



**T.C. İSTANBUL T CARET
 NİVERSİTESİ**

FEN B LİMLERİ ENSTİT S 

**MAKİNE   RENMESİ Y NTEMLERİ İLE İKİNCİ EL ARA LARDA
FİYAT TAHMİNİ**

Esra DERE

Dan şman

Do . Dr. Mustafa Cem KASAPBA I

**Y KSEK LİSANS TEZİ
B LGİSAYAR M HENDİSLİ İ ANAB LİM DALI
İSTANBUL-2023**

KABUL VE ONAY SAYFASI

Esra DERE tarafından hazırlanan "**Makine Öğrenme Yöntemleri ile İkinci El Araçlarda Fiyat Tahmini**" adlı tez çalışması 06/07/2023 tarihinde aşağıdaki jüri üyeleri önünde başarı ile savunularak, İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü **Bilgisayar Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman **Doç. Dr. Mustafa Cem KASAPBAŞI**
İstanbul Ticaret Üniversitesi

Jüri Üyesi **Dr. Öğr. Üyesi Ali AKMAN**
İstanbul Ticaret Üniversitesi

Jüri Üyesi **Doç. Dr. Buket DOĞAN**
Marmara Üniversitesi

Onay Tarihi: **15.09.2023**

İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsünün 15.09.2023 tarih ve 2023/395 numaralı Yönetim Kurulu Kararının 2. maddesi gereğince, ders yüklerini ve tez yükümlülüğünü yerine getirdiği belirlenen "Esra DERE" adlı öğrencinin mezun olmasına oy birliği ile karar verilmiştir.

Prof. Dr. Doğan KAYA
Enstitü Müdürü

AKADEMİK VE ETİK KURALLARA UYGUNLUK BEYANI

İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

15.09.2023

Esra DERE

İÇİNDEKİLER

	Sayfa
İÇİNDEKİLER.....	i
ÖZET	ii
ABSTRACT	iii
TEŞEKKÜR.....	iv
ŞEKİLLER DİZİNİ	v
ÇİZELGELER DİZİNİ	vi
1. GİRİŞ.....	1
2. LİTERATÜR ÖZETİ.....	3
3. MATERYAL VE METHOD.....	5
3.1. Makine Öğrenmesi Kavramı	5
3.2. Makine Öğrenmesi Tarihçesi.....	6
3.3. Makine Öğrenmesi Yöntemleri	7
3.4. Makine Öğrenmesi Uygulama Adımları.....	9
3.5. Regresyon	11
3.5.1. Basit Doğrusal (Linear) Regresyon	12
3.5.2. Çoklu Doğrusal (Linear) Regresyon.....	13
3.6. Rastgele Orman (Random Forest) Algoritması	13
3.7. Gradyan Yükseltme (Gradient Boosting) Algoritması	14
3.8. Ağaç Topluluğu (Tree Ensemble) Algoritması.....	14
3.9. Yapay Sinir Ağları (Artificial Neural Networks)	15
3.10. Knime.....	17
3.10. Veri Setinin Hazırlanması	18
3.12. Çalışma Kapsamı ve Akışı	24
4. ARAŞTIRMA BULGULARI VE TARTIŞMA.....	25
5. SONUÇ VE ÖNERİLER.....	38
KAYNAKLAR	39
ÖZGEÇMİŞ.....	41

ÖZET

Yüksek Lisans Tezi

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE İKİNCİ EL ARAÇLARDA FİYAT TAHMİNİ

Esra DERE

İstanbul Ticaret Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Mustafa Cem KASAPBAŞI

2023, 52 sayfa

Son yıllarda ikinci el araçların pazar hacmi artmıştır. Bu pazarda satıcı ve alıcı için doğru fiyatlandırma oldukça önemlidir. Son kullanıcı veya kurumlar için ikinci el araç değerlemesine ya da kiralanmasında yardımcı olacak sistemsel bir yapıya ihtiyaç bulunmaktadır. Kurulacak bu yapıyla hem araç alacak olan hem de aracını satacak olan kullanıcı, kilometresi, kaza kaydı, kasa tipi, benzin tipi, ağır hasar kaydı gibi fiyata etki edecek özelliklerine göre fiyat çıkarabilir. Bu çalışmada ilgili ikinci el araç ilanların yer aldığı sitelerden Selenium test aracı ile 13.000 üzerinde veri toplanmış ve bu veriler üzerinde veri ön işleme (temizleme, dönüştürme vs.) adımları uygulanmıştır. Makine öğrenme teknikleri KNIME Analytics Platform veri madenciliği programının 4.7.3 sürümünde uygulanarak ikinci el araç fiyatı tahmin edilmeye çalışılmış ve sonuçlar birbiriyle karşılaştırılmıştır. Performans ölçülürken R^2 kullanılmıştır. Doğrusal (Linear) regresyon 0,71 R^2 , Rastgele Orman (Random Forest) 0,85 R^2 , Gradyan artırıcı (GBoosted) 0,83 R^2 , ağaç topluluğu (Tree Ensemble) 0,85 R^2 ve yapay sinir ağları (Artificial Neural Network) 0,72 R^2 tahmin oranları elde edilmiştir. Elde edilen sonuçlarda en düşük tahmin başarısı doğrusal (linear) regresyon, en yüksek başarılı tahmin algoritması 0,85 R^2 ile rastgele orman algoritması olmuştur.

Anahtar Kelimeler: İkinci el araç fiyat tahmini, Knime, makine öğrenmesi.

ABSTRACT

M.Sc. Thesis

PRICE FORECASTING FOR USED CARS WITH MACHINE LEARNING METHODS

Esra DERE

**İstanbul Commerce University
Graduate School of Applied and Natural Sciences
Department of Computer Engineering**

**Supervisor: Assoc. Prof. Dr. Mustafa Cem KASAPBAŞI
2023, 52 pages**

The market volume of second-hand vehicles has increased in recent years. In this market, accurate pricing is very important for the seller and the buyer. There is a need for a systematic structure that will help end-users or institutions in the valuation or leasing of second-hand vehicles. With this structure to be established, both the user who will buy a vehicle and the user who will sell his vehicle can make a price according to the features that will affect the price such as mileage, accident record, body type, gasoline type, heavy damage record. In this study, over 13,000 data were collected from the relevant used car classifieds websites with the Selenium test tool and data preprocessing (cleaning, transformation, etc.) steps were applied on this data. Machine learning techniques were applied in version 4.7.3 of the KNIME Analytics Platform data mining program to predict the used car price and the results were compared with each other. R^2 was used to measure performance. Linear regression 0.71 R^2 , Random Forest 0.85 R^2 , Gradient Boosted 0.83 R^2 , Tree Ensemble 0.85 R^2 and Artificial Neural Network 0.72 R^2 prediction rates were obtained. In the results obtained, the lowest prediction success was linear regression and the highest successful prediction algorithm was the random forest algorithm with 0.85 R^2 .

Keywords: Knime, machine learning, second hand car price estimate.

TEŞEKKÜR

Bu araştırma için beni yönlendiren, karşılaştığım zorlukları bilgi ve tecrübesi ile aşmamda yardımcı olan değerli Danışman hocam Doç. Dr. Mustafa Cem KASAPBAŞI'na, projenin tüm sürecinde yanımda olan değerli meslektaşım Oğuzhan AYDIN'a ve Serkan KIRCA'ya teşekkürlerimi sunarım.

Tezimin her aşamasında beni yalnız bırakmayan aileme sonsuz sevgi ve saygılarımı sunarım.

Esra DERE
İSTANBUL, 2023



ŞEKİLLER

	Sayfa
Şekil 3.1. Makine Öğrenmesi	5
Şekil 3.2. Makine öğrenmesi yöntemleri	7
Şekil 3.3. Denetimli makine öğrenmesi.....	8
Şekil 3.4. Denetimsiz makine öğrenmesi.....	8
Şekil 3.5. Takviyeli makine öğrenmesi.....	9
Şekil 3.6. Makine öğrenmesi uygulama adımları.....	10
Şekil 3.7. Basit doğrusal regresyon.....	12
Şekil 3.8. Rastgele Orman ağaç yapısı.....	14
Şekil 3.9. Tek katmanlı yapay sinir yapısı	16
Şekil 3.10. Çok katmanlı yapay sinir ağ yapısı.....	16
Şekil 3.11. Veri hazırlanması akış şeması.....	18
Şekil 3.12. Selenium ayarları.....	18
Şekil 3.13. Markaların çekilmesi.....	19
Şekil 3.14. Alt markaların çekilmesi.....	20
Şekil 3.15. Detay bilgilerinin çekilmesi	21
Şekil 3.16. Markaların ilan sayı dağılımları	23
Şekil 3.17. Marka–Kilometre fiyat yoğunluk dağılımı	24
Şekil 3.18. Marka–Yıl fiyat yoğunluk dağılımı	25
Şekil 4.1. Sütun Filtresi (Column Filter) düğümü.....	27
Şekil 4.2. Değişkenler hedef değişken üzerindeki dağılımları bileşeni	27
Şekil 4.3. Marka (Brand) – Fiyat (Price) BoxPlot grafiği	28
Şekil 4.4. Kilometre – Fiyat (Price) BoxPlot grafiği	29
Şekil 4.5. Aykırı değerlerin tespit edilmesi.....	29
Şekil 4.6. Kural Motoru (Rule Engine) düğümü	30
Şekil 4.7. Uygulama	30
Şekil 4.8. Veri setinin eğitim test olarak ayrıştırılması.....	31
Şekil 4.9. Doğrusal (Linear) regresyon düğümü.....	32
Şekil 4.10. Rastgele Orman (Random Forest) düğümü.....	33
Şekil 4.11. Gradyan Artırıcı (Gradient Boosted Trees) düğümü	34
Şekil 4.12. Ağaç Topluluğu (Tree Ensemble) düğümü	35
Şekil 4.13. Yapay sinir ağ (Artificial Neural Network) düğümü	36
Şekil 4.14. Yapay sinir ağı öğrencisi ara katman ve nöron ayarları.....	36
Şekil 4.15. Yapay sinir ağı modelinin kolon ve tipleri	37
Şekil 4.16. Yapay sinir ağı nöron ve bias değerleri kısmi gösterimi	37

ÇİZELGELER

	Sayfa
Çizelge 3.1. Ham veri seti.....	22
Çizelge 3.2. Veri seti tanıtımı.....	23
Çizelge 4.1. Doğrusal (Linear) Regresyon sonucu	32
Çizelge 4.2. Rastgele Orman (Random Forest) sonucu	33
Çizelge 4.3. Gradyan Artırıcı (Gradient Boosted Trees) sonucu.....	34
Çizelge 4.4. Ağaç Topluluğu (Tree Ensemble) sonucu	35
Çizelge 4.5. Yapay sinir ağı (Artificial Neural Network) sonucu	38
Çizelge 5.1. Algoritma sonuçları.....	39



1. GİRİŞ

Günümüzde yüksek ÖTV oranları, bayilerde veya galerilerde araç bulunamaması gibi nedenlerden dolayı ikinci el araç satış hacmi Türkiye de ve dünyada artmıştır. Yeni otomobil satış hacmi, ikinci el satış hacminin gerisinde kalmıştır. Bu durum, ikinci el araç piyasasının gün geçtikçe daha da önem kazandığını ortaya koymaktadır. Dünyadan örnek verilecek olursa, ABD ekonomisinde ikinci el araç pazarı, sıfır araç pazar hacminden iki kat daha fazla ve bu fazlalık süreklilik göstermektedir (Lee, 2006). İkinci el araç pazarları, tüketicilere maksimum kalitede uygun fiyatlı araç bulma imkanı sağlamaktadır. Bu noktada aracın gerçek değerini belirlemek hem satıcı hem de alıcı için oldukça önemlidir.

Mevcut sistemde satıcı aracının gerçek değeri konusunda yani kaçta satması gerektiğini, alıcının da aracın ederinin ne olduğu konusunda bilgi sahibi değildir. Aracın fiyatın belirlenmesinde birçok faktör bulunmaktadır. Aracın yakıt türü, vites tipi, markası, modeli, hasar durumu, yaşı ve kaporta durumu gibi özellikleri bu faktörlerden bazılarıdır. Otomotiv sektörü, ekonomide yeri oldukça önemli bir konumda olan bir sektördür. Otomobil, motorlu araç üretim sektöründe 70 gibi büyük bir orana sahiptir. Bu bilgiler ışığında otomobil endüstrisinin, dünya ekonomisindeki payının büyüklüğü anlaşılabilmektedir (Onat, 2007).

Son yıllarda teknolojinin gelişmesiyle beraber veri kolayca toplanabilmektedir. Bu sayede veri depolama sistemlerinde geçmişe göre çok daha fazla veri tutulmaktadır ve depolanmaktadır. Veri madenciliği sayesinde tutulan verilerden anlamlı bilgiler çıkarılmakta ve çıkarılan bu bilgiler sayesinde karar alma süreçlerinde daha verimli hareket edilmektedir (Elshaw ve arkadaşları, 2018).

Bu çalışmada, doğrusal (Linear) regresyon, Rastgele Orman (Random Forest) regresyon, ağaç topluluğu (Tree Ensemble), Gradyan artırıcı (GBoosted) ve yapay sinir ağları (Artificial Neural Network) makine öğrenme algoritmaları kullanılarak KNIME aracı ile ikinci el araç ilanlarından yararlanılarak aracın fiyatı tahmin edilmiştir. İkinci el araç verileri Türkiye'deki ikinci el araç ilanlarının yayınlandığı siteden temin edilmiştir. Ham veri seti veri temizleme, dönüştürme gibi veri ön işleme adımlarından geçmiştir.

Veri önifleme adımından sonra veriler kendi arasında analiz edilmiş ve paylaşılmıştır. Elde edilen sonuçlar karşılaştırılmış ve ikinci el araç fiyatını belirlemedeki en doğru fiyat tahmini veren algoritma belirlenmiştir.

Tezin birinci bölümünde ilgili çalışmaların araştırılmasına yer verilmiştir. İkinci bölümde kullanılan materyal ve metotlar hakkında bilgiler verilmiştir ve veri seti tanıtılmıştır. Üçüncü bölüm olan uygulama bölümünde ise, veri önifleme adımları ve elde edilen analiz sonuçları, KNIME ortamında uygulama adımları detaylandırılmıştır. Dördüncü bölümde, KNIME aracı ile uygulanan makine öğrenmesi algoritmalarının gerçekleştirilmesi ve istatistiksel sonuçlarına yer verilmiştir.



2. LİTERATÜR ÖZETİ

Pazarlioglu ve Gunes, çalışmalarında hedonik fiyat modeli oluşturup üzerine tartışmışlardır. Deneysel analiz sonuçlarını paylaşmışlardır. Bulanık hedonik model ve normal model tahminleri karşılaştırılmıştır. Yüksek düzeyde bilgi sağlayan en iyi bilgi birleştirme yöntemi belirlenmiştir (Pazarlıoğlu ve Günes, 2000).

Asilkan ve Irmak, İkinci el araç fiyatının güncel ve gelecekteki fiyatını tahmin edebilmek için makine öğrenmesi yöntemlerinden Yapay Sinir ağları ile çalışmışlardır. Verileri ikinci el araç satış sitesinden elde etmişlerdir. Yapay sinir ağlarının sonuçlarını zaman serisi analizleri ile elde ettikleri sonuçlarla kıyaslamışlardır. Çalışma sonucunda Yapay sinir ağlarının tahminleme de başarılı ve güvenilir sonuçlar verdiği gözlenmiştir. YSA'nın otomobil sektöründe de kullanımını tavsiye etmişlerdir (Asilkan ve Irmak, 2009).

Namlı ve arkadaşları çalışmalarında, ikinci el araçlarda ikinci el araç satışlarında uygun bir fiyatlandırma yönteminin belirlenip belirlemeyeceğini araştırmışlardır. İstatiksel yöntemlerden olan Doğrusal Regresyon ve makine öğrenme tekniklerinden Destek Vektör Makineleri, Yapay Sinir Ağları kullanmışlardır. Algoritmalar, doğruluğunu kanıtlamak amaçlı çapraz doğrulama ve veri bölme yöntemi ile çalıştırmışlardır. En yüksek başarıyı DVM algoritması ile elde etmişlerdir (Namlı vd., 2018).

Pal ve arkadaşları çalışmalarında, ikinci el araba satış hacminin artmasıyla birlikte araç özelliklerine ve diğer araç fiyatlarına göre fiyat ataması yapabilmeyi amaçlamışlardır. Makine öğrenmesi gözetimli öğrenme yöntemlerinden olan Rastgele Orman (Random Forest) kullandılar. Eğitim için 500 karar ağacı oluşturdular. Eğitim doğruluğunu 95,82 ve test doğruluğunu 83,63 olarak elde ettiler (Pal vd., 2018).

Çelik ve Osmanoğlu, ikinci el araç pazarında hızlı satış yapabilmek, doğru fiyatlandırma yapma amaçlı ideal fiyatı makine öğrenme yöntemleri ile yakalamaya çalışmışlardır. İkinci el otomobile ait 5041 adet veri setindeki 78

değişkenden 23'ü belirleyerek model oluşturdular ve makine öğrenmesi tahminleme modelinin başarısının oranını 81,15 olarak elde ettiler. Daha fazla özniteliğe sahip veri seti ile daha iyi tahmin oranlarının ortaya çıkabileceğini savunmuşlardır (Çelik ve Osmanoğlu, 2019).

Gülmez ve Kulluk, araç alım satımında fiyat belirlemede makine öğrenme tekniklerinin kullanılabilir olduğunu belirtmişlerdir. Çalışmalarında GBT, Doğrusal, İzotonik regresyon, karar ağacı ve Random Forest algoritmaları ve veri seti için Apache Spark kullanmışlardır. Kullanılan beş farklı algoritma içinde fiyat tahmininde en yüksek başarı 0,887 R^2 değeriyle Random Forest algoritmasının olmuştur (Gülmez ve Kulluk, 2023).

Jha ve arkadaşları, cardekho.com'dan veri setini elde ettiler. Python, Flask ve HTML ve makine öğrenmesi kullanarak model oluşturdular. Oluşturdukları modeli makine öğrenmesi algoritmalarından Rastgele orman (Random Forest) algoritması kullanarak ikinci el aracın fiyatını tahmin etmeye çalışmışlardır (Jha vd., 2021).

Gaikwad ve arkadaşları çalışmalarında, doğrusal (linear) regresyon, Kement (Lasso) regresyon, destek vektör makinesi (Support Vector Machine), Rastgele orman (Random Forest) ve makine öğrenimini kullanmışlardır. Performans parametresi olarak R^2 parametresini kullanmışlardır. En yüksek performansı rastgele orman makine öğrenmesi algoritmasından elde etmişlerdir (Gaikwad vd., 2023).

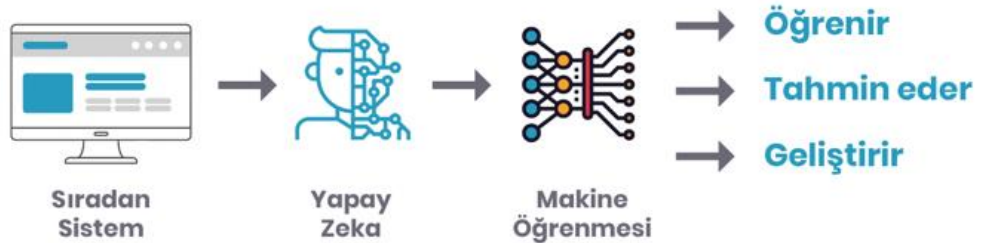
Viswapriya ve arkadaşları çalışmalarında, destek vektör makineleri (Support Vector Machine), yapay sinir ağları (Artificial Neural Networks) ve rastgele orman (Random Forest) algoritmalarını kullanmışlardır. En iyi fiyat tahmini elde ettikleri algoritmayı Java uygulamasına entegre etmişlerdir (Viswapriya vd., 2020).

3. MATERYAL VE METHOD

3.1. Makine Öğrenmesi Kavramı

Son yıllarda kişisel bilgisayar kullanımının artması, kablosuz iletişimin yaygınlaşması, teknolojinin hızla ilerlemesi, e-ticaretin insanın hayatının merkezine yerleşmesi gibi nedenlerden dolayı her alanda, her kullanıcı ciddi bir veri üretmektedir. Üretilen bu verilerden anlamlı bilgiler çıkarıp makinelere öğretmek insan hayatı kolaylaştırılabileceği ortaya konmuştur. Eğitim, reklam, ticaret, sağlık, bankacılık gibi birçok alanda makine öğrenmesi uygulanabilmektedir.

Yapay zekanın alt alanı, istatistiksel ve matematiksel teknikler yardımıyla elde edilmiş verilerden çıkarımlar yaparak bilinmeyen bulmaya çalışan bir veri madenciliği yöntemi olan Makine Öğrenmesi, birçok alanda uygulanmaktadır ve başarılı sonuçlar elde edilmektedir. Makine öğrenmesi geçmiş veri deneyimlerini kullanarak öğrenme ve geliştirme odaklı yeni veriye en uygun modeli istatistiksel ve algoritmalar ile bulmaya çalışır. Temel çalışma prensibi, bilinen veri üzerinden bilinmeyen veri üzerinde çıkarım yapma üzerinedir. Makine öğrenmesinin çalışma mantığı Şekil 3.1’de görüldüğü gibidir. Makine öğrenmesindeki amaç, veri ve deneyimden faydalanarak bilgisayara insan beynine benzer bir öğrenme yeteneği kazandırmaktır. Açıkça kodlamaya gerek olmadan öğrenme yeteneği sağlar.



Şekil 3.1. Makine Öğrenmesi (Turhost, 2021)

3.2. Makine Öğrenmesi Tarihçesi

Edmunda Berkeley, 1949 yılında yayımladığı Büyük Beyinler Düşünen Makineler (Giant Brains or Machine That Think) adlı eserinde, insan beynine benzer yapıda, düşünebilen, bilgiyi hesaplayabilen, çıktı üreten büyük makinelerden bahsetmiştir.

1950'li yıllarda makine öğrenimi izleri görülmeye başlamıştır. İngiliz matematikçi Alan Turing'in "Makinelerin İşleyişi ve Zekâ" isimli makalesinde makine öğrenimini tariflemektedir. Turing'in Taklit oyunu (Turing testi) olarak isimlendirdiği oyunu, bilgisayarın bir insan gibi düşünüp düşünemeyeceğine yönelik bir test atmıştır ortaya. 1952 yılında ise Arthur Samuel, otomatik öğrenmeyi keşfederek makine öğrenimini ortaya çıkarmış ve isimlendirmiştir. İlk makine öğrenmesi uygulaması Dama oyunudur.

Derin öğrenme ağlarının ilk temelleri ise 1965 yılında Matematikçi Alexey Ivakhnenko ve arkadaşları tarafından atılmıştır. En yakın komşu algoritması 1967 yılında oluşturulmuştur.

Denetimli ve denetimsiz makine öğrenmesi terimleri 1990 ve 2000'li yıllarında şekillenmeye başlamıştır. Bu yıllarda büyük veri üzerinde sonuç alabilmek ve öğrenme amaçlı programlar oluşturulmaya başlanmıştır. Schmidhuber, 1993 yılında çok derin öğrenme problemini çözmüştür. 1997'de Deep Blue (IBM'in geliştirdiği bilgisayar), satrançta dünya şampiyonunu yenmiştir. Destek Vektör Makineleri (Support Vector Machine) bu dönemde Vapnik tarafından geliştirilmiştir.

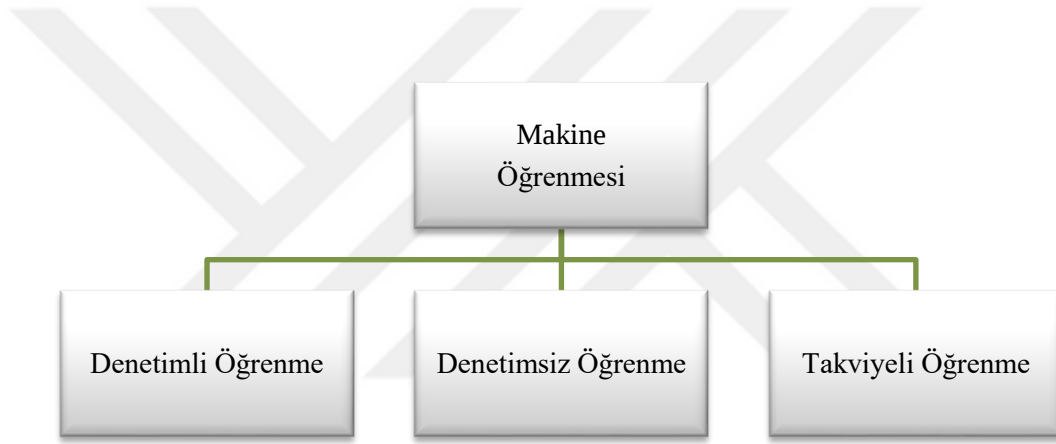
2006 yılında Kinect isimli; metni algılayan, resim veya videolardaki nesneleri algılayabilen hareket algılama cihazı geliştirilmiştir.

Facebook, 2014'te, fotoğraftaki kişileri insan faktörü ile doğrulayabilen DeepFace isimli algoritmayı oluşturmuştur.

Günümüzde de ciddi boyutlarda verinin akışının devam etmesi ve bunların yönetimi adına makine öğrenmesi ve yapay zeka gelişimi devam edecektir.

3.3. Makine Öğrenmesi Yöntemleri

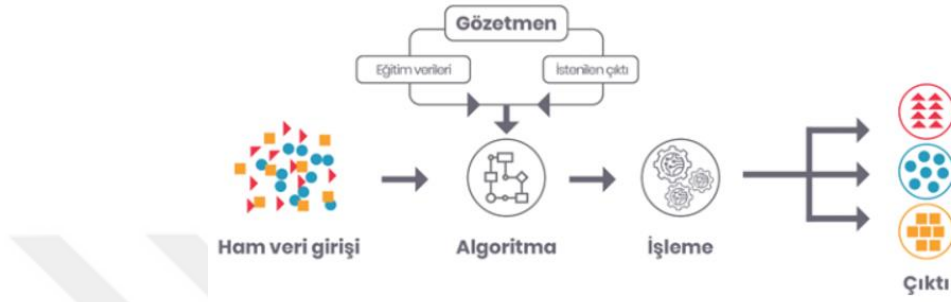
Makine öğrenme yöntemleri, problem tanımına göre uygun teknik ve algoritmayı kullanabilmek için geliştirilmiştir. Problem tanımına göre verinin sınıflandırılması, kümelenmesi veya tahmin edilmesi gerekebilir. Bu problemlere göre farklı algoritmalar bulunmaktadır. Bu algoritmaların uygulanabilmesi için makine öğrenme şekillere göre Şekil 3.2’de gösterilen yöntemlere ayrılır.



Şekil 3.2. Makine öğrenmesi yöntemleri

Denetimli (Supervised) öğrenme, girdi ve çıktının arasındaki ilişkiyi öğrenmek amaçlı etiketli verilerle çalışır. İlişkilerden anlamlı modeller çıkarmayı hedefler. Veri de, her bir girdi için tahmin eden x_i , $i=1,2,3,\dots,n$ değerleri için ise bir çıktı y_i değeri bulunmaktadır. Veri seti, eğitim ve test olarak ayrılır. Eğitim ve test verisi olarak ayırırken en iyi sonucu yakalamaya çalıştığımız için bu oranlar farklılık gösterebilir. Örneğin veri seti %70eğitim, %30 test veya %80 eğitim, %20 test olacak şekilde ayrılabilir. Eğitim için ayırdığımız kısım kullanacağımız gözetimli öğrenme ile makineye öğretilir ve model oluşturulur. Daha önce makinenin tanımadığı test verisi üzerinde bu model uygulanarak en iyi sonuç elde edilmeye çalışılır.

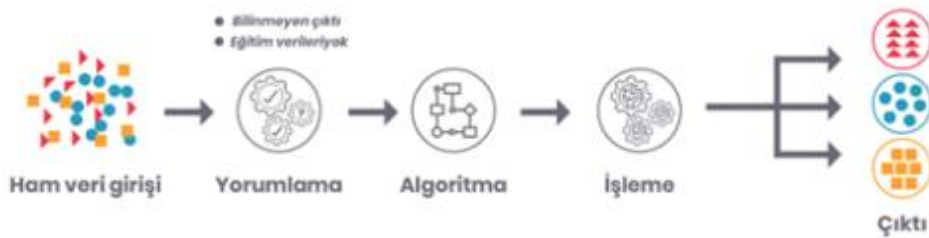
Şekil 3.3'te gözetimli öğrenmenin işleyişi görüldüğü gibidir. Sınıflandırma ve regresyon problemleri için kullanılır. Doğrusal (Linear) regresyon, Yapay Sinir Ağları (Artificial Neural Network), Destek Vektör Makineleri (Support Vector Machine), K-En Yakın Komşu (K-Nearest Neighbours), Naive Bayes Sınıflandırıcısı (Naive Bayes Classifier) algoritmaları gözetimli makine öğrenme tekniklerindendir.



Şekil 3.3. Denetimli makine öğrenmesi (Turhost, 2021)

Denetimsiz (Unsupervised) öğrenme, etiketsiz veri üzerinden algoritmalar kullanarak bağımsız olarak yapıları tahmin etmeye çalışır. Burada her bir girdi x_i ve çıktı değeri olan y_i değeri yoktur. Gözetimsiz öğrenme, veri setindeki verilerin arasındaki benzerliği veya desenleri insan müdahalesi olmadan uzaklık, yakınlık, aralarındaki ilişki gibi faktörlere bakarak keşfeder. Büyük verilerde veri analizi ve keşfinde kullanılabilir. Modelleri kümelendirme, ilişkilendirme ve boyut azaltma gibi işlevleri vardır.

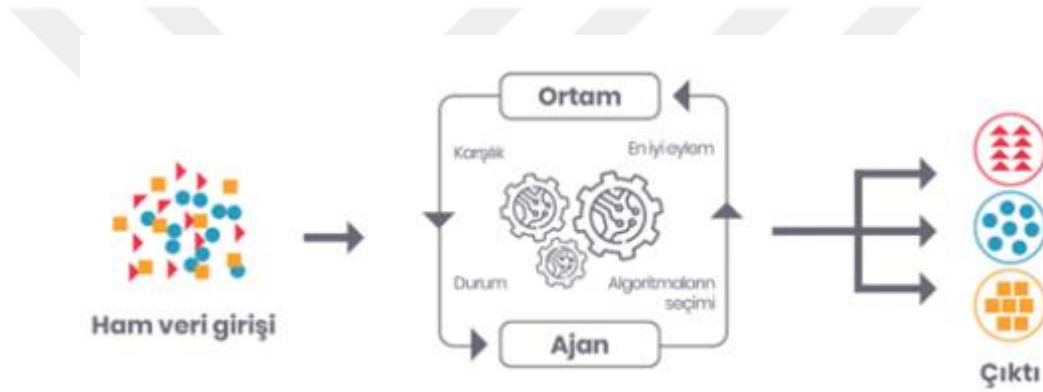
Gözetimsiz öğrenme; sahtekarlık tespiti, öneri sistemleri, doğal dil işleme ve derin öğrenme başlangıcı gibi alanlarda kullanılmaktadır. Şekil 3.4'te gözetimsiz öğrenmenin işleyişi görüldüğü gibidir.



Şekil 3.4. Denetimsiz makine öğrenmesi (Turhost, 2021)

Takviyeli öğrenme ise, oyunlarda, kontrol sistemlerinde veya robotiklerde karar verme veya karmaşık problemlerin çözümünde hangi eylemin en iyi sonuç verdiğini deneme yanılma yöntemini kullanarak çözmeye çalışır. Belirli bir ortamda hedefi en iyi şekilde gerçekleştirmeyi öğrenmeyi hedefler.

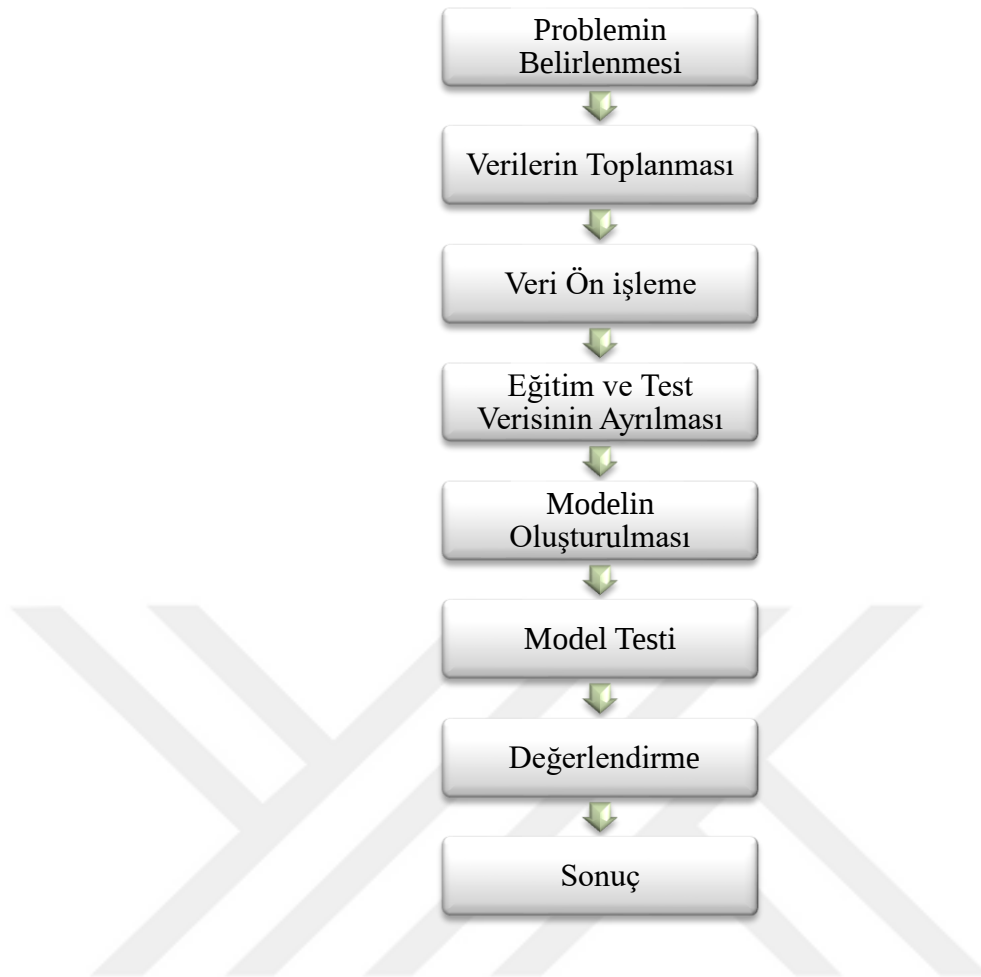
Bir ajan vardır ve bu ajan bir durum karşısında aksiyon seçer ve belirli olan ortamdan buna karşılık ödül ya da ceza alır, yeni aksiyona geçer. Bu adımları tekrarlayarak en iyi aksiyonları seçmeyi öğrenir. Takviyeli makine öğrenmesi, oyun, robotik ve otomasyon, sağlık, otomotiv, reklam ve pazarlama gibi birçok alanda kullanılmaktadır. Şekil 3.5'te takviyeli öğrenme işleyişi görülebilir.



Şekil 3.5. Takviyeli makine öğrenmesi (Turhost, 2021)

3.4. Makine Öğrenmesi Uygulama Adımları

Makine öğrenmesinin uygulanması, birtakım adımlardan oluşmaktadır. Her bir adım oldukça önemli ve doğru çalışması gerekmektedir. Bir sonraki adım bir önceki adımdan beslenmektedir. Şekil 3.6'da makine öğrenmesi uygulamasının genel çerçevesi görüldüğü gibidir.



Şekil 3.6. Makine öğrenmesi uygulama adımları

- Problemin belirlenmesi, uygulamanın en önemli ilk adımıdır. Çalışmanın amacı doğru belirlenmelidir. Probleme yönelik ön çalışma, kullanılacak metaryellerin belirlenmesi ve yol haritası hazırlanarak ilerlenmelidir.
- Verilerin toplanması adımı, çalışma amacına uygun veri toplanır. Problemi tanımına uygun özellikte veri toplanması önemlidir. Veri özelliği kadar veri boyutuda önemli bir kriterdir. Yanlış bilgiler içeren veri seti ilerleyen zamanlarda diğer adımlarında yanlış ilerlemesine sebep olacağından çalışmayı başa döndürebilir.
- Veri ön işleme adımı ise, verilerin çalışılabilir duruma getirilmesi gerekmektedir. Veri setinde eksik veri, farklı veri ölçüm değerleri, aykırı veri, farklı veri tiplerinin olması veri setinin analiz sürecine girmesi gerektiğini gösterir. Eğer veri seti birden fazla kaynaktan elde edilmişse

öncelikle tüm verilerin aynı formata getirilmesi gerekmektedir. Eksik değerler tespit edilmişse, eksik veri analizi uygulanmalı ya da etki derecesine göre veri setinden çıkarma gibi bir yol izlenebilir. Veri setinde aykırı değerler var ise kontrol gereklidir. Tüm bu adımlar doğru bir analiz sürecinin gerekliliğidir.

- Eğitim ve test verisinin ayrılması aşaması ise makineye öğretilecek kısım büyük oranda olacak şekilde eğitim ve test olarak ayrılır. Eğitim için ayrılan veri setiyle model oluşturulur. Oluşturulan model ve makinenin tanımadığı test verisi çalıştırılır. Bu adım eğitim ve test oranını değiştirerek en doğru sonucu alana kadar uygulanabilir. Bu adımın amacı doğru modeli kurup gelecek verilerde bu modeli kullanıp doğru sonuçlar alabilmektir.
- Makine öğrenmesi tekniklerinin uygulandığı bu adımda ise eğitim verisi üzerinde kurulan modelin parametreleri belirlenir. Model, test verisi üzerinde uygulanır. Model başarısı bazı ölçüm kriterlerine göre değerlendirilebilmektedir.
- Model değerlendirme ve yorumlama adımlarında ise, gerçek değerler ile test verisinden elde edilen sonuçlar değerlendirilir. Buradan modelin başarısı ölçülebilmektedir.

3.5. Regresyon

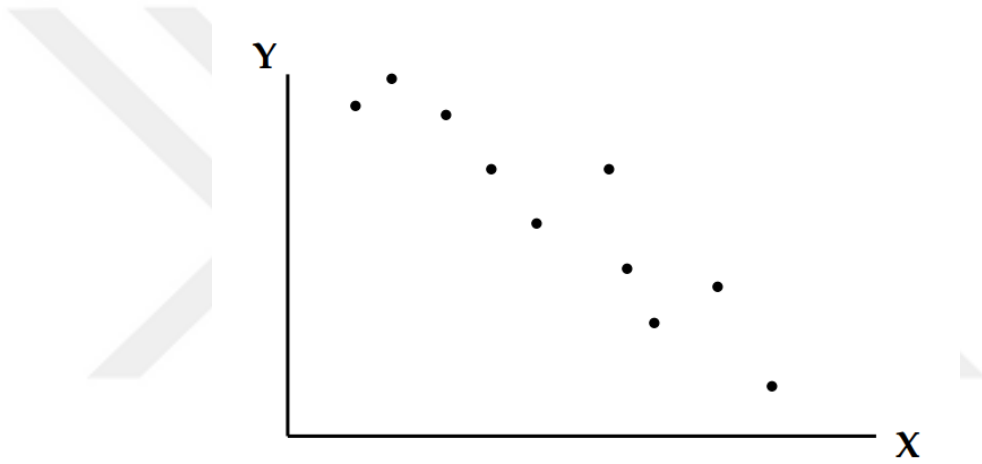
19. yüzyılın sonlarına doğru İngiliz istatistikçi Francis Galton, regresyon kavramını ortaya atmıştır. Galton'un araştırma konusu kalıttır. Anne ve babanın çocuklarının boyu arasındaki bağlantısı üzerine çalışmıştır.

Regresyon analizi belirli ve belirsiz değişkenler arasında bir yakınlık veya uzaklık yakalamaya çalışır. Eğer yakaladıysa ne tür bir uzaklık ya da yakınlık ilişkisi vardır, değişkenlerin birbiri üzerindeki etkisi nedir gibi sorulara cevap bulmaya çalışır.

Regresyonda bağımlı ve bağımsız değişkenler bulunmalıdır. Doğrusal regresyon Basit ve Çok değişkenli olmak üzere iki kısımdan oluşmaktadır. Bu çalışmada, Basit Doğrusal regresyon kullanılmıştır.

3.5.1. Basit Doğrusal Regresyon

Doğrusal ilişkiye sahip bir bağımlı değişken ile bir bağımsız değişken arasında kurulan desene basit doğrusal regresyon denir. Bilinen veriyle bilinmeyen veriyi analiz eder. Belirsiz değişkeni saptayabilmek için belirli değişkenden yola çıkarak tahmin eden bir model oluşturur. Şekil 3.7’de basit doğrusal regresyon grafiği verilmiştir.



Şekil 3.7. Basit doğrusal regresyon

X ve Y arasındaki basit doğrusal regresyon modeli aşağıdaki şekilde formüle edilmiştir.

$$Y_i = \beta_0 + \beta_1 X_i \quad i = 1, 2, \dots, n \quad (1)$$

Yukarıdaki formülde β_0 (kesim noktası) ve β_1 regresyon katsayısı, belirli değişken X ise belirsiz değişken Y üzerindeki değişimi ifade eder.

3.5.2. Çoklu Doğrusal Regresyon

Bir belirli değişkenin birden fazla belirsiz değişken üzerindeki sebep sonuç ilişkisini doğrusal bir modelle açıklar. Belirli değişken üzerinde belirsiz değişkenlerin ne derece etkisi var analiz edilebilir.

X bağımlı değişken $Y_1, Y_2, \dots, Y_p$ 'ler bağımsız değişkenler olmak üzere çoklu regresyon denklemi aşağıdaki şekilde formüle edilmiştir. (formül2)

$$X = \beta_0 + \beta_1 Y_1 + \beta_2 Y_2 + \dots + \beta_p Y_p + \varepsilon$$

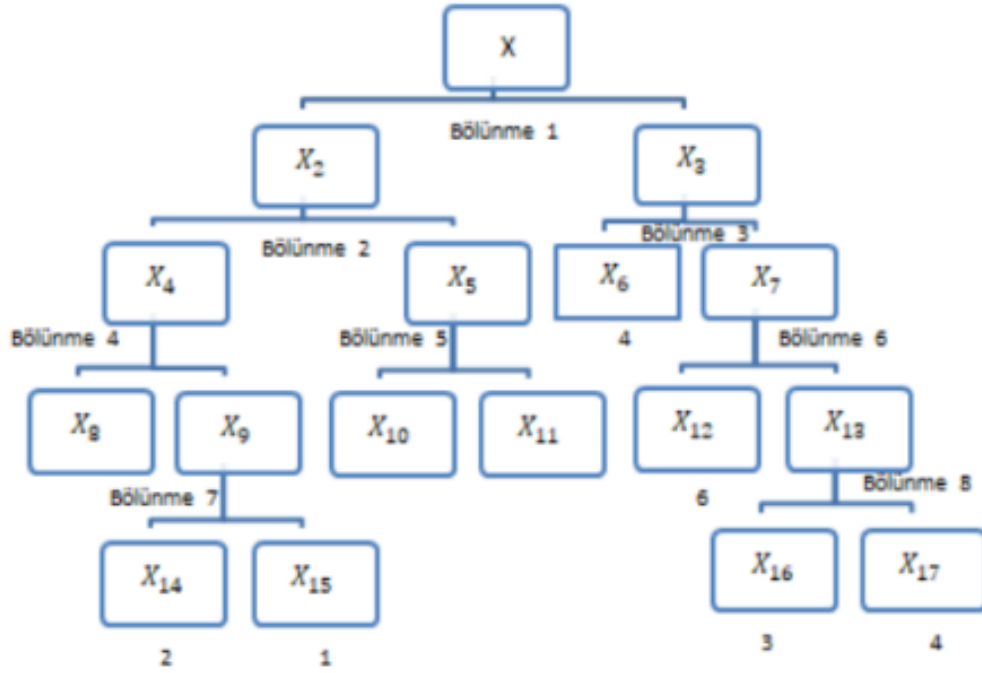
Yukarıdaki denklemde yer alan; β_j , $j = 0, 1, \dots, p$ değerlerine regresyon katsayılarını, Y_j 'deki değişime karşılık gelen X belirli değişkeninin değerini ifade eder.

3.6. Rastgele Orman (Random Forest) Algoritması

2001 yılında Leo Breiman tarafından oylama yapma ile tanıtılan Rastgele Ormanlar yöntemi, hızlı çalışan, yoğun olarak regresyon ve sınıflandırma problemlerinde kullanılan ve başarılı çıktılar veren bir algoritmadır.

Rastgele orman (Random Forest), veriyi rastgele parçalara böler. Birbirinden bağımsız birden fazla karar ağacını bir araya getirir ve ağaçlar arası rastgelelik ve çeşitlilik oluşturarak tahminde bulunmaya çalışır. Rastgele ormanların diğer ağaçlardan farkı çok sayıda karar ağacı oluşturmaktır. Bu ağaçları birleştirerek değerlendirme yapabilme yeteneğine sahiptir.

Şekil 3.8'de altılı sınıf için Rastgele Orman ağaç yapısı gösterilmiştir. Ormandaki tüm ağaçlar eğitim aşamasında en büyük derinliğe kadar büyür. Son tahmin ise tüm bireysel ağaçların oylamasıyla kazanan sınıf belirlenir.



Şekil 3.8. Rastgele orman ağaç yapısı (Kormaz vd., 2018)

3.7. Gradyan Yükseltme (Gradient Boosting) Algoritması

Jerome H. Friedman tarafından bulunan Gradyan Artırıcı (GBoosted) Ağaç, sınıflandırma ve regresyon problemlerinde kullanılan karar ağacı temelli bir algoritmadır. Karmaşık ilişkileri yakalar ve bu ilişkilerle doğru tahminler elde etmeye çalışır. Gradyan Artırıcı Ağaç algoritması ile zayıf tahmin modellerini sıralı olarak birleştirir ve bunlarla güçlü tahmin modellerine dönüştürmeyi hedefler, kusurları düzeltmeye çalışır. Her birleştirilen zayıf tahmin ağacı bir önceki ağacın hatasını en aza indirir. Yüksek tahmin performansına sahiptir.

3.8. Ağaç Topluluğu (Tree Ensemble) Algoritması

Ağaç Topluluğu (Tree Ensemble), 1996 yılında Breiman tarafından tanımlanmıştır. Başarılı bir tahmin modeli oluşturmak için birden fazla karar ağacını birleştirir. Her bir ağaç kendi kararını verir. Sınıflandırma ve regresyon

problemlerinde kullanılır. Başarılı sonuçlar elde edilebilen bir makine öğrenmesi yöntemidir.

3.9. Yapay Sinir Ağları (Artificial Neural Networks)

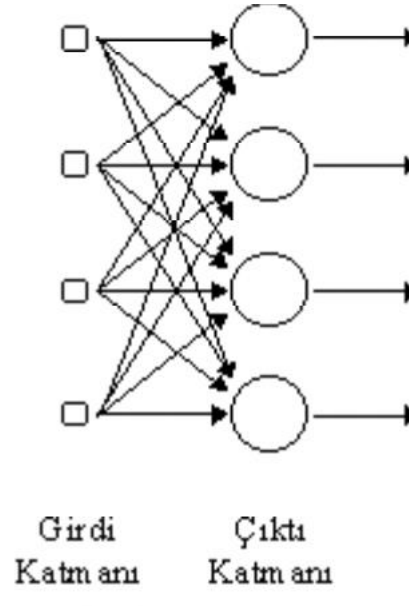
Yapay sinir ağlarına yönelik izler ilk olarak 1980 yılında görülmeye başlanmıştır. 1943'te nörofizyolog Warren McCulloch ve matematikçi Walter Pittsy ilk yapay sinir ağ modelini elektrik devresi olarak tanımlayarak oluşturmuşlardır. 1949 yılında Donald Hebb, The Organization of Behaviour isimli kitabında, yapay sinir ağlarını ileriye taşıyan Hebbian Öğrenme'nin temelinden bahsetmiştir.

Yapay sinir ağları (YSA), insan beyninin yapısından esinlenen, önceden öğrendiği ya da sınıflandırılmış bilgiler yardımıyla yeni bilgiler yaratabilen, keşfedebilen ve karar verme yeteneği olan makine öğrenmesi algoritmasıdır. Yapay sinir ağ çalışmaları ilk olarak nöronların modellenip bilgisayarda programlanması ile başlamıştır. Yapay sinir ağları, görüntü ve ses tanıma, tıp, doğal dil işleme, haberleşme, veri madenciliği veya kontrol gibi birçok alanda kullanılmaktadır. Temel olarak bir yapay sinir ağı, bir veya birden fazla girdisi olabilen uygulanan ağırlıklar sonucu tek bir çıktı üreten bir nöronudur. Problemin karmaşıklığına göre yapay sinir ağındaki katman sayısı değişebilir.

Yapay sinir ağları tek katmanlı ve çok katmanlı olmak üzere ikiye ayrılmaktadır.

➤ Tek Katmanlı Yapay Sinir Ağı

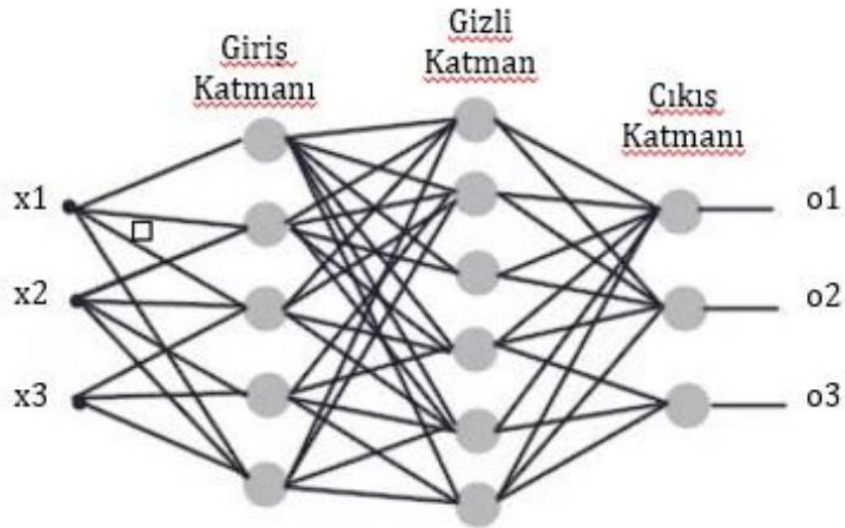
Girdi ve çıktı katmanından oluşan, doğrusal problemlerde kullanılabilen yapay sinir ağıdır. Şekil 3.9'da tek katmanlı yapay sinir ağı gösterilmiştir.



Şekil 3.9. Tek katmanlı yapay sinir ağ yapısı (Aydın, 2005)

➤ Çok Katmanlı Yapay Sinir Ağı

Girdi, gizli ve çıktı katmanlarından oluşan, tek katmanlı yapay sinir ağlarına çözüm olarak geliştirilen yapay sinir ağıdır. Şekil 3.10'da çok katmanlı yapay sinir ağının yapısı verilmiştir.



Şekil 3.10. Çok katmanlı yapay sinir ağ yapısı (Işıksan, 2020)

- Girdi katmanı: Bilgi girişinin olduğu katmandır.
- Gizli (Ara) katmanlar: Girdi katmanından aldığı bilgileri işleyen bu katmanda, doğrusal olmayan karmaşık problemlerin çözümünde birden fazla ara katman kullanılabilir.
- Çıktı katmanı: Girdi katmanından alınan bilginin işlenmesi ile oluşan çıktıya dış dünyaya iletir.

Eğitim esnasında ileri ve geri olmak üzere katmanlar arası geçişler olabilmektedir. Bu geçişler geri ya da ileri yayılım olarak adlandırılmaktadır. İleri yayılımında çıktı ve hata değeri üretilir, geri yayılımında ise hata değerinin düşük çıkması amaçlı ağırlıklar güncellenir.

Bu çalışmada, ikinci el araç fiyat tahmin etmek için çok katmanlı yapay sinir ağı uygulanmıştır.

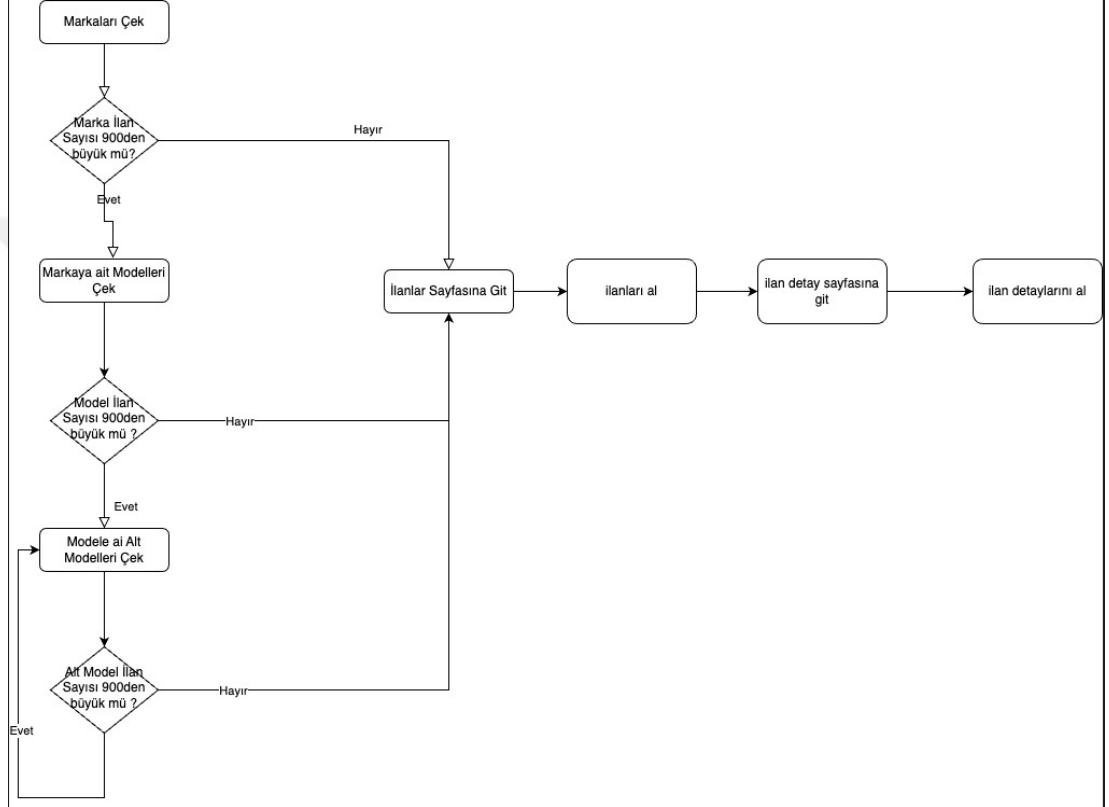
3.10. KNIME

KNIME (Konstanz Information Miner), veriyi analiz etme, görselleştirme, makine öğrenmesi modellemesi, istatistiksel işlevler, tahminleme, kümeleme ve sınıflandırma gibi akışlarımızı yapmamızı sağlayan açık kaynak kodlu bir veri analitiği platformudur. KNIME, veri analiz adımlarının gerçekleştirilebilmesi, verilerin görselleştirilmesi gibi imkanları sağlayan veri keşif ortamıdır. Düğümler aracılığıyla işlemler gerçekleştirilir. Bu düğümler analiz sürecini hızlandırır. Eklenen tüm düğümler sırasıyla çalışır ve her düğümün çıktısı görülebilir. KNIME içerisinde 1500 modül ve birçok örnek bulundurmaktadır.

3.11. Veri Setinin Hazırlanması

Türkiye’deki aktif bir ikinci el araç ilan sitesinde yer alan Mart 2023 ile Haziran 2023 tarihleri arasındaki güncel veriler çekilmiştir.

Veri setinin hazırlanma sürecinin akış şeması Şekil 3.11. de görüldüğü gibidir.



Şekil 3.11. Veri hazırlanması akış şeması

İlgili ilan sitesinden verilerin çekiliş sürecinden MongoDB veritabanına kadar olan yazım adımları aşağıdaki şekilde detaylandırılmıştır.

- Şekil 3.12' de yer alan kod satırında Selenium ayarları yapıldı.

```

Selenium Ayarları

void Init(int x = 0,int y = 0)
{
    new DriverManager().SetUpDriver(new ChromeConfig());
    ChromeOptions options = new ChromeOptions();
    options.AddArgument("--user-agent=Mozilla/5.0 (iPhone; CPU iPhone OS 6_0 like
Mac OS X) AppleWebKit/536.26 (KHTML, like Gecko) Version/6.0 Mobile/10A5376e
Safari/8536.25");
    options.AddAdditionalChromeOption("useAutomationExtension", false);
    options.AddExcludedArgument("enable-automation");
    options.AddArgument("headless");
    driver = new ChromeDriver(options);
    driver.Url = "https://www.sahibinden.com/kategori/otomobil";
    driver.Manage().Window.Size = new Size(375, 667);
    driver.Manage().Window.Position = new Point(x, y);
}

```

Şekil 3.12. Selenium ayarları

- Selenium kütüphanesini kullanarak hedef web sitesine giderek ilanlarda yer alan tüm markaları çekildi. Şekil 3.13'te yer alan kodlarla markalar çekildi. Selenium WebDriver aracılığıyla sayfa kontrol edilmiştir.

```

Markaları Alma

List<BrandEntity> mainBrands = new List<BrandEntity>();
void getBrands(string parentLink = "")
{
    Thread.Sleep(1000);
    var categories = driver.FindElement(By.ClassName("sub-
categories")).FindElements(By.TagName("li"));
    foreach (var category in categories)
    {
        try
        {
            var slug = category.GetAttribute("class").Replace(" branded", "");
            var name =
category.FindElement(By.TagName("a")).FindElement(By.TagName("h2")).Text;
            var count = category.FindElement(By.TagName("a")).GetAttribute("data-
count");

            var jse = (IJavaScriptExecutor)driver;
            string imageContent = (string)jse.ExecuteScript("return
window.getComputedStyle(arguments[0],':before').getPropertyValue('background-
image')", category);
            var imageSrc = imageContent.Replace("url(\"", "").Replace("\")", "");

            slug = parentLink != "" ? parentLink + "-" + slug : slug;
            mainBrands.Add(new BrandEntity { Slug = slug, Count =
int.Parse(count.Replace(".", "")), Name = name, Url = imageSrc, IsReading = false
});
            Console.WriteLine($"{slug} ---> {name} ---> {count} Adet İlan Var -->
{imageSrc}");
        }
        catch
        {
        }
    }
}

```

Şekil 3.13. Markaların çekilmesi

- Her markaya ait olan ilanların sayıları kontrol edildi. Eğer ilan sayısı 900'den fazla ise, o markanın alt markalarının da çekilmesi için markanın detay sayfasına erişilip alt markaların listesi Şekil 3.14'te görüldüğü şekilde elde edilmiştir.

```
Alt Markaları Alma

while (mainBrands.Any(x => x.Count > 900 && x.IsReading == false))
{
    var readingBrands = mainBrands.Where(x => x.Count > 900 && x.IsReading == false).ToList();
    foreach (var readBrand in readingBrands)
    {
        driver.Url = "https://www.sahibinden.com/kategori/" +
readBrand.Slug;
        getBrands(readBrand.Slug);
        readBrand.IsReading = true;
    }
}
```

Şekil 3.14. Alt markaların çekilmesi

- Yukarıda yapılan işlem ilan sayısı kontrol edilerek alt markalarında alt markalarını çekmek için işlemler tekrarlandı. Böylece hiyerarşik yapıdaki tüm alt markalar elde edilmiştir.
- Bu aşamada elde ettiğimiz markaların ilanlarının ve detaylarının çekilebilmesi için her marka özelinde ilan sayfalarına gidip ilanlar tek tek alındı. Alınan veriden başlık, açıklama, araçların tüm teknik özellikleri, fiyatları, resimleri ayrıştırılmıştır. İlanlarda yer alan resim işaretlemesi yapılmış ilanlarda resimdeki verilerden de yararlanarak aracın hasar durumu, boyalı, değişen veya orijinal parçası olup olmadığı bilgileri elde edildi. Şekil 3.15'te detay bilgilerinin çekildiği kodlara yer verilmiştir.



Şekil 3.15. Detay bilgilerinin çekilmesi

- Aracın hasar bilgisi veya kaza bedel bilgileri verilmemişse ilan detaylarından yapay zeka yardımıyla değişen veya orijinal parça tespiti yapılmıştır. Böylece ilanların detayları daha kapsamlı elde edilmiştir.
- Son aşamada ise, ayrıştırarak elde edilen tüm veriler MongoDB veritabanına yazıldı. MongoDB tasarımı marka, alt marka ve ilanlar için koleksiyonlar oluşturularak verilerin düzenli ve tutarlı olması sağlandı.

Ham verilerin ön işleme Python kütüphaneleri ile gerçekleştirilmiştir. Veriler aracın fiyatını belirlemede önemli olan yaşı, kilometre bilgisi, modeli, markası, tramer kaydı, kasa tipi, gövde tipi gibi bilgileri içermektedir. Veri seti çekilirken aynı ilanı çekip çekmediği veri alınırken kontrol edilip veritabanına yazılmamaktadır. Fiyat belirlemede etkili olmayan ham veri seti üzerindeki kriterler çıkarılmıştır. Tüm süreç sonunda en son olarak eksik değer kontrolü, filtreleme gibi işlemler KNIME ortamında uygulanmıştır.

Çizelge 3.1'de gösterilen ham veri setinin bir kısmı gösterilmiştir. Veri seti içerisinde 26.000 üzeri veri ve 28 öznitelik bulunmaktadır. Tabloda aracın markası, modeli, yılı, vitesi tipi, motor gücü, beygir gücü, kimden olduğu, durumu, gövde tipi, kilometre bilgisi, renk gibi bilgiler içermektedir. Ham veri setinde yer

alan açıklama, url, açıklama gibi özellikler fiyat tahmininde etken olmadığı için öznitelikler arasından çıkarılmıştır ve uygulama kısmında detaylandırılmıştır.

Çizelge 3.1. Ham veri seti

YarattığıTarih	Marka	Seri	Model	Yıl	YakıtTipi	VitesTürü	Durumu	KM	GövdeTipi	MotorGücü	Çekis	Renk
21 Şubat 2023	Aion	S	580	2021	Elektrik	Otomatik	İkinci El	14500	Sedan	126 - 150 HP	Önden Çekiş	Beyaz
20 Şubat 2023	Aion	S	580	2021	Elektrik	Otomatik	İkinci El	6900	Sedan	126 - 150 HP	Önden Çekiş	Beyaz
15 Şubat 2023	Aion	S	580	2021	Elektrik	Otomatik	İkinci El	6800	Sedan	126 - 150 HP	Önden Çekiş	Beyaz
07 Şubat 2023	Aion	S	580	2021	Elektrik	Otomatik	İkinci El	6500	Sedan	126 - 150 HP	Önden Çekiş	Beyaz
03 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2011	Benzin	Manuel	İkinci El	94000	Hatchback 5 kapı	170 hp	Önden Çekiş	Gümüş Gri
21 Şubat 2023	Alfa Romeo	Giulietta	1.6 JTD Distinctive	2014	Dizel	Manuel	İkinci El	94000	Hatchback 5 kapı	105 hp	Önden Çekiş	Kırmızı
09 Mart 2023	Alfa Romeo	Giulietta	1.6 JTD Progression Plus	2014	Dizel	Manuel	İkinci El	154000	Hatchback 5 kapı	105 hp	Önden Çekiş	Beyaz
09 Mart 2023	Alfa Romeo	556	2.0 JTS Distinctive	2004	Benzin & LPG	Yan Otomatik	İkinci El	250000	Sedan	165 hp	Önden Çekiş	Siyah
09 Mart 2023	Alfa Romeo	559	1.9 JTD Distinctive Plus	2010	Dizel	Yan Otomatik	İkinci El	207550	Sedan	150 hp	Önden Çekiş	Gri
08 Mart 2023	Alfa Romeo	547	1.6 TS Distinctive	2006	Benzin & LPG	Manuel	İkinci El	169000	Hatchback 5 kapı	120 hp	Önden Çekiş	Siyah
08 Mart 2023	Alfa Romeo	547	1.6 TS Progression	2004	Benzin & LPG	Manuel	İkinci El	295000	Hatchback 5 kapı	120 hp	Önden Çekiş	Kırmızı
08 Mart 2023	Alfa Romeo	Giulietta	1.6 JTD Progression Plus	2015	Dizel	Manuel	İkinci El	249000	Hatchback 5 kapı	105 hp	Önden Çekiş	Beyaz
08 Mart 2023	Alfa Romeo	547	1.6 TS Distinctive	2006	Benzin & LPG	Manuel	İkinci El	235000	Hatchback 5 kapı	120 hp	Önden Çekiş	Gümüş Gri
08 Mart 2023	Alfa Romeo	546	1.4 TS	1998	Benzin & LPG	Manuel	İkinci El	244000	Hatchback 5 kapı	103 hp	Önden Çekiş	Bordo
08 Mart 2023	Alfa Romeo	Giulietta	1.6 JTD Distinctive	2015	Dizel	Manuel	İkinci El	235000	Hatchback 5 kapı	105 hp	Önden Çekiş	Beyaz
08 Mart 2023	Alfa Romeo	556	1.6 TS Distinctive	2004	Benzin	Manuel	İkinci El	246000	Sedan	120 hp	Önden Çekiş	Mavi
08 Mart 2023	Alfa Romeo	556	2.0 TS Executive	1999	Benzin & LPG	Manuel	İkinci El	241500	Sedan	155 hp	Önden Çekiş	Gümüş Gri
08 Mart 2023	Alfa Romeo	546	1.4 TS Ritmo	1999	Benzin & LPG	Manuel	İkinci El	220000	Hatchback 5 kapı	103 hp	Önden Çekiş	Kırmızı
08 Mart 2023	Alfa Romeo	556	1.6 TS Progression	2004	Benzin & LPG	Manuel	İkinci El	287000	Sedan	120 hp	Önden Çekiş	Füme
08 Mart 2023	Alfa Romeo	556	1.6 TS Distinctive	2005	Benzin & LPG	Manuel	İkinci El	234000	Sedan	120 hp	Önden Çekiş	Füme
08 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2013	Benzin	Yan Otomatik	İkinci El	85000	Hatchback 5 kapı	170 hp	Önden Çekiş	Lacivert
08 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2012	Benzin	Yan Otomatik	İkinci El	116000	Hatchback 5 kapı	170 hp	Önden Çekiş	Beyaz
08 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2011	Benzin	Manuel	İkinci El	79000	Hatchback 5 kapı	170 hp	Önden Çekiş	Siyah
03 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2011	Benzin	Manuel	İkinci El	94000	Hatchback 5 kapı	170 hp	Önden Çekiş	Gümüş Gri
21 Şubat 2023	Alfa Romeo	Giulietta	1.6 JTD Distinctive	2014	Dizel	Manuel	İkinci El	94000	Hatchback 5 kapı	105 hp	Önden Çekiş	Kırmızı
09 Mart 2023	Alfa Romeo	Giulietta	1.6 JTD Progression Plus	2014	Dizel	Manuel	İkinci El	154000	Hatchback 5 kapı	105 hp	Önden Çekiş	Beyaz
09 Mart 2023	Alfa Romeo	556	2.0 JTS Distinctive	2004	Benzin & LPG	Yan Otomatik	İkinci El	250000	Sedan	165 hp	Önden Çekiş	Siyah
09 Mart 2023	Alfa Romeo	559	1.9 JTD Distinctive Plus	2010	Dizel	Yan Otomatik	İkinci El	207550	Sedan	150 hp	Önden Çekiş	Gri
09 Mart 2023	Alfa Romeo	559	1.9 JTD Distinctive Plus	2010	Dizel	Yan Otomatik	İkinci El	207550	Sedan	150 hp	Önden Çekiş	Gri
08 Mart 2023	Alfa Romeo	547	1.6 TS Distinctive	2006	Benzin & LPG	Manuel	İkinci El	169000	Hatchback 5 kapı	120 hp	Önden Çekiş	Siyah
08 Mart 2023	Alfa Romeo	547	1.6 TS Progression	2004	Benzin & LPG	Manuel	İkinci El	295000	Hatchback 5 kapı	120 hp	Önden Çekiş	Kırmızı
08 Mart 2023	Alfa Romeo	Giulietta	1.6 JTD Progression Plus	2015	Dizel	Manuel	İkinci El	249000	Hatchback 5 kapı	105 hp	Önden Çekiş	Beyaz
08 Mart 2023	Alfa Romeo	547	1.6 TS Distinctive	2006	Benzin & LPG	Manuel	İkinci El	235000	Hatchback 5 kapı	120 hp	Önden Çekiş	Gümüş Gri
08 Mart 2023	Alfa Romeo	546	1.4 TS	1998	Benzin & LPG	Manuel	İkinci El	244000	Hatchback 5 kapı	103 hp	Önden Çekiş	Bordo
08 Mart 2023	Alfa Romeo	Giulietta	1.6 JTD Distinctive	2015	Dizel	Manuel	İkinci El	235000	Hatchback 5 kapı	105 hp	Önden Çekiş	Beyaz
08 Mart 2023	Alfa Romeo	556	1.6 TS Distinctive	2004	Benzin	Manuel	İkinci El	246000	Sedan	120 hp	Önden Çekiş	Mavi
08 Mart 2023	Alfa Romeo	556	2.0 TS Executive	1999	Benzin & LPG	Manuel	İkinci El	241500	Sedan	155 hp	Önden Çekiş	Gümüş Gri
08 Mart 2023	Alfa Romeo	546	1.4 TS Ritmo	1999	Benzin & LPG	Manuel	İkinci El	220000	Hatchback 5 kapı	103 hp	Önden Çekiş	Kırmızı
08 Mart 2023	Alfa Romeo	556	1.6 TS Progression	2004	Benzin & LPG	Manuel	İkinci El	287000	Sedan	120 hp	Önden Çekiş	Füme
08 Mart 2023	Alfa Romeo	556	1.6 TS Distinctive	2005	Benzin & LPG	Manuel	İkinci El	234000	Sedan	120 hp	Önden Çekiş	Füme
08 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2013	Benzin	Yan Otomatik	İkinci El	85000	Hatchback 5 kapı	170 hp	Önden Çekiş	Lacivert
08 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2012	Benzin	Yan Otomatik	İkinci El	116000	Hatchback 5 kapı	170 hp	Önden Çekiş	Beyaz
03 Mart 2023	Alfa Romeo	Giulietta	1.4 TB MultiAir Distinctive	2011	Benzin	Manuel	İkinci El	94000	Hatchback 5 kapı	170 hp	Önden Çekiş	Gümüş Gri

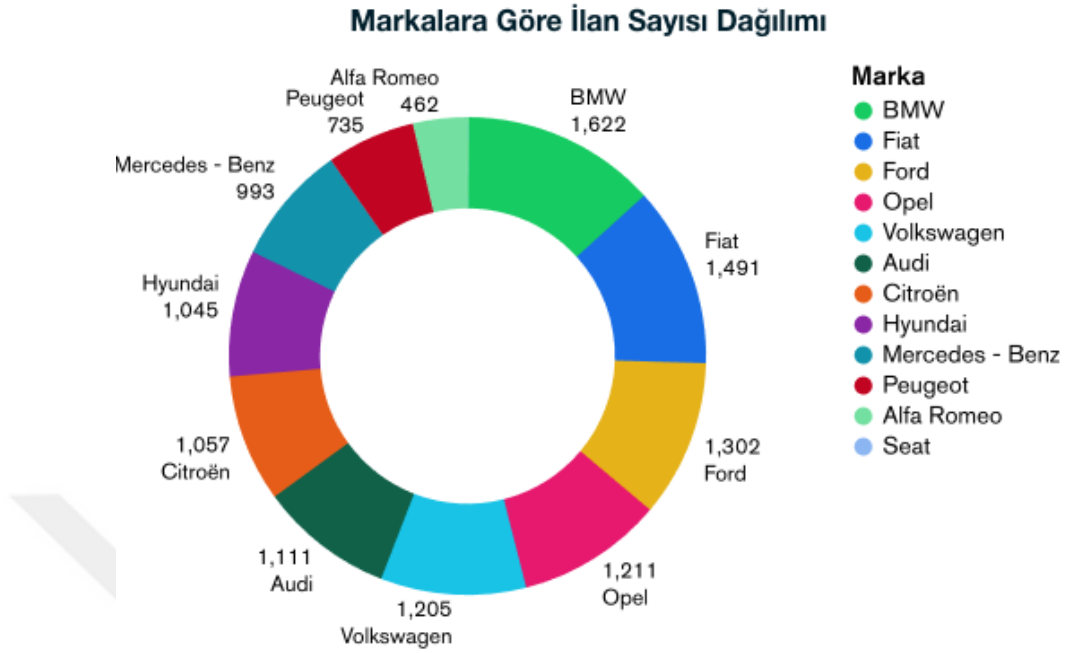
Çizelge 3.2’de veri setinde yer alan özniteliklerin açıklaması verilmiştir.

Çizelge 3.2. Veri seti tanıtımı

Öznitelik	Veri Türü	Açıklaması
KayıtTarihi	Object	İlanın Veri Setine Kayıt Tarihi
Kod	Object	İlanın Kodu
YaratılışTarihi	Object	İlanın Oluşturulma Tarihi
Marka	Object	Aracın Markası
Seri	Object	Aracın Model Serisi
Model	Object	Aracın Modeli
Yıl	Int64	Aracın Üretim Yılı
YakıtTipi	Object	Aracın Yakıt Tipi
Vites Tipi	Object	Aracın Vites Tipi
Durumu	Object	Aracın Durumu(İkinci El/Sıfır)
KM	Float64	Aracın kilometresi
KasaTipi	Object	Aracın Kasa Tipi
MotorGucu	Object	Aracın Motor Gücü
Cekis	Object	Aracın Çekişi
Renk	Object	Aracın Rengi
GarantiliMi	Object	Aracın Garanti Durumu
AgırHasarlıMi	Object	Aracın Ağır Hasarlı Olup/Olmadığı
Plaka	Object	Aracın Plaka Uyuşuğu
TakasDurumu	Object	Aracın Takas Durumu
Fiyat	Float64	Aracın Fiyatı
SatıcıTipi	Object	Aracın Kimden Olduğu
Satıcı	Object	İlan Sahibi Bilgileri
Adres	Object	İlanın Yayınlandığı İl, İlçe, Mahallesi
Açıklama	Object	İlanın Açıklaması
Resim	Object	Aracın Resimleri
Url	Object	İlan Linki

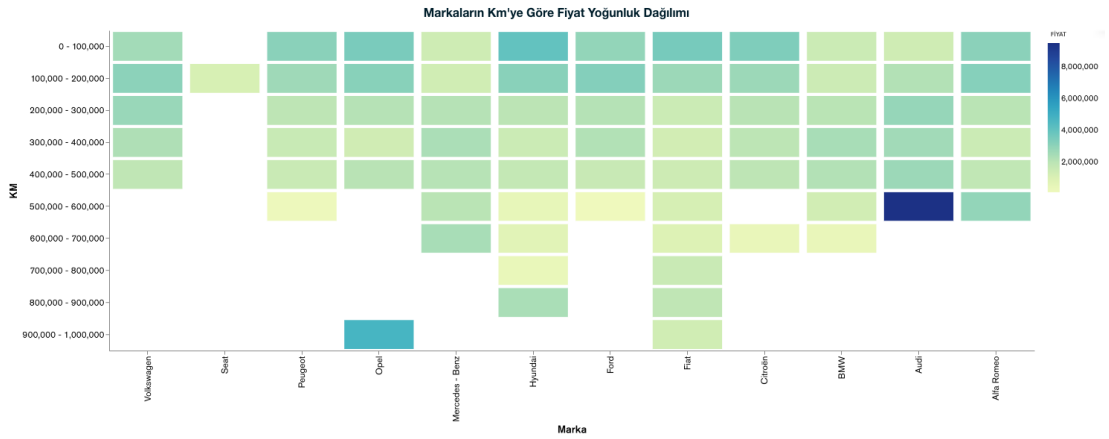
Ham veri setinde veri ön işleme adımından önce fiyat hesaplama da etken olan özniteliklerle keşifsel veri analizi yapılmıştır.

Markaların ilan sayısı gösterimi Şekil 3.16’da görüldüğü gibidir.



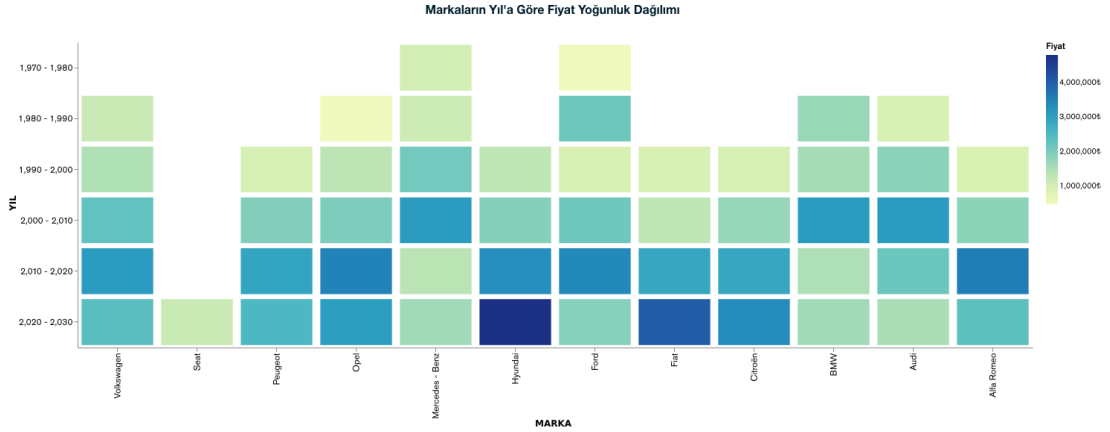
Şekil 3.16. Markaların ilan sayı dağılımları

Şekil 3.17’de markaların kilometreye göre yoğunluk dağılımı verilmiştir. Araç markalarının kilometresi arttıkça fiyatının arttığı gözlemlenmektedir.



Şekil 3.17. Marka-Kilometre fiyat yoğunluk dağılımı

Şekil 3.18'te ise markaların yıl içindeki fiyat yoğunluk dağılımı gözlemlenmektedir.



Şekil 3.18. Marka – Yıl fiyat yoğunluk dağılımı

3.12. Çalışma Kapsamı ve Akışı

Bu çalışmada, ikinci el araç ilanların yer aldığı sitenin ilan verileri kullanılmıştır. Veriler Selenium test aracıyla çekilip Mongo veritabanına yazılmıştır. Bu süreç veri hazırlanması bölümünde detaylı şekilde anlatılmıştır. Hedef değişken olan ücret ile veri setindeki diğer özellikler yorumlanmıştır.

Doğrusal (Linear) regresyon, Rastgele Orman (Random Forest), Gradyan Yükseltme (Gradient Boosting), Ağaç Topluluğu (Tree Ensemble) ve yapay sinir ağları (Artificial Neural Network) makine öğrenme algoritmaları KNIME aracı kullanılarak uygulanmıştır. Elde edilen sonuçlara sonuçlar kısmında yer verilmiştir.

4. ARAŞTIRMA BULGULARI VE TARTIŞMA

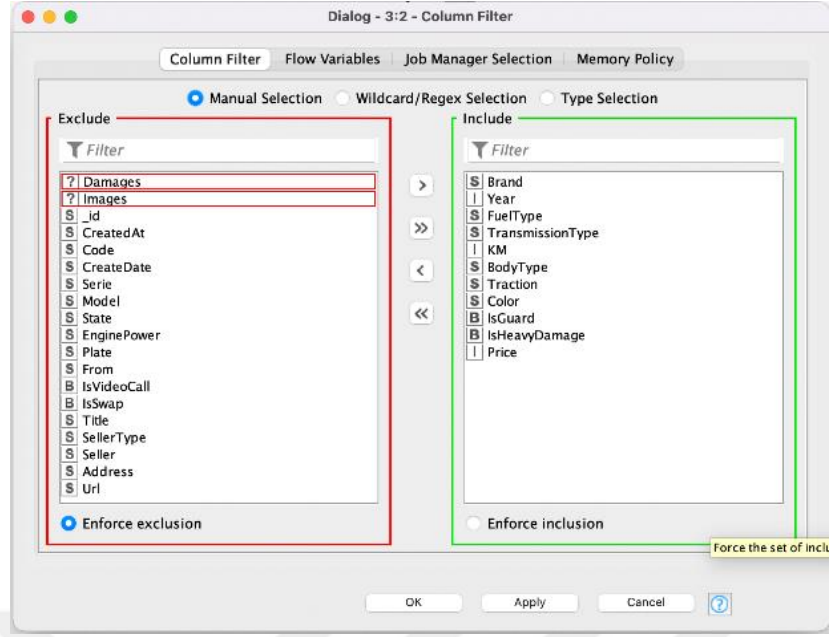
Türkiye’deki e-ticaret sitesinin ilanlarından elde edilen ham veri seti KNIME ortamında eksik veriler silinerek veri önışleme adımından geçmiştir. Veri önışleme adımıdan sonra aracın değerlemesinde etkili olan bağımsız özniteliklerle bağımlı öznitelik olan hedef değışken fiyat arasındaki ilişkiler incelenmiştir.

Çalışmanın amacı, ikinci el araç pazarında yer alan verilerinden yararlanılarak makine öğrenme teknikleri yardımıyla aracın gerçek fiyatını tahmin edebilmektir. Bu uygulamayla beraber hem alıcının hem de satıcının aracının gerçek fiyatı hakkında bilgi sahibi olabileceklerdir.

Uygulamada; dosya okuyucu (File Reader), sütun filtresi (Column Filter), veri önışleme için hazırlanan “Değışkenler hedef değışken üzerindeki dağılımları” isimli bir bileşen, veri hazırlama (Data Preparation), alan hesaplayıcı (Domain Calculater), birden çoğa (One to Many), normalleştirici (Normalizer), bölümleme (Partitioning), doğrusal (Linear) regresyon, rastgele orman (Random Forest), Gradyan artırıcı (GBoosted), ağaç topluluğu (Tree Ensemble) ve yapay sinir ağı (MLP) düğümleri kullanılmıştır.

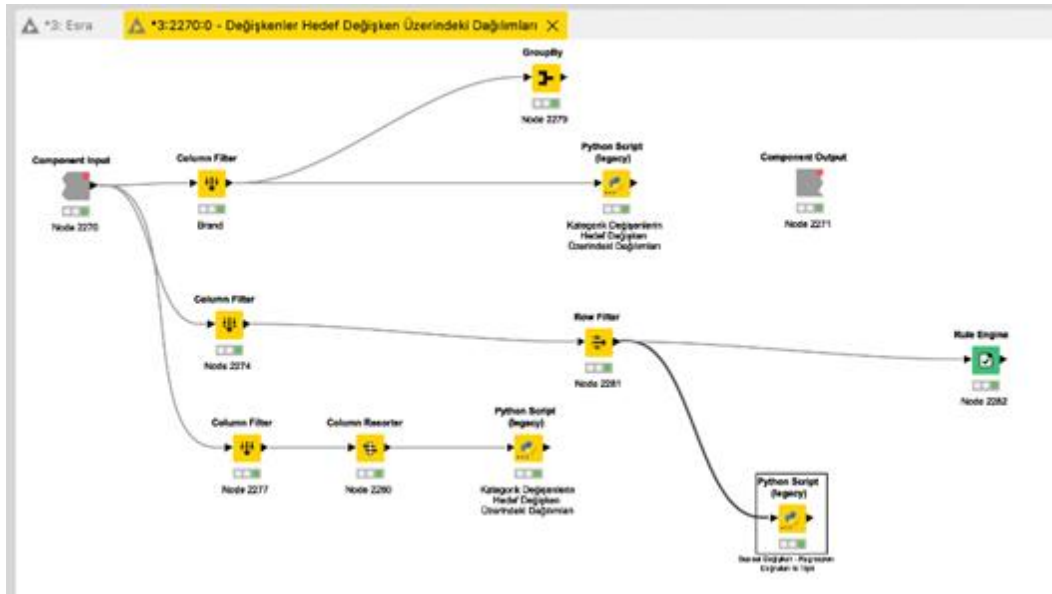
KNIME ortamında, excel olarak çıktısı alınmış ham veri seti dosya okuyucu (File Reader) düğümü kullanılarak ortama yüklenir. Dosya okuyucuya sağ tıklanarak yüklenen veriler gözlemlenebilir. Veri setinde yer alan özniteliklerin hepsi fiyata etki etmediğı için fiyat tahminini etkilemeyecek özniteliklerin kaldırılması için sütun filtresi (Column Filter) düğümü uygulanır.

Şekil 4.1’de sütun filtresi (Column Filter) düğümü görüldüğü gibidir. Hedef değışkene etki edecek marka (brand), yıl (year), benzin tipi (fuelType), vites tipi (transmissionType), kilometre(km), gövde tipi (body type), çekiş gücü (traction), renk (color), garantili mi (isGuard), ağır hasarlı mı (isHeavyDamage) ve fiyat (price) öznitelikleri seçilir.



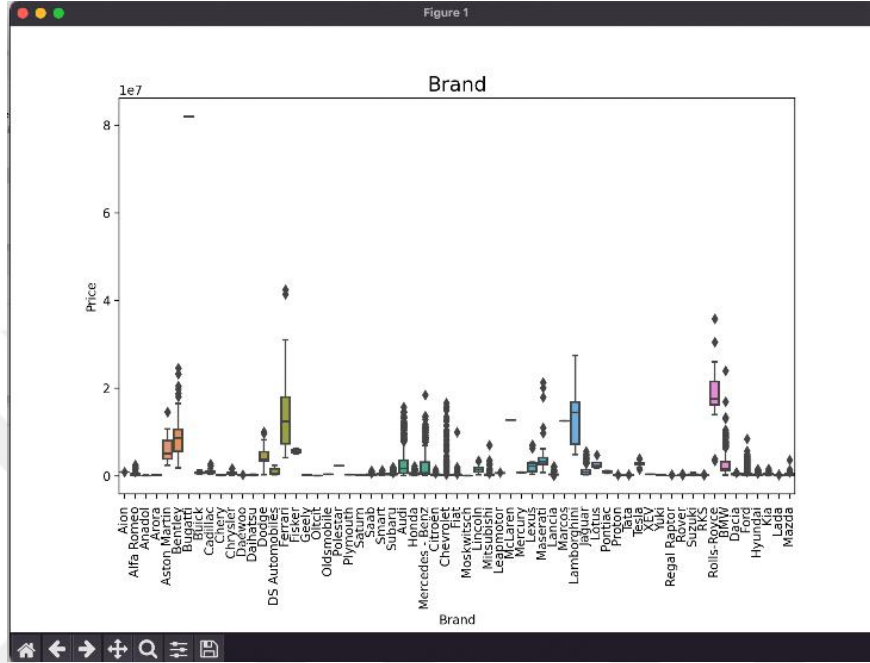
Şekil 4.1. Sütun Filtresi (Column Filter) düğümü

Sütun filtresi çalıştırıldıktan sonra özniteliklerin hedef değişken üzerindeki dağılımı incelemek üzere ve kural motoru (Rule Engine) düğümü kullanılarak yeni öznitelik oluşturmak için Şekil 4.2’de yer alan “Değişkenler hedef değişken üzerindeki dağılımları” isimli bileşen oluşturulur.



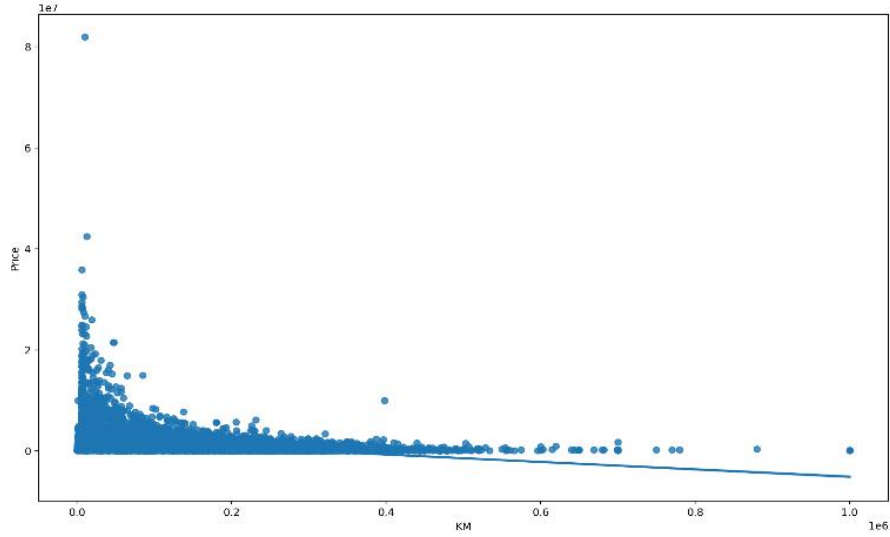
Şekil 4.2. Değişkenler hedef değişken üzerindeki dağılımları bileşeni

Burada ilk sütun filtresinde marka ve fiyat kolonları alınarak aralarında ilişki Python kütüphanesi olan BoxPlot ile grafikselleştirilmiştir. Şekil 4.3'te yer alan grafik incelendiğinde fiyatları birbirine yakın araç markalarının gruplanabilir olduğu gözlenmektedir.



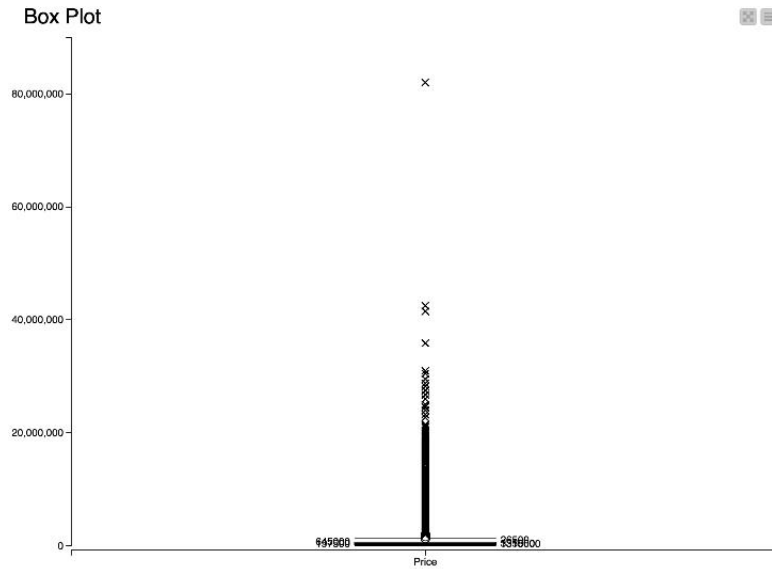
Şekil 4.3. Marka (Brand) – Fiyat (Price) BoxPlot grafiği

İkinci sütun filtresin de km ve fiyat kolonları alınmıştır. Kilometre kolonundaki eksik veri kontrolü için satır filtresi (Row Filter) düğümü çalıştırılmış, kilometre-fiyat arasındaki ilişki Python kütüphanesi olan BoxPlot ile grafikselleştirilmiştir. Çıkan sonuca göre aracın kilometresi artıkcça araç değerinin düştüğü Şekil 4.4'te görüldüğü gibidir.

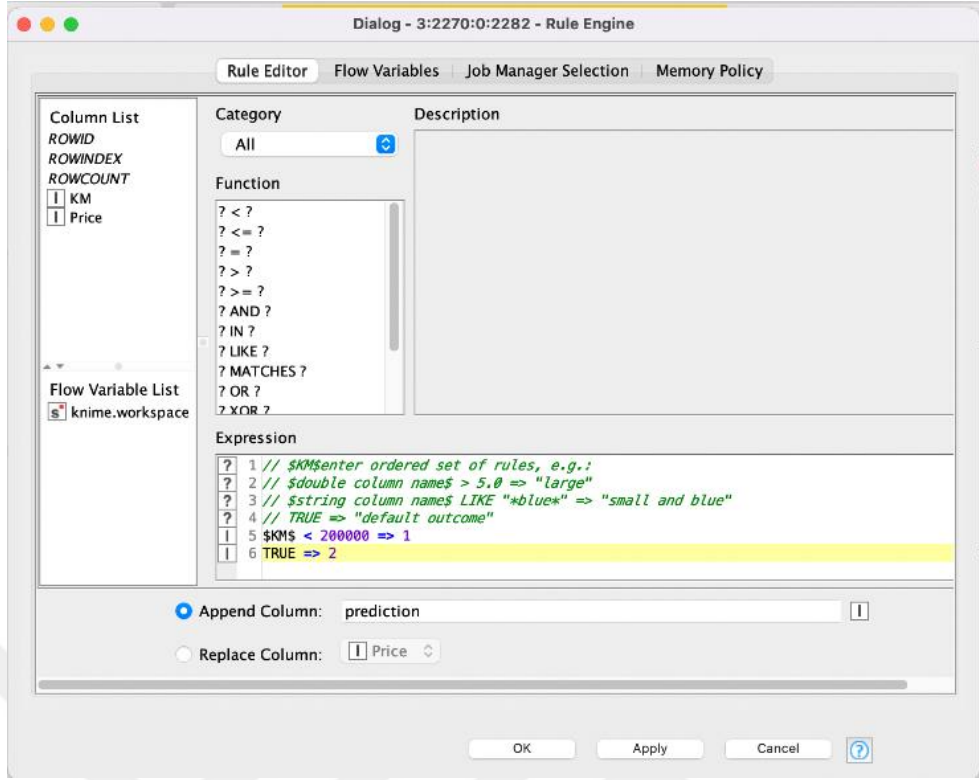


Şekil 4.4. Kilometre – Fiyat (Price) BoxPlot grafiği

İlk aşamada yer alan sütun filtresi (Column Filter) düğümü adımından sonra 'Box Plot' düğümü bağlanıp çalıştırıldığında veri setinde aykırı değerler(outliers) tespit edilmiştir. Şekil 4.5'te aykırı değer tespiti görülebilir. Aykırı değerler kural motoru (Rule Engine) düğümü ile fiyat tahminine etkisi en aza indirgenmiştir. Şekil4.6'da kural motoru (Rule Engine) düğümü görülebilir.

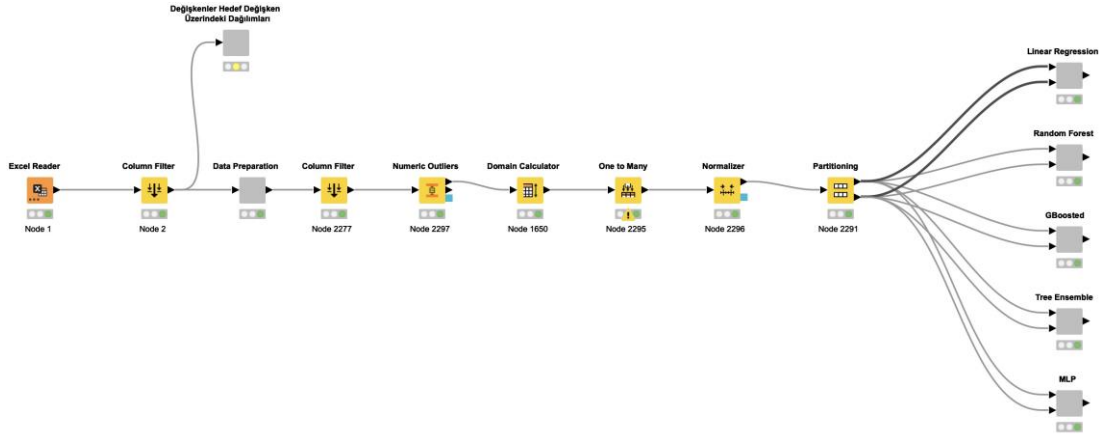


Şekil 4.5. Aykırı değerlerin tespit edilmesi



Şekil 4.6. Kural Motoru (Rule Engine) düğümü

Şekil 4.6’da yer alan kural motoru çalıştırıldıktan sonra Şekil 4.7’de yer alan düğümler sırasıyla çalıştırılır.



Şekil 4.7. Uygulama

- Data Preparation düğümü:

Verilerin hazırlanma kısmıdır. Eksik, kirli ve gürültülü verilerin temizlenme aşamasıdır.

- Numeric Outliers düğümü:

Ücret değerlerinde aykırı değerleri tespit eder ve işleme dahil etmez.

- Domain Calculator düğümü:

Bellekteki verileri temizler.

- One to Many düğümü:

Veri setindeki değerleri kategorikten nümerik değerlere çevirir.

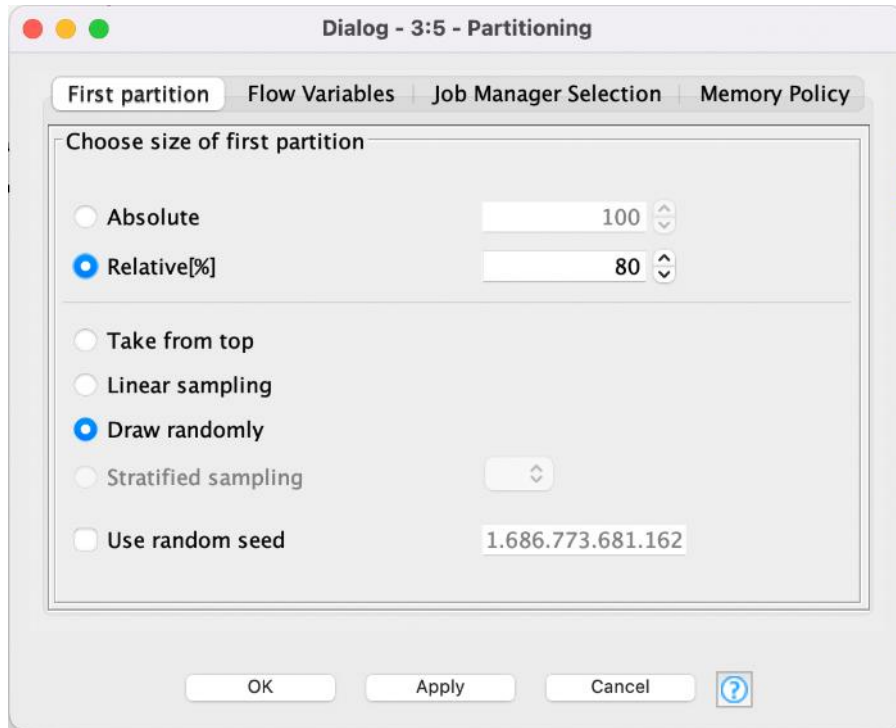
- Normalizer düğümü:

Sayının bulunan minimum değer ile farkını alır. Sayının o kolondaki minimum ve maksimum sayı farkına böler ve normalize eder.

- Partitioning düğümü:

Verilerin eğitim ve test olarak ayrılmasını sağlar.

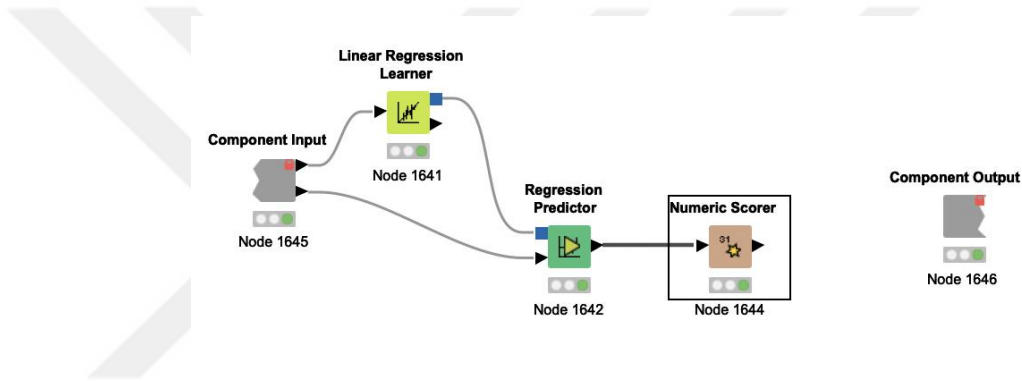
Şekil 4.8'de veri setinin eğitim ve test olarak ayrıştırılmasına ilişkin görseli verilmiştir.



Şekil 4.8. Veri setinin eğitim test olarak ayrıştırılması

Veri eğitim ve test datasına ayrıştırıldıktan sonra metot ve yöntemlerde anlatılan algoritmalar sırasıyla uygulanmıştır.

- ❖ Doğrusal (Linear) Regresyon algoritmasının düğümü Şekil 4.9'da görüldüğü gibidir. Doğrusal regresyon uygulama aşamasında, dosya okuyucu (File Reader), bölümlleme (Partitionig), doğrusal regresyon öğrencisi (Lineer Regression Learner), regresyon tahmincisi (Regression Predictor), sayısal puanlayıcı (Numeric Scorer) ve dağılım grafiği (Scatter Plot) düğümleri kullanılmıştır.



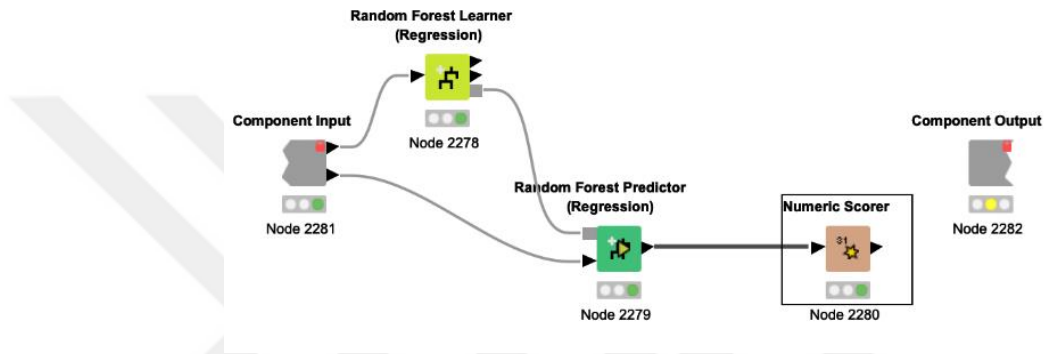
Şekil 4.9. Doğrusal (Linear) regresyon düğümü

- Çizelge 4.1'de Doğrusal (Linear) regresyonun istatistiksel çıktısı görülebilir.

Çizelge 4.1. Doğrusal (Linear) Regresyon sonucu

R^2	0.712
Ortalama Mutlak Hata (MAE)	83,923.731
Ortalama Karesel Hata (MSE)	15,257,528,194.531
Kök Ortalama Kare Hatası (RMSE)	123,521.367
Ortalama İşaretli Fark (MSD)	8,929.414
Ortalama Mutlak Yüzde Hata (MAPE)	0.319
Ayarlanmış (Adjusted)	0.712

- ❖ Rastgele Orman (Random Forest) düğümü Şekil 4.10'da görüldüğü gibidir. Algoritmanın uygulama aşamasında, dosya okuyucu (File Reader), bölümlleme (Partitionig), rastgele orman regresyon öğrenicisi (Random Forest Regression Learner), rastgele orman regresyon tahmincisi (Random Forest Predictor), sayısal puanlayıcı (Numeric Scorer) ve dağılım grafiği (Scatter Plot) düğümleri kullanılmıştır.



Şekil 4.10. Rastgele Orman (Random Forest) düğümü

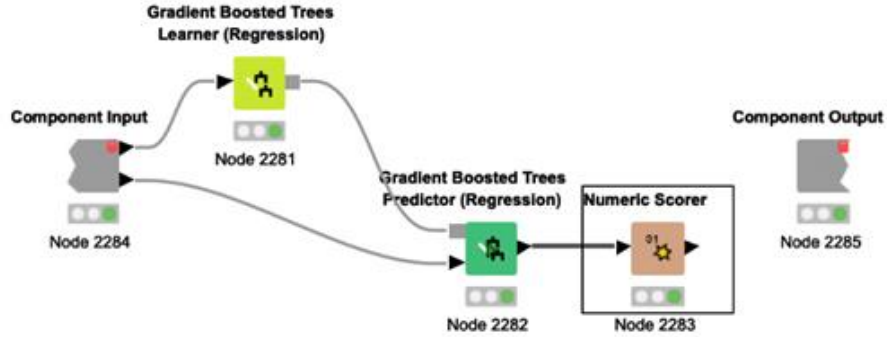
- Çizelge 4.2'de Rastgele Orman algoritmasının istatistiksel çıktısı görülebilir.

Çizelge 4.2. Rastgele Orman (Random Forest) sonucu

R^2	0.858
Ortalama Mutlak Hata (MAE)	57,279.448
Ortalama Karesel Hata (MSE)	7,547,078,427.062
Kök Ortalama Kare Hatası (RMSE)	86,873.923
Ortalama İşaretli Fark (MSD)	3,874.844
Ortalama Mutlak Yüzde Hata (MAPE)	0.213
Ayarlanmış (Adjusted)	0.858

- ❖ Gradyan Artırıcı (Gradient Boosted Trees) algoritmasının düğümü Şekil 4.11'de görüldüğü gibidir. Algoritmanın uygulama aşamasında, dosya okuyucu (File Reader), bölümlleme (Partitionig), Gradyan Artırıcı öğrenicisi (Gradient Boosted Trees Learner), Gradyan Artırıcı tahmincisi

(Gradient Boosted Trees Predictor), sayısal puanlayıcı (Numeric Scorer) ve dağılım grafiği (Scatter Plot) düğümleri kullanılmıştır.



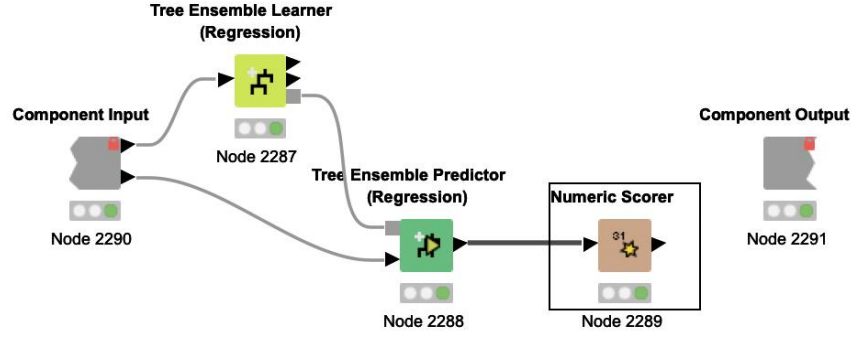
Şekil 4.11. Gradyan Artırıcı (Gradient Boosted Trees) düğümü

- Çizelge 4.3'te Gradyan Artırıcı (Gradient Boosted Trees) algoritmasının istatistiksel çıktısı görülebilir.

Çizelge 4.3. Gradyan Artırıcı (Gradient Boosted Trees) sonucu

R^2	0.834
Ortalama Mutlak Hata (MAE)	58,349.986
Ortalama Karesel Hata (MSE)	8,816,276,228.194
Kök Ortalama Kare Hatası (RMSE)	93,895.028
Ortalama İşaretli Fark (MSD)	-1,123.601
Ortalama Mutlak Yüzde Hata (MAPE)	0.198
Ayarlanmış (Adjusted)	0.834

- ❖ Ağaç Topluluğu (Tree Ensemble) algoritmasının düğümü Şekil 4.12'de görüldüğü gibidir. Algoritmanın uygulama aşamasında, dosya okuyucu (File Reader), bölümlleme (Partitionig), ağaç topluluğu öğrenicisi (Tree Ensemble Learner), ağaç topluluğu tahmincisi (Tree Ensemble Predictor), sayısal puanlayıcı (Numeric Scorer) ve dağılım grafiği (Scatter Plot) düğümleri kullanılmıştır.



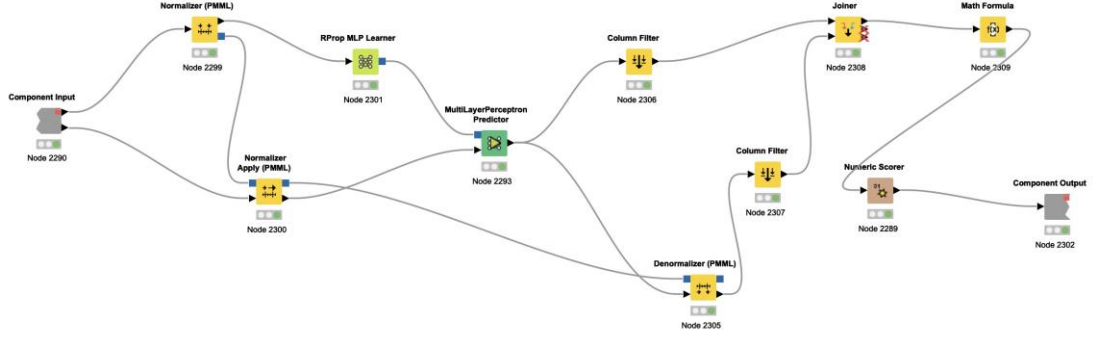
Şekil 4.12. Ağaç Topluluğu (Tree Ensemble) düğümü

- Çizelge 4.4'te Ağaç Topluluğu (Tree Ensemble) algoritmasının istatistiksel çıktısı görülebilir.

Çizelge 4.4. Ağaç Topluluğu (Tree Ensemble) sonucu

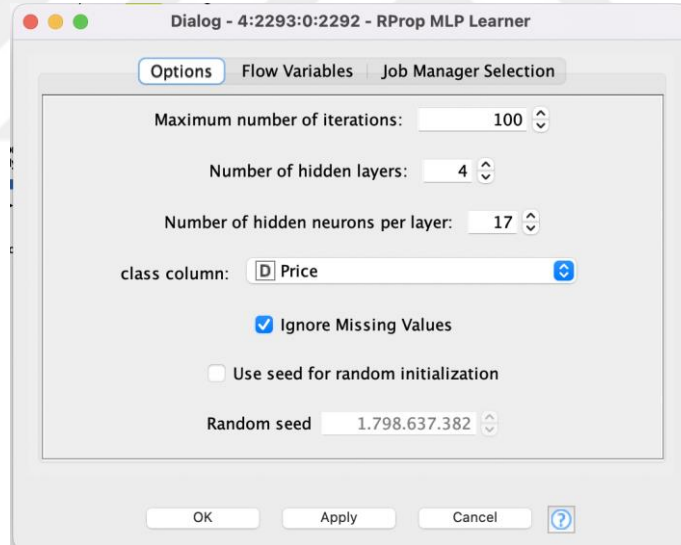
R^2	0.855
Ortalama Mutlak Hata (MAE)	57,533.699
Ortalama Karesel Hata (MSE)	7,694,663,102.183
Kök Ortalama Kare Hatası (RMSE)	87,719.229
Ortalama İşaretli Fark (MSD)	3,979.874
Ortalama Mutlak Yüzde Hata (MAPE)	0.215
Ayarlanmış (Adjusted)	0.855

- Yapay sinir ağının uygulama düğümü Şekil 4.13'te görüldüğü gibidir. Uygulama aşamasında, dosya okuyucu (File Reader), normalleştirici (Normalizer), normalleştirici uygulayıcı (Normalizer Apply), yapay sinir ağı öğrencisi (RProp MLP Learner), yapay sinir ağı tahmin edicisi (MultiLayer Perceptron Predictor) ve sayısal puanlayıcı (Numeric Scorer) kullanılmıştır.



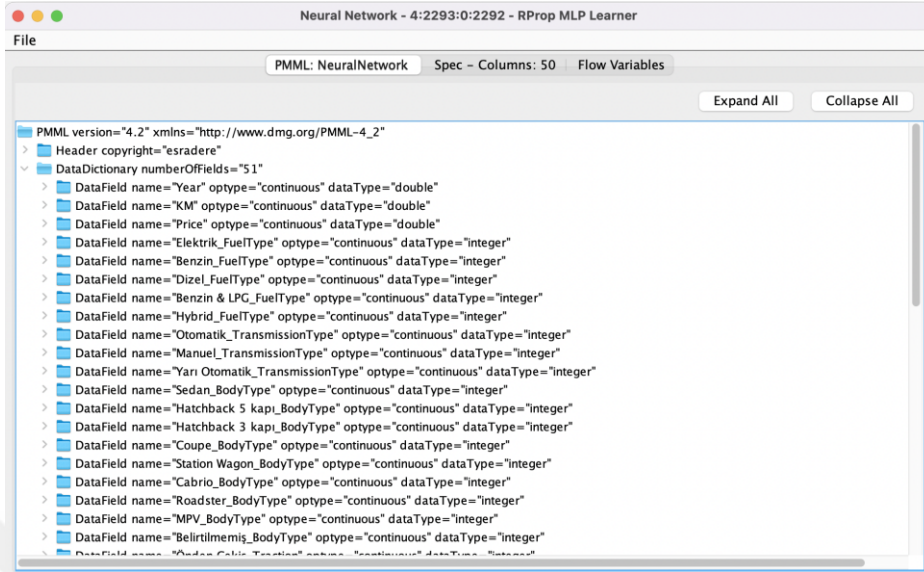
Şekil 4.13. Yapay sinir ağ (Artificial Neural Network) düğümü

Burada hedef değişkenimiz olan fiyat (price) özneliği normalize edildikten sonra yapay sinir ağı öğrenicisine bağlanmıştır. Yapay sinir ağı öğretilirken ara katman sayısı ve nöron sayıları değiştirilip denenerek başarı elde edilmeye çalışılmıştır. Bu düğümde Şekil 4.13'te yapay sinir ağı öğrenicisine verilen ara katman sayısı ve nöron değerleri görülmektedir.



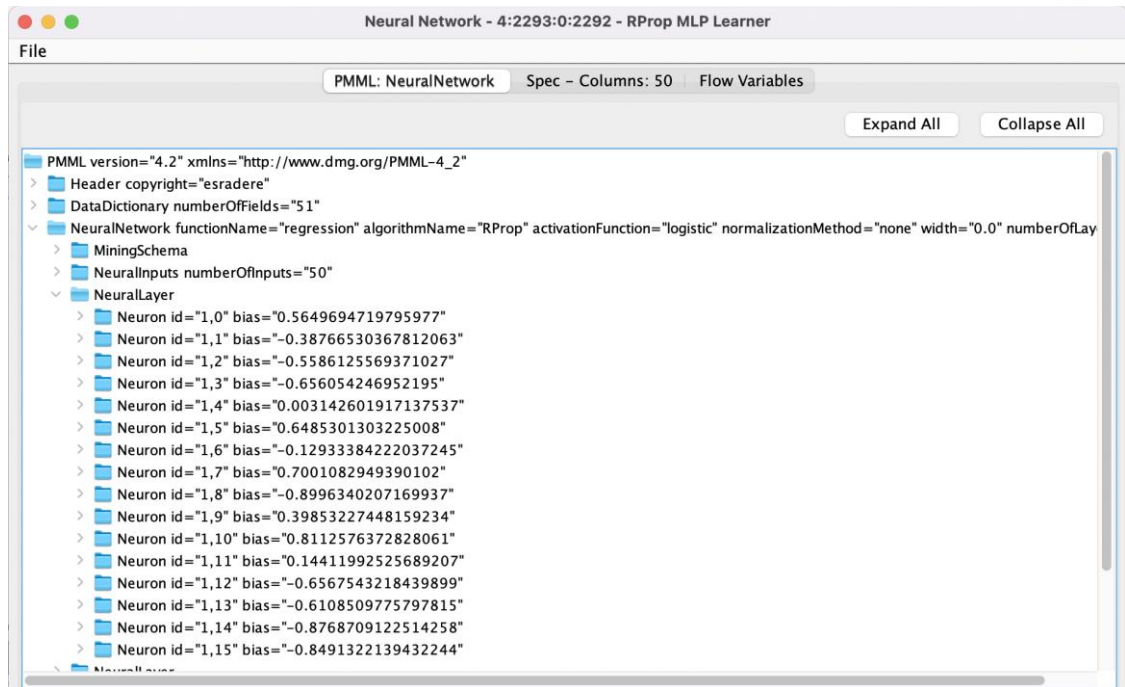
Şekil 4.14. Yapay sinir ağı öğrenicisi ara katman ve nöron ayarları

Şekil 4.15'te yapay sinir ağına uygulanan modelin kolonları ve kolonların tiplerinin detayı görülmektedir.



Şekil 4.15. Yapay sinir ağı modelinin kolon ve tipleri

Şekil 4.16'da yapay sinir ağına uygulanan modelde yer alan nöron ve bias (çıkı değerini kontrol etmeye yarayan sabit) değerlerinin bir kısmı gösterilmiştir.



Şekil 4.16. Yapay sinir ağı nöron ve bias değerleri kısmi gösterimi

- Çizelge 4.5'te yapay sinir ağı (Artificial Neural Network) düğümünün istatistiksel çıktısı görülebilir.

Çizelge 4.5. Yapay sinir ağı (Artificial Neural Network) sonucu

R^2	0.727
Ortalama Mutlak Hata (MAE)	78,041.211
Ortalama Karesel Hata (MSE)	14,455,346,057.32
Kök Ortalama Kare Hatası (RMSE)	120,230.387
Ortalama İşaretli Fark (MSD)	6,706.975
Ortalama Mutlak Yüzde Hata (MAPE)	0.305
Ayarlanmış (Adjusted)	0.727

5. SONUÇ VE ÖNERİLER

Bu çalışmada elde edilen ham veri, veri ön işleme adımları ile temizlenmiş ve analiz edilmiştir. İkinci el aracın fiyatını tahmin etmek için makine öğrenmesi yöntemlerinden Doğrusal (Linear) Regresyon, Rastgele Orman (Random Forest), Gradyan Artırıcı (GBoosted), Ağaç Topluluğu (Tree Emsemble) ve yapay sinir ağları (Artificial Neural Network) yöntemleri kullanılmıştır. Çalışma, KNIME Analytics Platform veri madenciliği programının 4.7.3 sürümünde uygulanmıştır. Uygulanan algoritmaların başarısını ölçmek için R^2 performans metriği kullanılmıştır. Doğrusal (Linear) regresyon 0,71 R^2 ile en düşük tahmin oranını vermiştir. Rastgele Orman (Random Forest) 0,85 R^2 en yüksek tahmin oranını vermiştir. Gradyan Artırıcı (GBoosted) 0,83 R^2 ve Ağaç Topluluğu (Tree Emsemble) 0,85 R^2 ve yapay sinir ağları 0,72 R^2 oranları elde edilmiştir.

İkinci el araç fiyat tahmininde en iyi performansa sahip rastgele orman algoritması olarak belirlenmiştir. Algoritmaların R^2 , ortalama mutlak hatası (Mean Absolute Error), ortalama karesel hatası (Mean Squared Error) ve kök ortalama kare hata (Root Mean Squared Error) çıktıları çizelge 5.1’de verilmiştir.

Çizelge 5.1. Algoritmaların istatistiksel sonuçları

Algoritmalar	R^2
Doğrusal (Linear) Regresyon	0.71
Rastgele Orman (Random Forest)	0.85
Gradyan Destekli (GBoosted)	0.83
Ağaç Topluluğu (Tree Ensemble)	0.85
Yapay Sinir Ağları (Artificial Neural Network)	0.72

Öneri olarak ikinci el araç satışlarında hem alıcıya hem de satıcıya gerçek fiyatlar sunmayı amaçlayan bu çalışma daha fazla veri setiyle geliştirilebilir. Mobil uygulama olarak çalışması devam eden bu çalışma da ilanların gerçek satış fiyatına ulaşarak tahmini fiyat karşılaştırılıp uygulamanın performansı ölçülebilir. Aynı zamanda yapay sinir ağı uygulanırken nöron ve bias değerleri denenerek daha iyi bir tahmin sonucu alınabilir.



KAYNAKLAR

- Asilkan, Ö., Irmak, S., “İkinci El Otomobillerin Gelecekteki Fiyatlarının Yapay Sinir Ağları ile Tahmin Edilmesi”, Suleyman Demirel University The Journal of Faculty of Economics and Administrative Sciences, No.2 pp.375-391, 2009
- Aydın, Ö., 2005. Erişim Tarihi: 11.07.2023.
<http://dspace.trakya.edu.tr:8080/xmlui/bitstream/handle/trakya/427/0039476.pdf?sequence=1&isAllowed=y>
- Çelik, Ö., Osmanoğlu, Ö., “Prediction of The Prices of Second-Hand Cars”, European Journal of Science and Technology, No. 16, pp. 77-83, August 2019
- Ecer, Fatih, “Türkiye’de 2. El Otomobil Fiyatının Tahmini ve Fiyat Belirleyicilerinin Tahmini”, Anadolu Üniversitesi Sosyal Bilimler Dergisi, 13(4), 101-112, 2013
- Elshawi R., Sakr S., Talia D., Trunfio P., Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service, Big Data Research, 14, 1-11, 2018
- Gaikwad, D. A., Suwarnakar, P. S., Mahajan, Y. R., Petkar, A.U., Theurkar, S. G., “Used Car Price Prediction Using Random Forest Algorithm”, International Journal for Multidisciplinary Research (IJFMR), Volume 5, Issue 3, May-June, 2023
- Gülmez B., Kulluk, S., “Analysis and price prediction of secondhand vehicles in Türkiye with big data and machine learning techniques”, Journal of the Faculty of Engineering and Architecture of Gazi University, 38:4, 2279-2289, 2023
- Jha, A., Singh, R., Manish., Saifi, I., Srivastava, S., “Used car price prediction”, International Journal of Advance Research, Ideas and Innovations in Technology, Volume 7, Issue 4-V7I4-1434, 2021
- Kormaz, D., Çelik, H. E., Kapar, M., “Sınıflandırma ve Regresyon Ağaçları ile Rastgele Orman Algoritması Kullanarak Botnet Tespiti: Van Yüzüncü Yıl Üniversitesi Örneği”, Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi, (3): 297-307, 2018
- Lee J., Empirical Analysis of Wholesale Used Car Auctions, A Dissertation in University of California, 2006, 120-121.
- Medium, 2020. Erişim Tarihi: 04.08.2023.
<https://isikhanelif.medium.com/multi-layer-perceptron-mlp-nedir-4758285a7f15>

- Namlı, E., Ünlü, R., Gül, E., “Fiyat Tahminlemede Makine Öğrenmesi Teknikleri ve Doğrusal Regresyon Yöntemlerinin Kıyaslanması; Türkiye’de Satılan İkinci El Araç Fiyatlarının Tahminlenmesine Yönelik Bir Vaka Çalışması”, Konya Mühendislik Bilimleri Dergisi, c.7, s.4, ss. 806-821, 2019
- Onat, M. G. Otomotiv Sektöründe Oranlar Yöntemi Aracılığı ile Finansal Analiz, Yüksek Lisans Tezi, Marmara Üniversitesi, İstanbul 2007
- Pal, N., Arora, P., Kohli, P., Sundararaman, D., & Palakurthy, S. S., “How Much Is My Car Worth? A Methodology for Predicting Used Cars’ Prices Using Random Forest”, In Future of Information and Communication Conference (pp. 413-422). Springer, Cham, 2018
- Pazarlıoğlu, M. V. & Gunes, M. “The Hedonic Price Model for Fusion on Car Market”, International Conference of Information Fusion, Paris, France, 4-13, <http://ieeexplore.ieee.org/document/862707/>, (Erişim: 4.08.2023) ,2000
- Turhost, 2021. Erişim Tarihi: 23.06.2023. <https://www.turhost.com/blog/makine-ogrenmesi-machine-learning-nedir/#serp>.
- Viswapriya, S. E., Sharma, D. S. S., Kiran, G. S., “Vehicle Price Prediction using SVM Techniques”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-9 Issue-8, June, 2020

ÖZGEÇMİŞ

Adı Soyadı : Esra DERE

Eğitim Durumu

Lise : Anadolu Lisesi, 2008

Lisans : Fırat Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü

Yüksek Lisans : İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı

Dere, E., Kasapbaşı, M.C., 2023. İkinci El Araç Değerlemesinde Makine Öğrenme Tekniklerinin Uygulanması, İstanbul Ticaret Üniversitesi Teknoloji ve Uygulamalı Bilimler Dergisi (Yayın sürecinde).