmiştir. RO modelinin sahip olduğu ağaç sayısı değiştirilmiş fakat sonuca etki etmemiştir. Model için tek bir eğitim ve test kümesi oluşturulmuştur. Model performansının değerlendirilmesi için yalnızca RMSE kullanılmıştır. Konut fiyatını etkileyen faktörler çalışma sonucunda belirtilmemiştir.

Ahtesham vd. (2020) Pakistan'ın Karaçi şehri için gerçekleştirilen çalışmada Gradyan Arttırma XGBoost modeli kullanılmıştır. 38961 adet konut verisinin yer aldığı veri kümesi Pakistan'ın open data sitesinden elde edilmiştir. Veri kümesinde yer alan yirmi adet öznitelik bilgisi Pearson Korelasyonu (PK) kullanılarak fiyat tahmini ile ilişkiye sahip olmayan öznitelikler çıkarılmış ve öznitelik sayısı on dörde düşürülmüştür. Oluşturulan model oluşturulurken eğitim ve test için çeşitli oranlara ayrılarak denenmiştir. Model için MAE kullanılarak farklı parametrelerin optimizasyonu sağlanmıştır. Sonuç olarak oluşturulan model %98'lik bir fiyat tahmin performansı göstermiştir.

Ho vd. (2020) çalışmalarında Hong Kong'a ait kapsamlı bir çalışma gerçekleştirmiştir. On sekiz yıllık zaman aralığına sahip veri kümesi içerisinde kırk bin adet konut verisi kullanılmıştır. Veri kümesinde yer alan öznitelikler KA ile ön işlemeden geçirilmiştir. Çalışmada RO, DVM ve Gradyan Arttırma Makinesi (GAM) algoritmaları kullanılarak farklı modeller oluşturulmuştur. Tahmin doğruluklarının karşılaştırılmasında RMSE, MAPE ve MSE hata ölçütleri kullanılmıştır. Tahmin gücü açısından RO ve GAM başarılı sonuçlar üretirken zaman maliyeti dikkate alındığında DVM daha başarılı yöntem olduğu görülmüştür.

Pai ve Wang (2020) Tayvan'ın Taichung şehri için üç yıllık bir zaman aralığını kapsayan 32215 konut verisi üzerinde çalışmıştır. Konutların bulunduğu konumlar kullanım durumuna göre üç bölgeye ayrılmıştır. Konutlar için kullanılan özniteliklerin sayısı PK uygulanarak on bire indirilmiştir. Regresyon Sinir Ağları (GRNN), Geri Yayılım Sinir Ağları (BPNN) ve En Küçük Kareler Destek Vektör Regresyonu (LSSVR) yöntemleriyle modeller oluşturulmuştur. Modeller için parametre ayarı Genetik Algoritma (GA) ile sağlanmıştır. Bu sayede Izgara Optimizasyonu gibi kaba kuvvet yaklaşımların hesaplama ve zaman maliyeti azaltılmıştır. Veriler her model için %80 eğitim ve %20 test verisi olarak ayrılmış ve beş kat çapraz doğrulama uygulanmıştır. Modeller Normalleştirilmiş Ortalama Mutlak Hata ve MAPE ile test edilmiştir. En az hatayı sağlayan model LSSVR olmuştur.

Jha vd. (2020) geniş zaman aralığını kapsayan çalışmalarında tahmin modelinin geliştirilmesi için 2010-2019 yıllarına ait 94530 adet konut verisi ve her konut için yirmi öznitelik kullanmıştır. Özniteliğin yanında gayri safi yurtiçi hasıla, tüketici fiyat endeksi, etkin federal fon oranı ve konut fiyat endeksi gibi sosyoekonomik değerler de veri kümesine eklenmiştir. Fiyata etki etmeyen özniteliklerin elenmesi PK ile yapılmıştır. Lojistik Regresyon, RO, Voting Sınıflandırıcısı ve XGBoost modelleri kullanılmıştır. Bu modeller geçerlilik testi ile eğitilmiş; doğruluk, kesinlik, geri çağırma, F1 puanı ve hata oranı metrikleri kullanılarak karşılaştırılmıştır. Karşılaştırma neticesinde XGBoost modeli hesaplama karmaşıklığı ve zaman maliyeti açısından en başarılı model olduğu ortaya konulmuştur. Çalışmada kullanılan veri kümesi özelliklerinin bölgelere göre değişiklik gösterdiği belirtilmiş bu yüzden bölge nazarında model kurulmasının tahmin performansını arttıracacağı vurgulanmıştır. Çalışmanın başarılı yönlerine bakıldığında öznitelik çıkarımı

yapılması, çapraz geçerlilik kullanılarak modellerin farklı kümeler üzerinde eğitim ve testi, model karşılaştırmasının farklı metriklerle yapılmış olması ve çalışmanın uzun yılları kapsaması elde edilen sonuçların genelleştirilebilirliğini arttırmıştır.

Zaman vd. (2021) Pakistan'ın başkenti İslamabat için satılık konut fiyatlarını farklı MÖ yöntemleri kullanarak tahmin etmeye çalışmıştır. Çevrimiçi emlak sitesinden web kazıyıcı kullanılarak oluşturulan veri kümesinde konut verileri için 13 öznitelik, çevresel ve coğrafi özellikler için 10 adet öznitelik yer almıştır. DVM, Elastik Net, YSA ve Doğrusal, Bayesian Ridge, Lasso, Pasif Agresif Regresyon, Gradient Boosting regresyon yaklaşımları kullanılmış ve RM'nin performanslarının değerlendirilmesinde MAPE, RMSE ve MAE hata ölçütleri kullanılmıştır. MAPE= 1918.49, MAE= 8595.60 ve RMSE= 18209.55 değerleriyle en başarılı tahmin sonucunu DVM ortaya koymuştur.

Tabar vd. (2021) gerçekleştirdiği çalışmada Tokat ilinin tek bir mahallesi için 176 adet konut verisi kullanmıştır. Veri kümesi emlak satış sitesinden elde edilmiştir. YSA ve RA ile oluşturulan model sonuçları MAPE kullanılarak karşılaştırılmıştır. İki modelden elde edilen doğruluk değerlerine göre YSA %96'lık bir fiyat tahmin başarısı sağlarken RA %95'lik başarı oranı vermiştir. Modeller tek bir metrik üzerinden karşılaştırılmıştır. Oluşturulan modellerin daha iyi karşılaştırılabilmesi için gerekli olan hassasiyet ve özgüllük gibi hata ölçütleri kullanılmamıştır.

Yazdani (2021) Colorado eyaletinin Boulder şehri için gerçekleştirdiği çalışmada HM ile YSA, RO ve KNN yöntemlerini karşılaştırmıştır. Modeller için oluşturulan veri kümesi, Çoklu Listeleme Hizmeti ve Devlet Okulu Derecelendirmeleri gibi farklı kaynaklardan elde edilmiş ve 2019 yılına ait 1061 satılık konut yer almıştır. Kategorik veri kullanılmayan modeller için one-hot kodlama tekniği uygulanmıştır. Veri kümesi %80 eğitim ve %20 test olarak ayrılmış ve ilgili modeller R Kare (R^2), MAPE ve RMSE hata metrikleri ile değerlendirilmiştir. RO modelinin konut fiyat tahmini için etkili bir model olduğu görülmüştür. RO, konut piyasalarının karmaşıklığını veya doğrusal olamama durumunu diğer modellerden daha iyi şekilde uygulamıştır.

Gerçekleştirilen çalışmada Türkiye’de satılık konut fiyatlarında sık sık meydana gelen fiyat dalgalanması sonucunda oluşan tutarsız fiyatlandırmanın sonlandırılması ve standartlaştırılmanın sağlanması için konut fiyatlarının gerçek değerinin belirlenmesi amaçlanmıştır. İlgili çalışmalar incelendiğinde satılık konut fiyat tahmini için oluşturulan modelin performansına etki eden en önemli faktörlerden biri veri kümesinin doğru bir şekilde hazırlanmasıdır. Bu çalışmada satılık konut verileri hazırlanırken güncel olarak kullanılan bir emlak sitesi üzerinden web kazıyıcı yardımıyla elde edilmiştir. Web kazıyıcı sayesinde veriler hem eksiksiz bir şekilde elde edilmiş hem de zaman maliyeti azaltılmıştır. Veriler belirli zaman aralığında çekilerek model sonuçları genellenebilir hale getirilmiştir. Veri kümesi veri ön işleme adımlarından geçirildikten sonra validasyon, eğitim ve test kümelerine ayrılmıştır. Validasyon kümesi, başarımlar için kullanılırken eğitim ve test kümesi için çok katmanlı çapraz doğrulama uygulanmıştır. Çok katmanlı çapraz doğrulama sayesinde veri kümesinin ezberlenilmesinden kaçınılmıştır. Bu sayede çıkan sonuçların güvenilirliği artırılmıştır. İncelenen çalışmalarda en başarılı ağaç temelli yöntemlerin, regresyon, YSA ve DVM olduğu görülmüştür. KA ve RO yöntemleri tahmin sonucunda oluşturduğu görselliği ile öne çıkarken doğrusal öğrenme problemleri için regresyon, doğrusal olmayan problemler için YSA ve DVM başarılı sonuçlar üretmiştir. Yürütülen bu tez çalışmasında kullanılan MÖ yöntemleri için en uygun parametre seçiminin sağlıklı bir şekilde gerçekleştirilebilmesi için Evrimsel Optimizasyon’dan (EO) yararlanılmıştır. İncelenen çalışmalarda her yöntem ayrı ayrı değerlendirilmiş veya birbirleri ile karşılaştırılmıştır. Geliştirilen meta model sayesinde konut fiyat tahmini tek bir yöntemle bağlı kalmaksızın YSA, DVM, KA ve RO yöntemi torbalama mantığı ile birleştirilmiş ve bu yöntemler ÇO’ya sokulmuştur. Bu iyileştirme sayesinde tahmin sonucu tek bir yöntemle bağlı kalmaksızın en yüksek oy konut fiyat tahmin sonucunu belirlemiştir.

ÜÇÜNCÜ BÖLÜM

MATERYAL VE METHOD

Bu bölümde tez çalışmasına konu olan meta modelin detayları ile ilgili modelin üzerinde çalışacağı veri kümesinin detayları alt başlıklar halinde sunulmuştur.

3.1. Materyal

Çanakkale il merkezi, 11 ilçe ve 105 mahalleye ait satılık konutlar için fiyat tahmininde kullanılan veri kümesi 1-11 Şubat 2022 tarihleri arasında aktif olarak kullanılan bir web sitesinden Python BeautifulSoup kütüphanesi yardımıyla web kazıyıcı oluşturularak elde edilmiştir. Elde edilen veri kümesi içerisinde 942 adet satılık konut ilanı ve her konut için 19 öznitelik yer almıştır. Veri kümesinde kullanılan öznitelikler ve özniteliklere ait veri türleri Tablo 1’de verilmiştir.

Tablo 1

Öznitelikler ve özniteliklere ait veri türleri

Öznitelik	Veri türü
Fiyat (TL)	Sayısal
Konut şekli	Kategorik
Oda + salon sayısı	Kategorik
Brüt / net m2	Sayısal
Bulunduğu kat	Sayısal
Bina yaşı	Sayısal
Isınma tipi	Kategorik

Tablo 1'in devamı

Kat sayısı	Sayısal
Krediye uygunluk	Kategorik
Eşya durumu	Kategorik
Banyo sayısı	Sayısal
Yapı tipi	Kategorik
Yapının durumu	Kategorik
Kullanım durumu	Kategorik
Tapu durumu	Kategorik
Cephe	Kategorik
Yakıp tipi	Kategorik
Mahalle	Kategorik
İlçe	Kategorik

Kategorik veri türüne sahip verilerin kategorileri Tablo 2'de verilmiştir.

Tablo 2

Özniteliklere ait kategoriler

Öznitelik	Özniteliğe ait kategoriler	Min frekans değeri	Max frekans değeri
Konut şekli	Daire, müstakil ev, villa, yazlık, köy evi, bina, dağ evi, çiftlik evi	1 (Çiftlik evi)	580 (Daire)

Tablo 2'nin devamı

Oda + salon sayısı	3+1, 2+1, 1+1, 4+1, 4+2, 5+1, 3+2, 5+2, 2+2, 6+1, 1+0, 4+3, 6+2, 6+4, 7+3, 8+1	1 (8+1)	298 (3+1)
Isınma tipi	Kombi, merkezi, soba, klima, kat kaloriferi, yerden ısıtma, ısıtma yok, güneş enerjisi, ısı pompası, doğalgaz sobası, belirtilmemiş	2 (Doğalgaz sobası)	369(Kombi)
Krediye uygunluk	Uygun, uygun değil, bilinmiyor	23 (Uygun değil)	705 (Uygun)
Eşya durumu	Eşyalı, eşyalı değil,	64 (Eşyalı)	693 (Eşyalı değil)
Yapı tipi	Betonarme, kagir, taş bina, prefabrik, yığma, çelik	1 (Çelik)	722 (Betonarme)
Yapının durumu	Sıfır, ikinci el, yapım aşamasında, betonarme	4 (Betonarme)	482 (İkinci el)
Kullanım durumu	Boş, ev sahibi oturuyor, kiracı oturuyor, belirtilmemiş	18 (Belirtilmemiş)	499 (Boş)
Tapu durumu	Kat irtifakı, kat mülkiyeti, arsa	19 (Arsa)	459 (Kat irtifakı)
Cephe	Kuzey, güney, doğu, batı, güney-doğu, güney-batı, kuzey-güney, doğu-batı, kuzey-doğu, kuzey-batı, güney-doğu-batı, kuzey-güney-batı, kuzey-doğu-batı, kuzey-güney-doğu, kuzey-güney-doğu-batı	10 (Kuzey-güney-doğu)	374 (Kuzey-güney-doğu-batı)

Tablo 2'nin devamı

Yakıt tipi	Doğalgaz, kömür-odun, elektrik, akaryakıt	2 (Akaryakıt)	521 (Doğalgaz)
------------	---	---------------	----------------

Model oluşturmak için elde edilen veri içerisinde sekiz özniteliğe ait toplam 1552 adet eksik veri tespit edilmiştir. Eksik verilerin sahip olduğu öznitelikler ve eksik veri sayısı Tablo 3'de gösterilmiştir.

Tablo 3

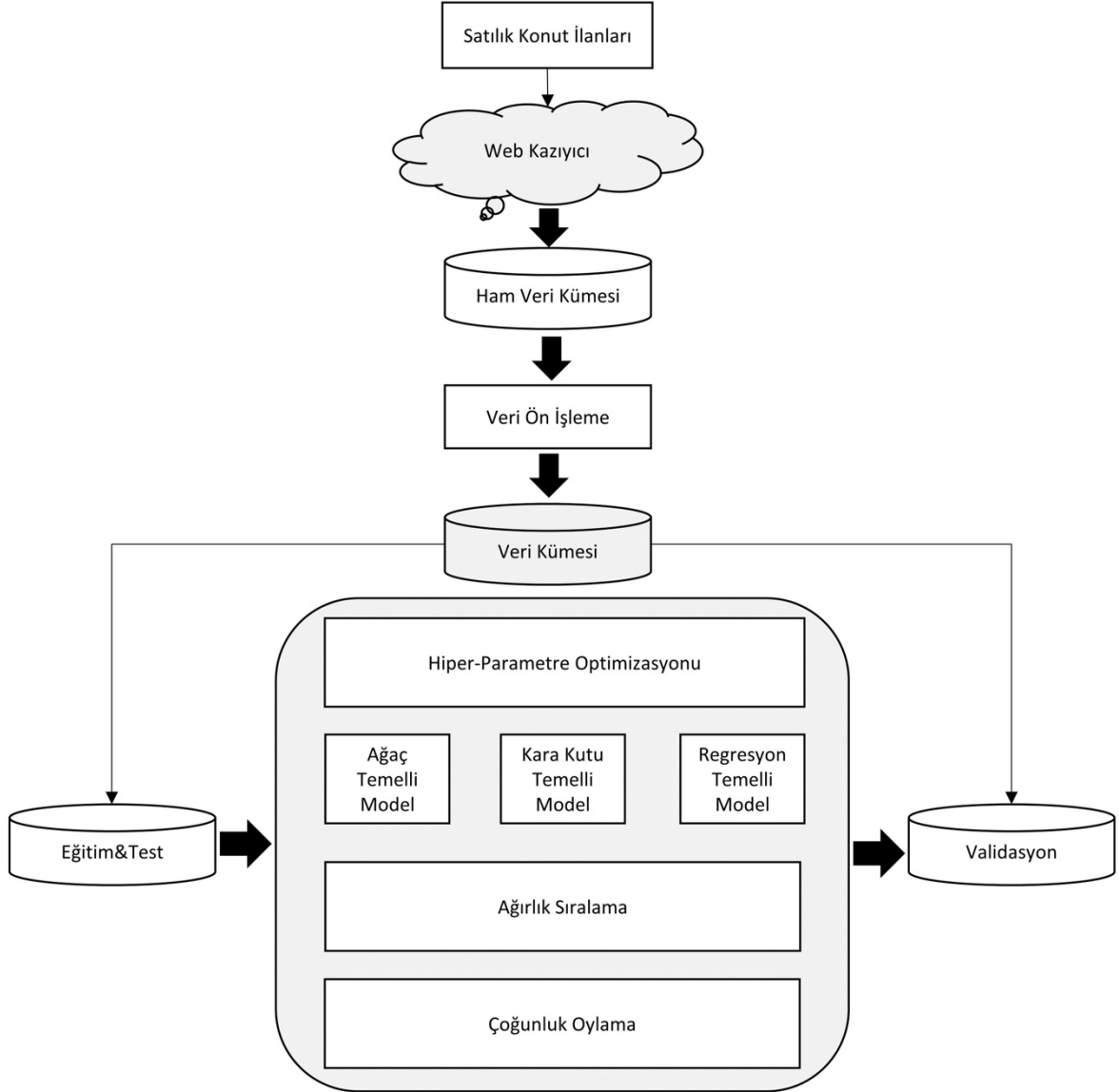
Öznitelik ve özniteliğe ait eksik veri sayısı

Öznitelik	Eksik veri sayısı
Cephe	369
Bulunduğu kat	299
Tapu durumu	237
Yapı tipi	174
Yakıt tipi	164
Yapının durumu	146
Kullanım durumu	140
Kat sayısı	16

Tespit edilen eksik verilerin temizlenmesi için gerekli adımların detayları Veri Ön İşleme alt başlığında detaylı bir şekilde sunulmuştur.

3.2. Metot

Çalışmanın bu bölümünde her bölgeye uygulanabilecek satılık konut fiyatlarının tahmin edilmesi için geliştirilen Şekil 1’de genel hatları sunulmuş meta modelin detayları verilmiştir.



Şekil 1. Satılık konut fiyat tahmini için geliştirilen meta model akış şeması.

İlgili meta model YSA, DVM, KA, RO algoritmalarını torbalama mantığı ile birleştirerek sınıflama kararı ve sayısal sonuç tahmini için ÇO kullanmaktadır. Model içerisindeki her bir yöntemin en iyi performansı verebilmesi için bu yöntemler EO

algoritması ile hiper parametre optimizasyonuna sokulmuştur. Meta modelin eğitimi, testi ve validasyonu için geliştirilen web kazıyıcı yardımıyla çekilip veri ön işleme adımından geçirilen veri kümesi kullanılmıştır. Eğitim ve test adımlarında çok katmanlı geçerlilik kullanılmış, modelin validasyonu için bu adımlarda daha önce kullanılmamış veri kümesi tercih edilmiştir.

3.2.1 Web Kazıyıcı

Geliştirilen modelin sağlıklı sonuçlar üretmesi için satılık konut verilerinin hatasız olarak elde edilmesi oldukça önem arz etmektedir. HTML ve XML dosyalarının ayrıştırılabilmesi için geliştirilen Python BeautifulSoup kütüphanesi yardımıyla web kazıyıcı oluşturulmuştur. Web kazıyıcı yardımıyla satılık konut verilerinin elde edilme işlemi otomatikleştirilerek hem zaman maliyeti azaltılmıştır hem de veriler hatasız bir şekilde elde edilmiştir. Aktif olarak kullanılan bir emlak sitesi üzerinden Çanakkale iline ait satılık konut verileri oluşturulan web kazıyıcı ile elde edilmiş ve bu veriler excel dosya formatında saklanmıştır. Fiyat tahmin sonuçlarının genellenebilmesi için satılık konut verileri belirli tarih aralığında çekilmiştir. Elde edilen ham veri kümesi veri ön işleme bölümünde anlatılan işlemlerle veri ön işleme adımlarından geçirilmiştir.

Veri ön işleme adımlarından geçirilen veri kümesi MÖ yöntemlerinde kullanılmak üzere hazır hale getirilmiştir. Oluşturulan meta modelde kullanılan KA, RO, YSA ve DVM yöntemleri devam eden alt başlıklarda sunulmuştur.

3.2.2. Veri Ön İşleme

Genel olarak ham veri kümesi içerisinde gürültülü, fazla, eksik, aralık dışı, hatalı veya tutarsız veriler yer almaktadır (Çetin ve Yıldız, 2021). Veri içerisindeki bu hatalar, modelin tahmin sonuçlarını negatif yönde etkilenmesi ve dolayısıyla güvenilirliğini azaltması gibi dezavantajlar getireceğinden verinin ilgili hatalardan mümkün olduğunca arındırılması gerekmektedir (Keskin vd., 2019; Oğuzlar, 2003). Verideki mevcut hataların ayıklanması veri ön işleme adımı olarak adlandırılır.

Yapılan bu çalışmada Python BeautifulSoup kütüphanesi kullanılarak oluşturulan web kazıyıcı ile elde edilen ham veri kümesinin model oluşturulurken kullanılabilmesi için ham veri kümesi içerisinde yer alan aykırı verilerin tespit edilip kaldırılması ve eksik verilerin giderilmesi ve verinin uygun formata getirilmesi gerekmektedir.

Veri ön işleme ilk adımı olarak aykırı değerlerin tespiti yapılmıştır. Aykırı veriler Çeyrekler Arası Mesafe (IQR) yardımıyla tespit edilmiştir. Q_1-Q_3 , birinci ve üçüncü çeyreklik ve IQR veri noktalarının bulunduğu çeyrekler arasında yer alan mesafedir (Whaley, 2005). Özniteliklerin bulunduğu kolona ait değerler küçükten büyüğe doğru sıralanmış ve beş sayı özetini içeren minimum, maksimum, birinci çeyreklik, ortanca ve üçüncü çeyreklik değerleri hesaplanmıştır (Keskin vd., 2019). Alt ve üst limit değerleri belirlenerek bu aralık dışında kalan değerler aykırı değer olarak alınmıştır. IQR, alt ve üst limit değerlerinin nasıl hesaplandığı aşağıda gösterilen eşitliklerle (Eşitlik 3.1-3.2-3.3) belirlenmiştir (Kannan vd., 2015).

$$IQR = Q_3 - Q_1 \quad (3.1)$$

$$\text{Alt limit} = Q_1 - 1,5 \times IQR \quad (3.2)$$

$$\text{Üst limit} = Q_3 + 1,5 \times IQR \quad (3.3)$$

Öznitelikler ve bu özniteliklere ait minimum, maksimum, birinci çeyreklik, ortalama ve üçüncü çeyreklik değerleri Tablo 4’de verilmiştir.

Tablo 4

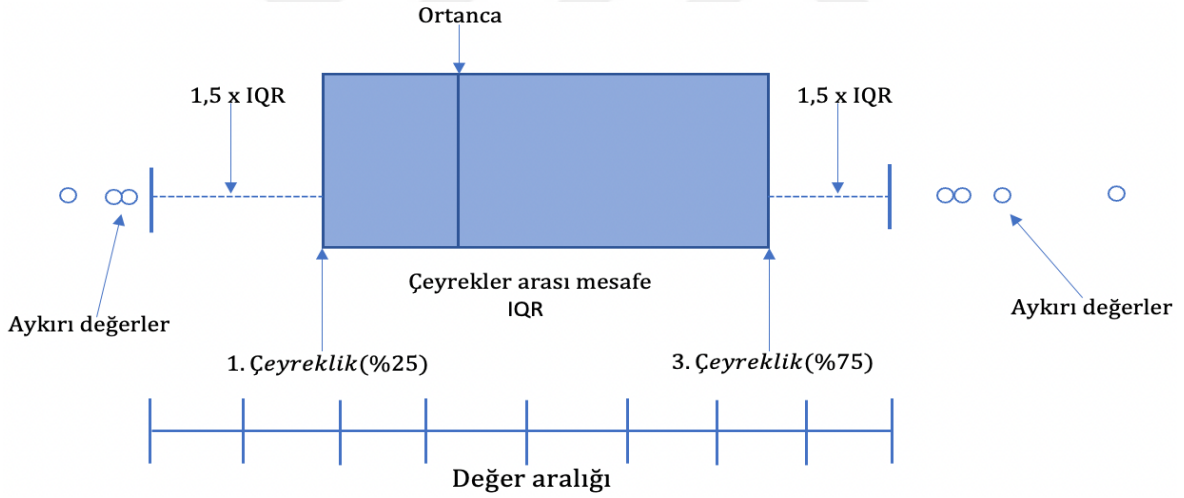
Özniteliklere ait minimum, ortalama, maksimum ve çeyreklik değerleri

	Fiyat	Brüt / Net M2	Bina yaşı	Kat sayısı	Banyo sayısı
Minimum	62500	35	0	1	1
Q1	550000	100	0	2	1

Tablo 4'ün Devamı

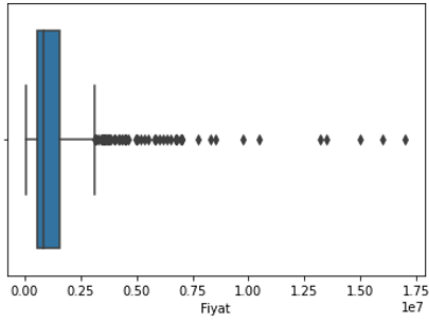
Ortalama	1348081	204,811	10,88	3,644	1.64
Q3	1570000	200	17	5	2
Maksimum	17000000	4638	217	10	20

Veri dağılımını çeyrek değerler yardımıyla görselleştirme ihtiyacını karşılayan box-plot yani kutu grafiğinde yer alan alt ve üst kenarlar Q_1-Q_3 , aradaki çeyrekler arası açıklık IQR, kutu grafiğinin ortasında yer alan çizgi ortanca (medyan) değerini, en alt ve üst çizgilerin dışında kalan noktalar ise aykırı değerleri temsil etmektedir. Örnek box-plot grafiği Şekil 2'de sunulmuştur.

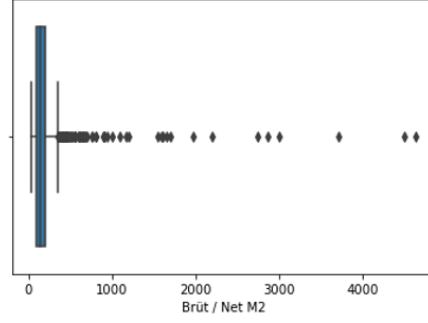


Şekil 2. Örnek box-plot grafiği.

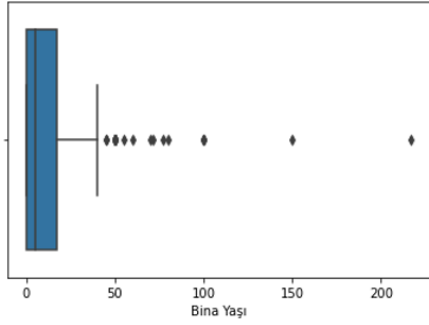
Aykırı değerlerin yer aldığı fiyat, brüt / net m2, bina yaşı, kat sayısı ve banyo sayısı özniteliklerine ait değer aralıklarını görselleştiren box-plot grafiği aşağıda yer alan Şekil 3'de (A-B-C-D-E) gösterilmiştir.



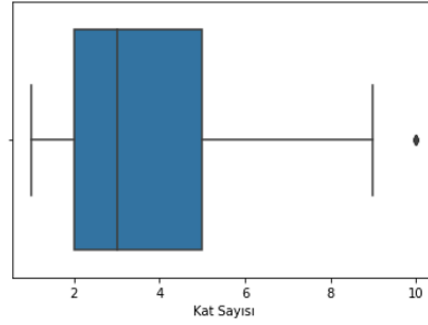
A-) Fiyat özniteliğine ait değer aralığı



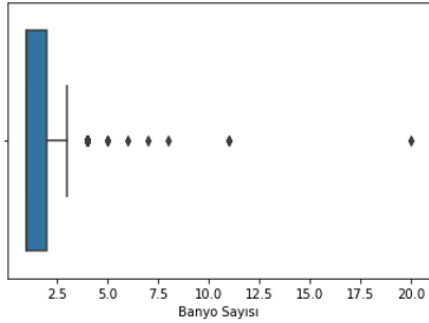
B-) Brüt / Net M2 özniteliğine ait değer aralığı



C-) Bina yaşı özniteliğine ait değer aralığı



D-) Kat sayısı özniteliğine ait değer aralığı



E-) Banyo sayısı özniteliğine ait değer aralığı

Şekil 3. Özniteliklerin değer aralığını gösteren box-plot grafiği.

Aykırı değerlerin tespit edilmesi için fiyat, brüt/net m2, bina yaşı, kat sayısı ve banyo özniteliklerine ait hesaplanan çeyreklik değerleri, IQR, ortanca ve aykırı değerler Şekil 3'de yer alan grafiklerle sunulmuş, en alt ve üst çizgiler dışında kalan aykırı değerler grafikte nokta şeklinde görselleştirilmiştir.

Özniteliklere ait deęer aralıkları belirlendikten sonra aykırı veriler veri kümesinden kaldırılmıştır. Aykırı deęerler veri kümesinden kaldırıldıktan sonra konut veri adedi 757 olmuştur. Kaldırılan aykırı verilere ait öznitelik ve veri adedi Tablo 5’de gösterilmiştir.

Tablo 5

Öznitelik ve her öznitelięe ait aykırı veri adedi

Öznitelik	Aykırı veri adedi
Fiyat	69
Brüt / net m2	76
Bina yaşı	26
Kat sayısı	6
Banyo sayısı	27

Veri ön işleme aşamasında, ikinci adım olarak eksik verilerin giderilmesi üzerinde durulmuştur. Veri türü sayısal olan eksik veriler, bulunduğu kolon içerisinde yer alan verilerin ortalaması alınarak doldurulmuştur. Veri türü kategorik olan veriler ise bulunduğu kolon içerisinde yer alan verilerin frekansı en yüksek olanla doldurulmuştur.

3.2.3. Karar Ağacı

KA yöntemi genel olarak sınıflandırma problemleri için tercih edilmekte ve sayısal sonuca ulaşmak amacıyla kullanılan RA için de kullanılmaktadır. Bir karar destek aracı olan karar ağacı tahmin veya sınıflandırma sonucunu görsel olarak ağaç yapısına benzer bir şekilde sunmaktadır (Uysal ve Güyer, 2014). KA’da yer alan bağımsız deęişkenler bilgi kazancına göre ayrılmakta ve tahmin işleminde tahmin edilen deęer hangi aralıkta ise o aralığın sahip olduęu ortalama deęer sonucu döndürmektedir (Akşehir ve Kılıç, 2019).

KA'nın en üst kısmında yer alan düğüm "kök" olarak nitelendirilmektedir. Kök düğüm karar verme sürecinde en yüksek etkiye sahiptir ve kök düğümden aşağı dallara ve yapraklara doğru inildikçe bu etki azalmaktadır. Ağaca ait kök düğüm ve dallar belirleyici nitelik ile ayrılır ve bu belirleyici nitelik gini indeksi veya bilgi kazancı hesaplamasıyla elde edilir. Entropi hesabı yardımıyla bilgi kazancı bulunur (Akçetin ve Çelik, 2014). Aynı seviyede yer alan dalların parametreleri aynı etkiye sahiptir.

Buna göre N adet değer alan bir A değişkeninin sınıflandırılması için gerekli olan bilgi miktarının hesaplanması Eşitlik 3.4'de verilmiştir. S_i sınıfına ait örnekler \ddot{O}_i ile belirtilmiştir. Tek bir i ögesinin S_i sınıfına ait olma olasılığı P_i ile gösterilmiştir.

$$Bilgi(\ddot{O}_1, \ddot{O}_2, \dots, \ddot{O}_n) = - \sum_{i=1}^n p_i \log p_i \quad (3.4)$$

Bir A değişkeni üzerinden eldeki veri kümesini homojen bir şekilde n adet alt kümeye bölünebilmesi için gerekli olan entropi hesaplaması Eşitlik 3.5'de verilmiştir.

$$Entropi(A) = \sum_{j=1}^n \frac{\ddot{O}_{1j} + \ddot{O}_{2j} + \dots + \ddot{O}_{nj}}{\ddot{O}} * Bilgi(\ddot{O}_{1j}, \ddot{O}_{2j}, \dots, \ddot{O}_{nj}) \quad (3.5)$$

Son kısımda bir değişken için gerekli olan bilgi kazancı hesabı Eşitlik 3.6'da verilmiştir.

$$Bilgi\ Kazancı = Bilgi(\ddot{O}_1, \ddot{O}_2, \dots, \ddot{O}_n) - Entropi(A) \quad (3.6)$$

Tüm değişkenler için bilgi kazancı hesabı yapılarak en yüksek değere sahip değişken ağacın ilgili düğümüne ataması gerçekleştirilir. Her düğüm için kalan değişkenler üzerinden hesaplama işlemi sürdürülür. RA her düğüm için tahmin edilen toplam varyans azaltılır ve ayırım gerçekleştirilerek oluşturulur (Akçetin ve Çelik, 2014).

Sayısal ve kategorik değerlerle çalışabilmesi, doğrusal olmayan problemlerin çözümünde başarılı sonuçlar üretmesi ve tahmin sonucunun arkasında yer alan yapıyı görsel olarak sunabilmesi KA avantajlarındanır.

3.2.4. Rassal Orman

RO yöntemi Breiman tarafından sunulmuştur (Breiman, 2001). Yöntem birden fazla KA'nın meydana getirdiği ağaç topluluğudur ve hem sınıflandırma hem de regresyon problemleri için kullanılabilir. Topluluk içerisinde yer alan her bir KA veri kümesinden rastgele şekilde seçilmiş olan alt kümeler karşılık gelmektedir (Akşehir ve Kılıç, 2019).

Dört adımdan meydana gelen RO yöntemi için ilk adım KA sayısının belirlenmesidir. İkinci adımda ağaçların sahip olduğu her düğüme rassal bir şekilde m adet değişken seçilir ve dallanma için gini hesaplaması kullanılarak en iyi dal belirlenir. Gini indeks sıfırlanmaya kadar düğümler iki alt dala ayrılır ve her bir yaprak düğüm bir sınıfa karşılık geldiğinde dallanma sona erer (Okumuş ve Aydemir, 2017).

Tahmin sonucu, ağaç topluluğunda yer alan ağaçlardan elde edilen oy sonucuna göre belirlenir (Ustalı vd., 2021). En fazla oy tahmin sonucu oluşturur. Oluşturulan modelde yeterli sayıda KA olması veri kümesinin ezberlenme sorununu ortadan kaldırmaktadır (Breiman, 2001).

Her bir KA'da yer alan her bir özneliğin gini önemi Eşitlik 3.7'de verilmiştir. \bar{O}_j bir j düğümün önemi, A_j düğüme ulaşan örneklerin ağırlıklandırılmış sayısı, S düğümün safsızlık katsayısı, düğüme ait sağ dallanma ayrımı S_a ve sol dallanma ayrımı S_o olarak gösterilmiştir.

$$\bar{O}_{ij} = A_j S_j - A_{S_{oj}} S_{S_{oj}} - A_{S_{aj}} S_{S_{aj}} \quad (3.7)$$

RO yönteminde hiper parametrelere farklı değerler atanarak birden fazla ağaç oluşturulmaktadır. Oluşturulan bu ağaç topluluğu sayesinde KA hiper parametrelerinin sahip olduğu tekillikten kaynaklanan performans değişikliği ortadan kalkmaktadır (Schonlau ve Zou, 2020).

3.2.5. Destek Vektör Makinesi

DVM verileri en uygun şekilde ayırmak amacıyla n boyutta bir hiper düzlem oluşturur (Kavzoğlu ve Çölkesen, 2010) ve bunun için bir çekirdek (kernel) fonksiyonu kullanır. Bu sayede hem regresyon hem de sınıflandırma problemlerinde kullanılabilir. Hiper düzlem oluşumunda sınıf ayırıcıları arasındaki mesafenin maksimum olmasına çalışır (Kavzoğlu ve Çölkesen, 2010). Bu maksimum mesafe maksimum marjin olarak nitelendirilmektedir. Hinge Loss fonksiyonu ile bu mesafenin maksimum olması sağlanmaktadır (Xing ve Ji, 2018). İlgili fonksiyon Eşitlik 3.8'de gösterilmiştir.

$$(x, y, f(x)) = \begin{cases} 0, & \text{eğer } y * f(x) > 1 \\ 1 - y * f(x), & \text{ise} \end{cases} \quad (3.8)$$

DVM'de eğitim örnekleri regresyon için Eşitlik 3.9'da gösterilmiştir ve bu problemlerde yanıt değişkeni olarak $y_i \in \mathbb{R}$ şeklinde tanımlanmaktadır (Daş vd., 2019).

$$[(x_i - y_i)]_{i=1}^N = 1 \quad (3.9)$$

DVM, eğitim kümesinde yer alan verilerin benzer ve bağımsız bir şekilde dağıldığını farz etmektedir. Çekirdek algoritması ve ε hatasının bulunmadığı fonksiyon DVM'nin temelini oluştururken doğrusal olmayan regresyon problemlerinde öğrenme yöntemi Eşitlik 3.10'da gösterilmiştir. ω ağırlık vektörü, (\cdot) iç çarpım işlemi ve b eşik değeridir (Li vd., 2009).

$$f(x) = \omega \cdot \phi(x) + b \quad (3.10)$$

Yapısal risk ilkesi göz önüne alındığında riskin minimum seviyeye indirilmesi için ε hatası göz önünde bulundurularak minimum ω vektörü bulunarak vektörün normu minimize edilir ve $f(x)$ fonksiyonu bulunur (Bilişik, 2011). Bazı kısıtlamalar sebebiyle ξ_i, ξ_i^* ile gösterilen serbest değişkenler eklenmiştir. Fonksiyon ve kısıtlar Eşitlik 3.11 ve 3.12'de belirtilmiştir (Smola ve Schölkopf, 2004). Sabit $C > 0$ değeri f ile ε 'den daha büyük sapmalar olduğu zaman göz ardı edilen miktar arasında kurulan dengeyi belirler. Bu parametre kullanıcı tarafından belirlenirken çok büyük veya çok küçük seçilmesi

sınıflandırma başarısını azaltır ve $C = \infty$ olma durumu yalnızca doğrusal problemler için uygun bir seçimdir (Kavzoğlu ve Çölkesen, 2014).

$$\text{Min: } \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (3.11)$$

$$\begin{aligned} \gamma_i - \langle \omega, x_i \rangle + b &\leq \varepsilon + \xi_i \\ \text{Kısıtlar: } \langle \omega, x_i \rangle + b - \gamma_i &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned} \quad (3.12)$$

DVM için kullanılacak kernel fonksiyonu ve uygun parametrelerin belirlenmesi gerekmektedir. DVM’de kullanılan kernel fonksiyonları ve matematiksel karşılığı Tablo 6’da sunulmuştur. γ, d, σ ve ω değerleri sırasıyla kernel boyutu, polinom derecesi ve pearson genişliği parametreleridir (Daş vd., 2019).

Tablo 6

Kernel fonksiyonları ve matematiksel eşitliği

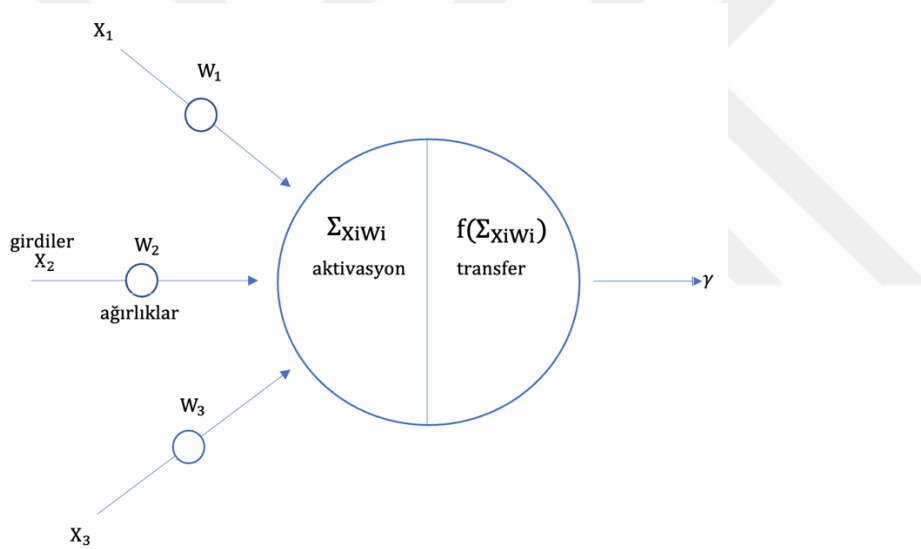
Kernel fonksiyonu	Matematiksel eşitlik
Polinom kernel	$K(x, y) = ((x \cdot y) + 1)^d$
Normalleştirilmiş polinom kernel	$K(x, y) = \frac{((x \cdot y) + 1)^d}{\sqrt{((x \cdot y) + 1)^d ((y \cdot y) + 1)^d}}$
Radyal tabanlı fonksiyon kernel	$K(x, y) = e^{-\gamma \ x - x_i\ ^2}$
Person evrensel kernel (PUK)	$K(x, y) = \frac{1}{\left[1 + \left(\frac{2 \cdot \sqrt{\ x - y\ ^2 \sqrt{2^{(1/\omega)} - 1}}}{\sigma} \right)^2 \right]^\omega}$

Doğrusal olmayan durumlarda DVM kernel fonksiyonları sayesinde orijinal giriş uzayından yüksek boyutlu uzaya geçiş yaparak kolay bir şekilde sınıflandırma yapar. Yüksek boyutlu uzayda değer bulma işlemi bütün değerleri kapsayacak şekilde çarpma işlemi yerine kernel fonksiyonunda yerine konularak sağlanır (Yakut vd., 2014). DVM doğrusal olmayan problemler için gerçekleştirdiği başarılı ve hızlı işlemlerle ön plana çıkar.

3.2.6. Yapay Sinir Ağı

İnsan sinir hücrelerine benzer bir yapıda tasarlanan YSA, mantıksal işlemlerin sayısal olarak modellenebileceğini ortaya koymuştur. YSA verilen girdileri çıktı ürettiği kara kutu modeli olarak nitelendirilmesi sinir hücrelerinde bilgi akışının aksonlar yardımıyla taşınmasına benzetilmektedir (Şen, 2004).

Yapay sinir hücresinde giriş kısmından alınan veriler sinir üzerindeki etkiyi belirleyen ağırlıklar ile çarpılarak toplama fonksiyonu yardımıyla eşik değeri ile toplanır ve toplam sonucu aktivasyon fonksiyonuna sokularak çıktı elde edilir (Budak ve Erpolat, 2012). Yapay sinir hücre yapısı Şekil 4’de sunulmuştur (İon vd., 2009).



Şekil 4. Yapay sinir hücresi.

YSA’da toplam, çarpım, en küçük, en büyük, kümülatif ve çoğunluk fonksiyonları toplam fonksiyonu olarak tercih edilebilir. Bu toplam fonksiyonlarına ait eşitlikler Tablo 7’de verilmiştir.

Tablo 7

Toplam fonksiyonlarına ait matematiksel eşitlik

Fonksiyon	Matematiksel eşitlik
Toplam	$TG = \sum_{i=1}^m G_i * A_i$
Çarpım	$TG = \prod_{i=1}^m G_i * A_i$
En küçük	$TG = \min (G_i * A_i)$
En büyük	$TG = \max (G_i * A_i)$
Kümülatif	$TG = TG(\text{Eski}) + \sum_{i=1}^m G_i * A_i$
Çoğunluk	$TG = \sum_{i=1}^m \text{sgn} (G_i * A_i)$

Sonuç nöronunun genel çıktısının elde edilmesi için aktivasyon fonksiyonundan geçirilir ve genellikle aktivasyon fonksiyonu için 0-1 arasında çıktı değeri üreten sigmoid seçilir (Budak ve Erpolat, 2012). YSA'da doğrusal olmayan aktivasyon fonksiyonlarının tercih edilmesi çok farklı ve karmaşık problemlerin çözümünü sağlar (Yavuz ve Deveci, 2012). Daha karmaşık problemler için tercih edilen sigmoid ve hiperbolik tanjant fonksiyonları Eşitlik 3.13 ve 3.14'de verilmiştir.

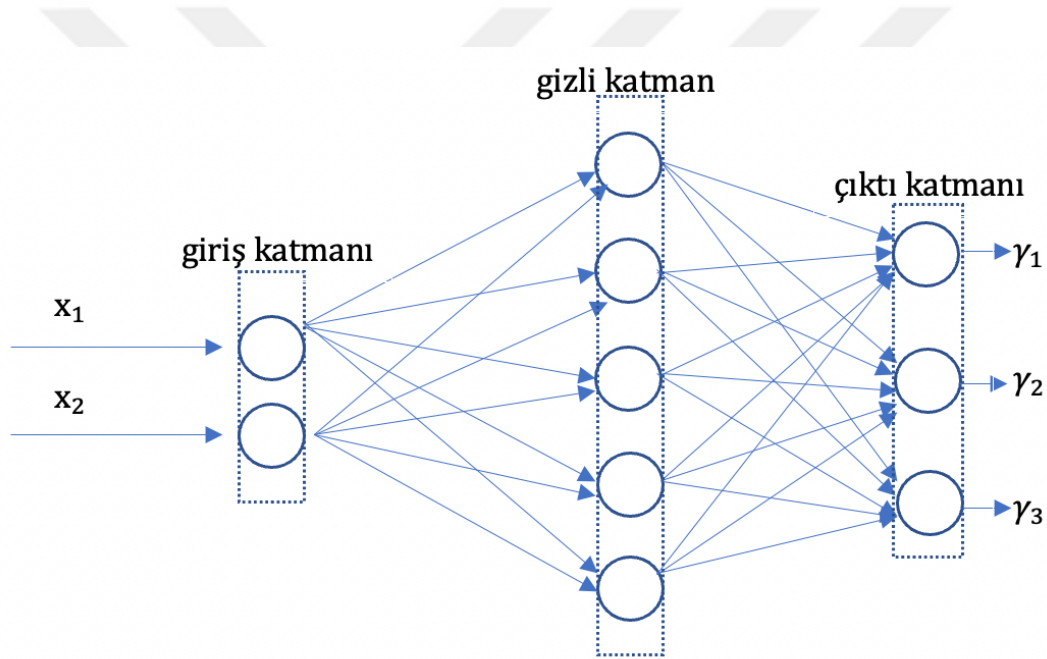
$$\text{Sigmoid: } \alpha = \frac{1}{1+e^{-z}} \quad (3.13)$$

$$\text{Hiperbolik tanjant: } \alpha = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (3.14)$$

Yalnızca girdi ve çıktı katmanına sahip olan YSA tek katmanlı yapıdadır ve karmaşık işlemleri hesaplayabilmek için en az bir ara katmana ihtiyaç duyar (Yılmaz vd., 2018).

Bilgi akışı bakımından yapay sinir hücreleri ileri ve geri beslemeli olarak ayrılmaktadır. Bilgi akışı ileri yönlü olan ileri beslemeli ağlar girdi, bir veya iki gizli katman ve çıktı katmanına sahip olurken bu ağlar çok katmanlı ileri beslemeli ağlar olarak da adlandırılmaktadır (Takma vd., 2012). Geri beslemeli sinir ağı ise en az bir sinir hücresi sonrasında gelen hücrelerce de beslenir ve veri akışı yalnızca ileriye doğru değil geriye doğru da olabilir (Yavuz ve Deveci, 2012).

Çok katmanlı yapıya örnek olarak iki giriş ve üç katmana sahip sinir ağı Şekil 5'te gösterilmiştir. Her katmanda bir veya daha fazla nöron olarak nitelendirilen yapay sinir hücresi mevcuttur (Yılmaz vd., 2018).



Şekil 5. İki giriş ve üç katmana sahip sinir ağı.

Eğitilmiş bir YSA'nın eksik veya hatalı verilerle doğru sonuca ulaşabilmesi en büyük avantajlarından. Bilgi sistemlerinde en küçük hatalara karşı hassasiyet çok fazladır ve sistemde meydana gelecek bu hatalar hatalı sonuçlar üretilmesine sebep olur. Ancak YSA'da bazı nöronlarında oluşacak hasar çıktı sonuçlarını benzer şekilde etkilemez. Sinir ağında işlem elemanları paralel çalışma sayesinde hızlı sonuç üretir (Aktaş, 2003). YSA doğrusal olmayan problemler için de başarıyla ön plana çıkmaktadır (Taşar vd., 2018).

3.2.7. Hiper Parametre Optimizasyonu

Satılık konut fiyatlarının başarılı bir şekilde tahmin edilebilmesi ve yüksek boyuta sahip veri kümesinin neden olduğu hesaplama maliyetinin düşürülmesi için parametre optimizasyonu oldukça önemli bir rol oynamaktadır.

Parametre sayısı fazla ve hesaplama maliyeti yüksek olan yöntemler için özellikle derin öğrenme algoritmalarında, Evrimsel Algoritmalar (EA) çok sık tercih edilmektedir (Tanyıldızı ve Demirtaş, 2019). Rastgele aramaya benzer bir çalışma yapısına sahip EA ile belli parametrelerin değeri sabit tutularak her denemede az sayıda değer değiştirilerek model sonuçları gözlenir ve en iyi sonucu veren parametreler seçilir (Tanyıldızı ve Demirtaş, 2019).

GA da evrim sürecinin taklidini yapan ve evrime ait mutasyon, kalıtım, çaprazlama ve seçim gibi teknikler kullanılarak en iyi parametrelerin tespit edilmesini sağlayan bir yöntemdir. Parametre seçimi için GA yardımıyla parametrelerden meydana gelen alt kümeler oluşturulur. Parametrelerin yer aldığı her alt küme bir birey olarak kabul edilir. EA sayesinde en iyi parametrelerin belirlenmesi için bireyler arası özellikler çaprazlama şeklinde değiştirilebilir (Arunadevi ve Nithya, 2016). GA, parametre optimizasyonu için problemin yapısını dikkate almadan çok karmaşık yapıya sahip problemler için çözüm üretebilir (Kahraman ve Özdağlar, 2004).

3.2.8. Çoğunluk Oylaması

Topluluk Öğrenmesi (TO) yöntemlerinden biri olan ÇO sınıflandırma yöntemlerinin birleştirilmesinde yaygın olarak kullanılır. ÇO basitliği ile ön plana çıkmaktadır. Bireysel yöntemlerden elde edilen sonuçlar arasında oylama yapılarak en fazla oya sahip sınıf etiketi sonuç sınıfını belirler (Onan, 2021). ÇO yardımıyla sınıflandırma performansında iyileştirme gerçekleştirilir.

ÇO ile yanlış karar yalnızca yapılan hata aynı olduğunda veya yanlış sınıf etiketi atayan yöntemlerin sayısı fazla olduğunda oluşur (Lam ve Suen, 1997).

Tahmin sonucu ile sınıf etiketinin karşılaştırılması sonucunda en fazla ağırlığa sahip çoğunluk oyu sonucu belirler. ÇO'ya ait denklem Eşitlik 3.15'de sunulmuştur (Rojarath vd., 2016).

$$Sonuç = Maksimum \sum_{i=1}^{Sınıf} SınıfOyu \quad (3.15)$$

3.3. Meta Model

Gerçekleştirilen çalışmada satılık konut fiyatlarının tahmin edilmesi için Metot bölümünde anlatılan yöntemler kullanılarak her bölgeye uygulanabilecek meta model geliştirilmiştir ve ilgili model yapısı Şekil 1'de sunulmuştur.

Satılık konut fiyatlarının tahmin edilmesinde başarılı sonuçlar elde edilmesi için modelde kullanılacak veri kümesinin hazırlanması oldukça önemlidir. Gerçekleştirilen çalışmada Çanakkale iline ait satılık konut verileri kullanılmıştır. Çanakkale iline ait bu güncel veriler aktif olarak kullanılan bir emlak sitesi üzerinden web kazıyıcı oluşturularak elde edilmiştir. Web kazıyıcı yardımıyla elde edilen satılık konut verileri belirli tarih aralığında çekilmiştir ve bu sayede tahmin sonuçları genellenebilir olmuştur. Oluşturulan web kazıyıcı sayesinde verilerin elde edilme işlemi otomatikleştirilerek veri elde etme süreci azaltılmış ve veriler hatasız olarak elde edilmiş ve excel formatında saklanmıştır. Elde edilen ham veri kümesi eksik ve gürültülü veriler içerdiği için ham veri kümesine veri ön işleme adımları uygulanmıştır. Veri ön işleme adımı sayesinde veri kümesinin kalitesi artırılarak tahmin başarısı da doğru orantılı olarak artırılmıştır. Veri ön işleme sonrasında elde edilen veri kümesiyle kullanılan yöntemlerin uygulanması veri madenciliği, metin madenciliği ve MÖ gibi amaçlar için geliştirilen RapidMiner aracıyla sağlanmıştır. Veri kümesi eğitim-test ve validasyon kümelerine ayrılmıştır.

İlgili çalışmalar incelendiğinde meta modelde tercih edilen KA ve RO yöntemi fiyat tahmini arkasında yer alan yapıyı ağaç oluşturarak sunduğu görsellik ile YSA ve DVM yöntemleri de doğrusal olmayan problemlerdeki başarılı sonuçları ile ön plana çıkmaktadır. Tahmin başarısının artırılması için KA, RO, DVM ve YSA'ya ait hiper parametrelerin optimize edilmesi önemlidir. Her bir yöntem EO'ya sokularak en iyi sonuç verecek hiper parametreler belirlenmiştir. Bulunan bu optimum hiper parametreler kullanılarak MÖ

yöntemlerine k katlamalı çapraz doğrulama uygulanmış ve bireysel tahmin sonuçları elde edilmiştir. Çapraz doğrulama için kullanılan k değeri 10 olarak seçilmiştir. Çapraz doğrulama ile kullanılan veri kümesinin ezberlenmesinin önüne geçilerek gerçekleştirilen tahmin doğruluğunun artırılması amaçlanmıştır. Meta modelde kullanılan KA, RO, DVM ve YSA yöntemleri torbalama mantığıyla birleştirilmiş ve TO yöntemlerinden ÇO uygulanmıştır. Bireysel yöntemlerden elde edilen sonuçlar arasında en yüksek oyu alan tahmin sonucu belirlemiştir. Bu sayede tahmin sonucu tek bir yöntemle bırakılmamış ve tahmin başarısı ÇO kullanılarak artırılmıştır.

3.4. Validasyon

Model performansının değerlendirilmesi için MÖ yöntemlerinin değerlendirilmesinde sıklıkla tercih edilen R^2 ve korelasyon katsayısı kullanılmış ve bu performans metrikleri alt başlıklarda sunulmuştur.

3.4.1. R Kare

Model performansının değerlendirilmesi için kullanılan hata metriği R^2 tahmin edilen değer ve gerçek değer arasındaki bağlantıyı ortaya koyar. R^2 , 0 ve 1 değerleri arasında değişmektedir. Bu değer 1'e ne kadar yakınsa tahmin edilen değer ile gerçek değer arasında bulunan ilişki de o kadar yüksektir. Genel olarak R^2 belirleme katsayısı bir regresyonda bulunan ilişkinin sahip olduğu gücü gösteren bir ölçüttür (Kasuya, 2019). Bu ölçüt Eşitlik 3.16'da sunulmuştur (Yıldız ve Abut, 2022). Y gerçek değer, Y' tahmin değeri ve \bar{Y} gerçek değerlerin ortalamasını göstermektedir.

$$R^2 = 1 - \frac{\sum(Y-Y')^2}{\sum(Y-\bar{Y})^2} \quad (3.16)$$

3.4.2. Korelasyon Katsayısı

Korelasyon katsayısı iki değişken arasında bulunan ilişkinin gücünü ortaya koymaktadır. -1 ve +1 değerleri arasında değişen korelasyon katsayısı ters yönde bir

korelasyon oluřtuęunda -1, aynı ynde olduęunda ise +1 deęerini almaktadır (Abay, 2013). Gerekleřtirilen alıřmada korelasyon katsayısı satılık konut fiyatlarının gerek deęeri ile tahmin edilen deęer arasındaki iliřkiyi ortaya koymak iin kullanılmıřtır.

Korelasyon katsayısı +1 deęerine yaklařtıķa iliřkinin gc anlamlı bir korelasyon deęerine ulařmaktadır. Korelasyon katsayısı forml Eřitlik 3.17’de verilmiřtir (ılgın vd., 2020).

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}} \quad (3.17)$$

DÖRDÜNCÜ BÖLÜM

ARAŞTIRMA BULGULARI

Gerçekleştirilen çalışmanın bu bölümünde KA, RO, DVM, YSA yöntemlerine ait hiper parametrelerin EO ile optimize edilmesi sonucunda elde edilen değerler, bireysel yöntemler ve geliştirilen meta modelin başarı durumunun değerlendirilmesi için kullanılan R^2 ve korelasyon katsayısı metriklerine ait sonuçlar ortaya konulmuştur.

4.1. Hiper Parametre Optimizasyon Değerleri

Çalışmada kullanılan MÖ yöntemleri ÇO'ya sokulmadan önce tahmin başarısını arttırmak ve hiper parametre değerlerini optimize etmek için EO ile hiper parametre optimizasyonu gerçekleştirilmiştir. KA, RO, DVM ve YSA yöntemlerine ait varsayılan hiper parametre değerleri ve optimizasyon işlemi sonucunda oluşan hiper parametre değerleri Tablo 8' de sunulmuştur.

Tablo 8. Optimizasyon sonucu elde edilen hiper parametre değerleri ve varsayılan değerler

Yöntem	Yönteme ait parametre	Hiper parametre değeri	Varsayılan değer
KA	Maksimum derinlik	7	10
	Minimum gain	0,032034	0,01
	Minimum yaprak boyutu	9	2
	Minimum bölme boyutu	27	4
	Alternatif ön budama sayısı	56	3
RO	Ağaç sayısı	87	100
	Makimum derinlik	72	10

Tablo 8'in devamı

SVM	Kernel type	Dot	Dot
	Maximum iterasyon	88	100000
	Kernel cache	68	200
	C	0,744815	0,0
	Covergence epsilon	0,189686	0,001
	L_pos	0,239976	1,0
	L_neg	0,315595	1,0
	Epsilon	0,810368	0,0
	Epsilon_plus	0,263150	0,0
	Epsilon_minus	0,441278	0,0
YSA	Training cycles	141	200
	Learning rate	0,134307	0,01
	Momentum	0,134449	0,9
	Error epsilon	0,140711	1,0E-4

Meta model ve bireysel yöntemlerin tahmin performanslarının değerlendirilmesi için R^2 ve korelasyon katsayısı metrikleri kullanılmıştır. Elde edilen sonuçlar devam eden alt başlıklarda sunulmuştur.

4.2. Korelasyon Sonuçları

Gerçekleştirilen modelde optimizasyon sonucunda oluşan hiper parametre değerleri kullanılarak KA, RO, DVM, YSA yöntemleri uygulanmış ve bu yöntemler torbalama mantığıyla birleştirilerek ÇO gerçekleştirilmiştir. Bunun sonucunda oluşan bireysel yöntemler ve ÇO'ya ait korelasyon katsayısı değerleri Tablo 9'da gösterilmiştir.

Tablo 9

Korelasyon katsayısına ait değerler

Yöntem	Varsayılan değerler ile elde edilen korelasyon katsayısı	Optimizasyon değerleri ile elde edilen korelasyon katsayısı
KA	0,650	0,702
RO	0,814	0,825
DVM	0,786	0,796
YSA	0,787	0,797
ÇO	0,811	0,836

Varsayılan değerler ile gerçekleştirilen bireysel yöntemler (KA, RO, DVM, YSA) ve ÇO ile elde edilen korelasyon katsayısı incelendiğinde RO yöntemi 0,814 korelasyon katsayısıyla en başarılı sonucu üretmiştir. Bireysel yöntemler EO ile optimize edildikten sonra elde edilen optimizasyon değerleri kullanılarak gerçekleştirilen bireysel yöntemler ve geliştirilen meta model sonucu 0,836 korelasyon katsayısı ile en başarılı konut fiyat tahmin sonucu üreten model olmuştur. Ayrıca optimizasyon parametreleri kullanılarak meydana gelen tahmin sonuçları varsayılan değerler kullanılarak meydana gelen tahmin sonuçlarına göre daha başarılı konut fiyat tahmini yapmıştır.

4.3. R Kare Sonuçları

Çalışma sonucunda bireysel yöntemler ve ÇO ile meydana gelen R^2 hata metriğine ait sonuçlar Tablo 10'da verilmiştir.

Tablo 10

Çalışma sonucu oluşan R^2 değerleri

Yöntem	Varsayılan değerler ile elde edilen R^2 değeri	Optimizasyon değerleri ile elde edilen R^2 değeri
KA	0,426	0,496
RO	0,665	0,685
DVM	0,621	0,636
YSA	0,620	0,637
ÇO	0,659	0,701

Tablo 10'da verilen R^2 değerleri incelendiğinde varsayılan değerler kullanılarak gerçekleştirilen yöntemlerin ürettiği sonuçlar, optimizasyon işlemi sonucunda meydana gelen hiper parametre değerleri ile gerçekleştirilen bireysel yöntemler ve geliştirilen meta modelin ürettiği sonuca göre daha başarısız olmuştur. Geliştirilen meta model 0,701 R^2 değeri ile en başarılı tahmin sonucunu üretmiştir.

BEŞİNCİ BÖLÜM

SONUÇ VE ÖNERİLER

Bu tez çalışmasında, konut fiyatlarında meydana gelen tutarsızlaşmanın sonlandırılması ve fiyat standartının sağlanması için farklı MÖ yöntemlerini birleştiren bir meta model oluşturulmuştur. Bu meta modelde KA, RO, DVM ve YSA yöntemleri sağladığı avantajlara göre tercih edilmiştir. Ağaç temelli yöntemler tahmin sonucunun arkasında yer alan mantığı görsel olarak sunabilmektedir. KA'nın tekil bir ağaca olan bağlılığı nedeniyle oluşan dezavantajı RO ile giderilmiştir. DVM, kullandığı kernel fonksiyonları sayesinde doğrusal olmayan problemleri hızlı ve başarılı bir şekilde sınıflandırmaktadır. YSA ise paralel çalışması, hata toleransı ve doğrusal olmayan problemleri hızlı ve başarılı bir şekilde çözmesiyle tercih edilmiştir.

Meta modelin eğitimi ve testi için Çanakkale iline ait satılık konut verileri kullanılmış ve bu veriler güncel olarak kullanılan bir emlak sitesinden belirli bir tarih aralığında web kazıyıcı oluşturularak elde edilmiştir. Web kazıyıcı sayesinde zaman maliyeti düşürülmüş, kullanılan veriler eksiksiz ve hatasız olarak elde edilmiştir. Verilerin belli bir tarih aralığında çekilmesiyle model sonucunun genellenebilir olması sağlanmıştır.

Meta modelden en iyi tahmin performansını almak adına KA, RO, DVM ve YSA yöntemlerine ait hiper parametreler EO ile optimize edilmiştir. Optimizasyon sonucunda oluşan parametreler kullanılarak yöntemler torbalama mantığıyla birleştirilmiş ve ÇO'ya sokulmuştur. Bu sayede tahmin sonucu tek bir yöntemle bırakılmamış, bireysel yöntemlerden elde edilen tahmin sonuçları arasında oylama yapılarak en fazla oy ile tahmin sonucu belirlenmiştir. Ek olarak bireysel modellerden kaynaklı tekil hataların önüne geçilerek meta modelin tahmin başarısını artırılmıştır. Model performansı korelasyon katsayısı ve R^2 ile değerlendirilmiştir.

Tez çalışmasında geliştirilen meta model 0,836 korelasyon katsayısı ve 0,701 R^2 değeriyle en başarılı satılık konut fiyat tahmin sonucu üretmiştir. KA ise 0,702 korelasyon katsayısı ve 0,496 R^2 değeriyle en düşük tahmin performansına sahip yöntem olmuştur. Geliştirilen modelde EO ile gerçekleştirilen hiper parametre optimizasyonunun model performansının başarısına olan etkisini göstermek adına aynı adımlar varsayılan hiper

parametre deęerleriyle de gerekleřtirilmiřtir. Hiper parametre optimizasyonu ile elde edilen deęerler kullanarak gerekleřtirilen her yntem (KA, RO, DVM, YSA, O) varsayılan hiper parametre deęerleri kullanarak oluřturulan yntemlere gre daha bařarılı tahmin sonuları retmiřtir. Elde edilen korelasyon sonuları hiper parametre optimizasyonun bireysel model bazında en dřk %1,27 en yksek %8’lik bir performans artıřı saęladığıını gstermiřtir. Meta model iin bu artıř %3,08’dir. Benzer karřılařtırma R^2 metrięi ile yapıldığında bireysel modellerin performans artıřı %2,74-%16 arası deęiřirken bu artıř meta model iin %6,37 olmuřtur.

Elde edilen sonular, birden fazla farklı M yntemini birleřtiren meta modelin tek modelli yaklařımlara gre bařarım performansının daha yksek olduęunu gstermektedir. Bunun yanında hem alt modeller hem de meta model iin hiper parametre optimizasyonunun tahmin performansını arttırdığı grlmřtir.

Geliřtirilen bu meta model, emlak alanındaki tm paydařlar iin kullanılabilir niteliktedir. Satıcı ve alıcı taraf iin emlaęın belirli zellikleri girildięinde standartlařmıř bir fiyat tahmini yapmanın yanı sıra bankalar, tapu kadastro vb. paydařlar iin de ekspertiz ihtiyaını karřılama potansiyeli bulunmaktadır.

Modelin hali hazırdaki stn bařarısı, gelecek alıřmalarda kullanılacak veri kmesinin daha geniř bir tarih aralıęına yayılması, dviz kurları, borsa endeksleri, gayrisafi milli hasıla vb. ekonomik faktrlerin ve emlaklar iin konum bilgisinin de eklenmesi ile daha da artırılabilir.

KAYNAKÇA

- Abay, R. (2013). "Markowitz karesel programlama ile portföy seçimi: İMKB 30 endeksinde riskli portföylerin seçimi". *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 22 (2), 175-194.
- Akçetin, E. ve Çelik, U. (2014). "İstenmeyen elektronik posta (spam) tespitinde karar ağacı algoritmalarının performans kıyaslaması". *Journal of Internet Applications and Management*, 5 (2), 43-56. 10.5505/iuyd.2014.43531.
- Akşehir, Z. D. ve Kılıç, E. (2019). "Makine öğrenmesi teknikleri ile banka hisse senetlerinin fiyat tahmini". *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 12 (2), 30-39.
- Aktaş, R., Doğanay, M. M. ve Yıldız, B. (2003). "Mali başarısızlığın öngörülmesi: istatistiksel yöntemler ve yapay sinir ağı karşılaştırılması". *Ankara Üniversitesi SBF Dergisi*, 58 (04). 10.1501/SBFder_0000001691.
- Arunadevi, J. ve Nithya, M. J. (2016). "Comparison of feature selection strategies for classification using rapid miner". *International Journal of Innovative Research in Computer and Communication Engineering*, 4 (7), 13556-13563. 10.15680/IJIRCCCE.2016. 0407088.
- Bilişik, M. T. (2011). "Destek vektör makinesi çoklu regresyon ve doğrusal olmayan programlama ile perakendecilik sektöründe gelir yönetimi için dinamik fiyatlandırma". XI. Üretim Araştırmalar Sempozyumu, 23-24 Haziran 2011, İstanbul Ticaret Üniversitesi, İstanbul. 785-799.
- Breiman, L. (2001). "Random forests". *Machine Learning*, 45 (1), 5-32. 10.1023/A:1010933404324.
- Budak, H. ve Erpolat, S. (2012). "Kredi riski tahmininde yapay sinir ağları ve lojistik regresyon analizi karşılaştırılması". *AJIT-e: Bilişim Teknolojileri Online Dergisi*, 3 (9), 23-30. 10.5824/1309-1581.2012.4.002.x.
- Çetin, V. ve Yıldız, O. (2022). "Veri analizinde veri ön işleme teknikleri üzerine kapsamlı bir inceleme". *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 28 (2), 299-312. 10.5505/pajes.2021.62687.

- Çılgın, C., Ünal, C., Alıcı, S., Akkol, E. ve Gökşen, Y. (2020). “Metin sınıflandırmada yapay sinir ağları ile bitcoin fiyatları ve sosyal medyadaki beklentilerin analizi”. *Mehmet Akif Ersoy Üniversitesi Uygulamalı Bilimler Dergisi*, 4 (1), 106-126. 10.31200/makuubd.651904.
- Daş, M., Balpetek, N., Akpınar, E. K. ve Akpınar, S. (2019). “Investigation of wind energy potential of different provinces found in Turkey and establishment of predictive model using support vector machine regression with the obtained results”. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 34 (4), 2203-2213. 10.17341/gazimmfd.432590.
- Ecer, F. (2014). “Türkiye’deki konut fiyatlarının tahmininde hedonik regresyon yöntemi ile yapay sinir ağlarının karşılaştırılması”. In International Conference On Eurasian Economies, 1-2 Temmuz 2014, Skopje, Macedonia. 1-10.
- Goodman, A. C. (1998). “Andrew Court and the invention of hedonic price analysis”. *Journal of urban economics*, 44 (2), 291-298. 10.1006/juec.1997.2071.
- Ion, R. M., Munteanu, D. ve Cocina, G. C. (2009). “Concept of artificial neural network (ANN) and its application in cerebral aneurism with multi walls carbon nanotubes (MWCNT)”. Proceedings of the 10th WSEAS international conference on Neural networks, 23-25 Mart 2009, Prag, Çek Cumhuriyeti. 104-107.
- Kahraman, A. M. ve Özdağlar, D. (2004). “Su dağıtım sistemlerinin genetik algoritma ile optimizasyonu”. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 6 (3), 1-18.
- Kannan, K. S., Manoj, K. ve Arumugam, S. (2015). “Labeling methods for identifying outliers”. *International Journal of Statistics and Systems*, 10 (2), 231-238.
- Kasuya, E. (2018). “On the use of r and r squared in correlation and regression”. *WILEY Ecological Research*, 34 (1), 235-236. 10.1111/1440-1703.1011.
- Kavzoğlu, T. ve Çölkesen, İ. (2010). “Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi”. *Harita Dergisi*, 144 (7), 73-82.
- Keskin, S., Aydın, F. ve Yurdagül, H. (2019). “Eğitsel veri madenciliği ve öğrenme analitikleri bağlamında e-öğrenme verilerinde aykırı gözlemlerin

- belirlenmesi”. *Eğitim Teknolojisi Kuram ve Uygulama*, 9 (1), 292-309. 10.17943/etku.475149.
- Lam, L. ve Suen, S. Y. (1997). “Application of majority voting to pattern recognition: an analysis of its behavior and performance”. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27 (5), 553-568. 10.1109/3468.618255.
- Li, W., Zhao, Y., Meng, W. ve Xu, S. (2009). “Study on the risk prediction of real estate investment whole process based on support vector machine”. 2009 Second International Workshop on Knowledge Discovery and Data Mining, 23-25 Ocak 2009, Moscow, Russia. 167-170.
- Oğuzlar, A. (2003). “Veri ön işleme”. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, (21), 67-76.
- Okumuş, H. ve Aydemir, Ö. (2017). “Random forest classification for brain computer interface applications”. 2017 25th signal processing and communications applications conference (SIU), 15-18 Mayıs 2017, Antalya, Türkiye. 1-4.
- Onan, A. (2021). “Haber metinlerinden sosyo-ekonomik ve epidemiyolojik konuların metin madenciliğine dayalı belirlenmesi”. *Avrupa Bilim ve Teknoloji Dergisi*, 26, 295-300. 10.31590/ejosat.957004.
- Rojarath, A., Songpan, W. ve Pong-inwong, C. (2016). “Improved ensemble learning for classification techniques based on majority voting”. 7th IEEE international conference on software engineering and service science (ICSESS), 26-28 Ağustos 2016, Beijing, China. 107-110.
- Schonlau, M. ve Zou, R. Y. (2020). “The random forest algorithm for statistical learning”. *The Stata Journal*, 20 (1), 3-29. 10.1177/1536867X20909688.
- Schulz, R. ve Werwatz, A. (2004). “A state space model for Berlin house prices: estimation and economic interpretation”. *The Journal of Real Estate Finance and Economics*, 28 (1), 37-57. 10.1023/A:1026373523075.
- Selim, H. (2009). “Determinants of house prices in Turkey: hedonic regression versus artificial neural network”. *Expert systems with Applications*, 36 (2), 2843-2852. 10.1016/j.eswa.2008.01.044.

- Shimizu, C., Takatsuji, H., Ono, H. ve Nishimura, K. G. (2010). "Structural and temporal changes in the housing market and hedonic housing price indices: a case of the previously owned condominium market in the Tokyo metropolitan area". *International Journal of Housing Markets and Analysis*, 3 (4), 351-368. 10.1108/17538271011080655.
- Smola, A. J. ve Schölkopf, B. (2004). "A tutorial on support vector regression". *Statistics and computing*, 14 (3), 199-222. 10.1023/B:STCO.0000035301.49549.88.
- Şen, Z. (2004). Yapay sinir ağları ilkeleri. Su Vakfı: İstanbul.
- Takma, Ç., Atıl, H. ve Aksakal, V. (2012). "Çoklu doğrusal regresyon ve yapay sinir ağı modellerinin laktasyon süt verimlerine uyum yeteneklerinin karşılaştırılması". *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 18 (6), 941-944. 10.9775/kvfd.2012.6764.
- Tanyıldızı, E. ve Demirtaş, F. (2019). "Hiper parametre optimizasyonu hyper parameter optimization". 2019 1st International Informatics and Software Engineering Conference (UBMYK), 6-7 Kasım 2019, Ankara, Türkiye. 1-5.
- Taşar, B., Üneş, F., Demirci, M. ve Kaya, Y. Z. (2018). "Yapay sinir ağları yöntemi kullanılarak buharlaşma miktarı tahmini", *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 9 (1), 543-551.
- Ustalı, N. K., Tosun, N. ve Tosun, Ö. (2021). "Makine öğrenmesi teknikleri ile hisse senedi fiyat tahmini". *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 16 (1), 1-16. 10.17153/oguibf.636017.
- Uysal, M. ve Güyer, T. (2014). "İstatistiksel veri analizine ilişkin genişleyebilir bir karar ağacı tasarımı". *Bilişim Teknolojileri Dergisi*, 7 (3), 33-43. 10.12973/bid.2022.
- Whaley III, D. L. (2005). The interquartile range: theory and estimation. YL Tezi, East Tennessee State University, Mathematics, Tennessee, ABD.
- Xing, H. J. ve Ji, M. (2018). "Robust one-class support vector machine with rescaled hinge loss function". *Pattern Recognition*, 84, 152-164. 10.1016/j.patcog.2018.07.015.
- Yakut, E., Elmas, B. ve Yavuz, S. (2014). "Yapay sinir ağları ve destek vektör makineleri yöntemleriyle borsa endeksi tahmini". *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 19 (1), 139-157.

- Yavuz, S. ve Deveci, M. (2012). “İstatiksel normalizasyon tekniklerinin yapay sinir ađın performansına etkisi”. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 40, 167-187.
- Yazdani, M. (2021). Machine Learning, deep learning, and hedonic methods for real estate price prediction. Eriřim: 28 Mayıs, 2022, <https://arxiv.org/pdf/2110.07151.pdf>
- Yıldız, İ. ve Abut, F. (2022). “Türkiye’de mutluluk düzeyini etkileyen faktörlerin makine öğrenme ve nitelik seçme algoritmaları ile belirlenmesi”. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 11 (1), 1-9. 10.28948/ngumuh.873199.
- Yılmazel, Ö., Afşar, A. ve Yılmazel, S. (2018). “Konut fiyat tahmininde yapay sinir ađları yönteminin kullanılması”. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, 20, 285-300. 10.18092/ulikidince.341584.