



T.C.
Uludağ Üniversitesi
Fen Bilimleri Enstitüsü

SAC MALZEME SATIN ALMA SÜREÇLERİNDE
VERİ MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ
UYGULAMALARI

Seray MİRASÇI

Yüksek Lisans Tezi

SAC MALZEME SATIN ALMA SÜREÇLERİNDE
VERİ MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ
UYGULAMALARI

Seray MİRASÇI



T.C.
BURSA ULUDAĞ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**SAC MALZEME SATIN ALMA SÜREÇLERİNDE VERİ MADENCİLİĞİ VE
MAKİNE ÖĞRENMESİ UYGULAMALARI**

Seray MİRASÇI
0000-0003-4654-6474

Doç. Dr. Aslı AKSOY
(Danışman)

YÜKSEK LİSANS
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI

BURSA – 2023
Her Hakkı Saklıdır

TEZ ONAYI

Seray Mirasçı tarafından hazırlanan “Sac Malzeme Satın Alma Süreçlerinde Veri Madenciliği Ve Makine Öğrenmesi Uygulamaları” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Bursa Uludağ Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı’nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman : Doç. Dr. Aslı AKSOY

Başkan : Doç. Dr. Aslı AKSOY
0000-0002-2971-2701
Bursa Uludağ Üniversitesi,
Mühendislik Fakültesi,
Endüstri Mühendisliği Anabilim Dalı İmza

Üye : Prof. Dr. Seda ÖZMUTLU
0000-0002-2744-2744
Bursa Uludağ Üniversitesi,
Mühendislik Fakültesi,
Endüstri Mühendisliği Anabilim Dalı İmza

Üye : Doç. Dr. Olcay POLAT
0000-0003-2642-0233
Pamukkale Üniversitesi,
Mühendislik Fakültesi,
Endüstri Mühendisliği Anabilim Dalı İmza

Yukarıdaki sonucu onaylarım

Prof. Dr. Hüseyin Aksel EREN
Enstitü Müdürü
.././.....

Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- 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 üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

09/01/2023

Seray MİRASÇI

TEZ YAYINLANMA FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezin/raporun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma izni Bursa Uludağ Üniversitesi'ne aittir. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet hakları ile tezin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları tarafımıza ait olacaktır. Tezde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederiz.

Yükseköğretim Kurulu tarafından yayınlanan “**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**” kapsamında, yönerge tarafından belirtilen kısıtlamalar olmadığı takdirde tezin YÖK Ulusal Tez Merkezi / B.U.Ü. Kütüphanesi Açık Erişim Sistemi ve üye olunan diğer veri tabanlarının (Proquest veri tabanı gibi) erişimine açılması uygundur.

Doç. Dr. Aslı AKSOY
09.01.2023

Seray MİRASÇI
09.01.2023

İmza

Bu bölüme kişinin kendi el yazısı ile okudum
anladım yazmalı ve imzalanmalıdır.

İmza

Bu bölüme kişinin kendi el yazısı ile okudum
anladım yazmalı ve imzalanmalıdır.

ÖZET

Yüksek Lisans

SAC MALZEME SATIN ALMA SÜREÇLERİNDE VERİ MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ UYGULAMALARI

Seray MİRASÇI

Bursa Uludağ Üniversitesi
Fen Bilimleri Enstitüsü
Endüstri Mühendisliği Anabilim Dalı

Danışman: Doç. Dr Aslı AKSOY

Firmaların varlıklarını sürdürebilmeleri için, belli karlılık hedeflerini tutturmaları gerekmektedir. Firmalarda karlılık hedeflerine doğrudan etki eden faaliyetlerden biri de satın alma faaliyetleridir. Değişen dünya koşullarında satın alma süreçlerinin kritik malzeme grupları için çevik ve stratejik olması gerekmektedir. Bu çalışmada, çelik malzeme ürün grubunda fiyat tahminleme çalışması için veri madenciliği ve makine öğrenmesi yöntemleri ortaya konmuştur. Veri setinde bulunan gürültülü veriler tespit edilerek veri madenciliği teknikleri ile temizlenmiştir. Temizlenen veri seti makine öğrenmesi tekniklerinden hiyerarşik kümeleme ve k-ortalama yöntemleri kullanılarak ideal küme sayıları tespit edilmiştir. Bu analizde bulunan ideal küme sayısının doğrulanması farklı performans ölçütlerine göre doğrulanmıştır. Belirlenen kümede yer alan referanslar için hem yapay sinir ağları ile hem adaptif ağ yapısına dayalı bulanık çıkarım sistemleri (ANFIS) ile fiyat tahminlemesi yaparak hangi yöntemin diğerine göre daha üstün olduğu ortaya konmuştur.

Bu çalışmada önerilen analizler ile satın alma süreçlerinde, çalışan kaynaklı hataların satın alma stratejileri geliştirme süreçlerindeki etkileri azaltılmış, satın alma çalışanlarının uzun zaman harcayarak yaptığı analizler, endüstri mühendisliğinin teknikleri içerisinde yer alan veri analizi ve makine öğrenmesi yöntemleri ile gerçekleştirilmiştir.

Anahtar Kelimeler: Stratejik Satın alma, Fiyat Tahminleme, Yapay Sinir Ağları, ANFIS, Veri Madenciliği, Makine Öğrenmesi

ABSTRACT

MSc Thesis

MACHINE LEARNING AND DATA MINING APPLICATIONS IN STEEL MATERIAL PURCHASING PROCESSES

Seray MİRASÇI

Bursa Uludağ University
Graduate School of Natural and Applied Sciences
Department of Industrial Engineering

Supervisor: Associate Professor Dr Aslı AKSOY

For the sustainability of companies, it is necessary to meet certain profitability targets. Purchasing function is one of the crucial functions that directly affects the profitability targets of companies. It has become a necessity to act agile and strategically for critical material groups in changing world conditions of purchasing functions. In this study, data mining and machine learning methods has presented to forecast price of steel material product groups. The noisy data in the data set was revealed and cleaned with data mining techniques. The data set was analyzed by clustering analysis such as hierarchical clustering and k-means methods. Optimal number of clusters was determined and validated by different methods. For the references in the selected cluster, price forecasting models was presented by using artificial neural networks (ANN) and adaptive network based fuzzy inference systems (ANFIS) The proposed forecasting model aims to reduce the effects of purchasing employee related errors in strategy development process for purchasing decisions and the analyses made by purchasing experts by spending a long time were carried out with industrial engineering methods such as data mining and machine learning algorithms.

Key words: Strategic Purchasing, Price Estimation, Artificial Neural Network, ANFIS, Data Mining, Machine Learning.

TEŐEKKÜR

Bu tez alıřmamın oluřma s¼recinde, yardımlarını hibir zaman esirgemedен, y¼nlendirici, anlayıřlı ve potansiyelimi ortaya ıkararak beni teřvik eden ok deęerli danıřmanım Do. Dr. Aslı Aksoy'a ve tez konumun fikir babası olan Prof. Dr. H¼seyin Cenk ¼zmutlu'ya sonsuz teřekk¼rlerimi sunarım.

Bana lisans sıralarında azimle ve inatla inandıęım yolda y¼r¼meyi ¼ğreten deęerli hocam Do. Dr. Olcay Polat'a, gerek lisans gerek y¼ksek lisans eęitimimde her zaman desteęini arkamda hissettiren Prof. Dr. Ařkıner G¼ng¼r'e ve bana malzeme bilimini ¼ğreten ve bu tezin ortaya ıkmasında emeęi b¼y¼k olan Prof. Dr. Cemal Meran'a teřekk¼r¼ bor bilirim.

Bug¼nlere gelmemde payı ok b¼y¼k olan, beni eęitimimin her ařamasında sonsuz bir ¼zveri ile destekleyen aileme sonsuz teřekk¼rlerimi sunarım.

Son olarak, bu ařamaya gelene kadar ¼zerimde emeęi olan t¼m hocalarıma ve arkadařlarıma teřekk¼r ederim.

Seray MİRASI
Ocak, 2023

İÇİNDEKİLER

	Sayfa
ÖZET.....	i
ABSTRACT.....	ii
TEŞEKKÜR.....	iii
SİMGELER VE KISALTMALAR DİZİNİ.....	v
ŞEKİLLER DİZİNİ.....	vi
ÇİZELGELER DİZİNİ.....	vii
1. GİRİŞ.....	1
2. KURAMSAL TEMELLER ve KAYNAK ARAŞTIRMASI.....	4
2.1. Veri Madenciliği.....	4
2.2. Makine Öğrenmesi.....	7
2.3 YSA ve ANFIS Sistemleri.....	11
3. MATERYAL ve YÖNTEM.....	15
3.1. Veri Madenciliği Yöntemleri.....	15
3.2. Makine Öğrenmesi Yöntemleri.....	18
3.2.1. Apriori Algoritması.....	18
3.2.2. K-ortalamar Yöntemi.....	19
3.3. Satın Alma Fiyatlarına Etki Eden Kriter Belirleme Yöntemleri.....	23
3.4. Yapay Sinir Ağları.....	24
3.4.1. Yapay Sinir Ağı Çeşitleri.....	28
3.4.2. Matlab Programında Yapay Sinir Ağları.....	28
3.5. Bütünleşik Bulanık Sistemler.....	30
3.6. Adaptif Ağ Yapısına Dayalı Bulanık Çıkarım Sistemleri (ANFIS).....	31
4. BULGULAR VE TARTIŞMA.....	35
4.1. Veri Madenciliği Yöntemleri ile Yapılan Çalışmalar.....	36
4.2. Makine Öğrenmesi Yöntemleri Sınıflandırılma.....	40
4.2.1. Kümeleme Çalışmaları Sonrasında Geliştirilen Stratejiler.....	44
4.3. Fiyat Tahmin Modelinin Oluşturulması.....	45
4.3.1. Yapay Sinir Ağları ile Fiyat Tahmin Modeli.....	46
4.3.2. ANFIS ile Fiyatların Tahminlenmesi.....	52
4.4. YSA ve ANFIS ile Satın Alma Stratejilerinin Oluşturulması.....	56
5. SONUÇ.....	57
KAYNAKLAR.....	60
EKLER.....	64
ÖZGEÇMİŞ.....	73

SİMGELER ve KISALTMALAR DİZİNİ

Kısaltmalar Açıklama

YSA	Yapay Sinir Ağları
ANFIS	Adaptif Ağ Yapısına Dayalı Bulanık Çıkarım Sistemleri
VBO1	Veri Bölünme Oranı 1 (%70'i eğitim, %15 validasyon ve %15 test)
VBO2	Veri Bölünme Oranı 2 (%80'i eğitim, %10 validasyon ve %10 test)
FIS	Bulanık Çıkarım Sistemi
BM	Bulanık Mantık
KBÖS	Karşılıklı Bilgi Öznitelik Seçim Algoritması
HR	Sıcak Haddelenmiş Çelik Malzeme
HDG	Sıcak Galvanizlenmiş Çelik Malzeme
CR	Soğuk Haddelenmiş Çelik Malzeme
MAE	Mutlak Ortalama Hata
MSE	Hata Kareler Ortalaması
OMS	Ortalama Mutlak Sapmaların Yüzdesi

ŞEKİLLER DİZİNİ

	Sayfa
Şekil 1. Sık tercih edilen makine öğrenmesi algoritmaları	7
Şekil 2. YSA yapısı	25
Şekil 3. Yapay sinir hücresi modeli (Aksoy, 2012)	25
Şekil 4. Bulanık çıkarım sistemi (Yücel, 2010)	30
Şekil 5. İki girdili bir çıktılı ANFIS yapısı (Karaboga vd., 2019)	32
Şekil 6. Gürültülü veriler barındıran veri setinin fiyat-tüketim trafiği	36
Şekil 7. Gürültülü veriler barındıran veri setinin ciro-tüketim grafiği	36
Şekil 8. Gürültülü verilerin temizlenmesinden sonra fiyat-tüketim grafiği	37
Şekil 9. Gürültülü veriler temizlenmesinden sonra ciro-tüketim grafiği	37
Şekil 10. Gürültülü veriler temizlenmeden önceki duruma ait grafik	38
Şekil 11. Gürültülü veriler temizlendikten sonraki duruma ait grafik	38
Şekil 12. Kümeleme uygulamaları	39
Şekil 13. Farklı küme sayıları için k-ortalamalar yöntemi ile kümeler	41
Şekil 14. Hiyerarşik kümeleme dendogramı	42
Şekil 15. MRMR algoritması sonuçları	45
Şekil 16. Simulink model örneği	48
Şekil 17. ANFIS hata değeri	54

ÇİZELGELER DİZİNİ

	Sayfa
Çizelge 1. Biyolojik sinir sistemi elemanlarının, yapay sinir ağı karşılıkları	24
Çizelge 2. Matlab’te bulunan eğitim algoritmaları (Beale vd., 2018)	28
Çizelge 3. Ağın eğitimini durdurma kriterleri (Beale vd. 2018).....	29
Çizelge 4. Farklı küme sayıları için validasyon değerleri.....	40
Çizelge 5. Optimal küme sayıları.....	40
Çizelge 6. K-ortalamalar yöntemi ile belirlenen kümelerin proje bazlı ayrılması.	41
Çizelge 7. Hiyerarşik kümeleme ile belirlenen kümelerin proje bazlı ayrılması..	42
Çizelge 8. MRMR algoritma değerleri.....	45
Çizelge 9. VBO1 ile en iyi R değerine sahip 5 Deneme	47
Çizelge 10. VBO2 ile en iyi R değerine sahip 5 Deneme	47
Çizelge 11. VBO1 ve VBO2 için üç test sonucunda değerler	49
Çizelge 12. Tek gizli katman için en iyi R değerine sahip ağlar.....	50
Çizelge 13. Çift gizli katman için en iyi R değerine sahip ağlar.....	50
Çizelge 14. 28 Adet verinin Tahmin Sonuçları.....	51
Çizelge 15. Üyelik fonksiyonları	53
Çizelge 16. 28 adet verinin ANFIS ile test sonuçları.....	54

1. GİRİŞ

İşletmelerin hayatta kalabilmesi için kritik öneme sahip faaliyetlerden bir tanesi de satın alma faaliyetleridir. Satın alma faaliyetleri, tedarikçi ve müşteri arasında bir köprü görevi görmekle birlikte, rekabet, maliyet, kalite, zaman planlaması gibi konuları da kapsamaktadır.

Değişen dünya koşullarında, satın alma fonksiyonlarının çevik ve doğru kararlarla yönetilmesi gerekmektedir. Aksi halde istenmeyen maliyetlerin doğmasına, üretim fonksiyonlarında verimlilik kayıplarına veya üretim duruşlarına yol açabilir.

Satın alma yapılan kalemler arasında, kritik öneme sahip birçok malzeme vardır. Önemli grupta yer alan malzemelerden biri de sac malzemelerdir. Sac malzemeler, otomotiv, beyaz eşya, havacılık, madencilik gibi birçok sektörde farklı amaçlarla kullanılabilir.

Sac malzemeler, büyük bobinler halinde üretilip sonrasında istenilen ebatlarda kesilerek ilgili süreçlerden geçtikten sonra talep edilen formlara dönüştürülmektedir. Ancak karmaşık ve çok operasyonlu bir süreç olması sebebiyle doğru yönetilmesi gerekmektedir.

Çalışmanın amacı, sac malzeme satın alma süreçlerinde karlılık sağlayacak stratejilerin hızlı ve kolay ortaya konabilmesi için fiyat tahminlemesi yapan modeli oluşturmaktır.

Bu kapsamda, otomotiv sektöründe yer alan sac malzeme grubu için, doğru satın alma stratejilerini oluşturabilmek amacıyla bir uygulama yapılmıştır. Bu uygulama üç temel basamaktan oluşmaktadır:

1. Adım: Gürültülü veriler içeren satın alma veri setinin, veri madenciliği teknikleri ile temizlenmesi ve sonucunda stratejik öneme sahip kümenin belirlenmesi
2. Adım: Makine öğrenmesi yöntemleri ile, seçilen stratejik projelere ait fiyatların tahminlenmesi ve satın alma stratejilerinin geliştirilmesi

3. Adım: Tahminlenen fiyatların ikinci bir yöntem ile kıyaslanarak en doğru yöntemin ortaya konması ve satın alma stratejilerinin belirlenmesi.

Literatürde yapılan çalışmalar incelendiğinde seçilen konularda birtakım eksiklikler saptanmıştır. Satın alma için, Türkçe literatürde müşterilerin satın alma davranışları üzerinden pek çok çalışma mevcutken, satın almayı kendi içinde ve dinamikleriyle inceleyen çalışmaların yeterli sayıda olmadığı tespit edilmiştir. Köylüoğlu vd. (2018) çalışmasında akademisyenlerin araç satın alma davranışları üzerinde çalışmıştır. Bu çalışmada SPSS programını kullanarak, regresyon ve Anova gibi istatistiksel analiz ve testler kullanılmıştır. Kırgız (2021) ise yaptığı çalışmada, müşterilerin lüks otomobil satın alma kararlarının hangi faktörlerin etkilediği üzerinde korelasyon yöntemiyle çalışmıştır. Ertaylan vd. (2021) çalışmasında elektrik alanında gelecek dönem piyasa takas fiyatlarını yapay sinir ağları (YSA) ile tahminlemiştir.

Bu çalışma ile hedeflenen, günlük hayatta da sık karşılaşılan satın alma problemlerine, veri madenciliği ve makine öğrenmesi yöntemleri ile çözüm oluşturmak ve karlı stratejiler oluşturmak için en doğru fiyat tahminlemelerini gerçekleştirmektir.

Bu çalışmada satın alma departmanında çalışanlar için mesleki tecrübelerine göre iki ayrı fiyat tahminleme yöntemi sunulmuştur. Tecrübeli satın alma çalışanları adaptif ağ yapısına dayalı bulanık çıkarım sistemleri (ANFIS) yöntemi ile tecrübelerine dayanarak kuralları belirledikten sonra fiyat tahminlemelerini gerçekleştirmektedir. Ancak tecrübesiz bir satın alma çalışanı, ANFIS'te kuralları belirlerken hata yapma riskine sahiptir ve kurallar yanlış tanımlanırsa fiyat tahminlerinde hatalar gözlemlenebilir. YSA ile fiyat tahminlerken herhangi bir satın alma tecrübesi gerekmemektedir. Bu sebepten tecrübesiz satın alma çalışanları için yapay sinir ağları ile fiyat tahminlemesi yapılmıştır.

Bu tezin ortaya çıkmasında günlük hayatta satın alma departmanlarında karşılaşılan durumlar yol gösterici olmuştur. Örneğin tedarikçilerden gelen tekliflerin anlamlı bir fiyat düzeyinde olup olmadığını anlamak için farklı firmalardan alternatif teklifler alınmalı, piyasa araştırmaları yapılmalı veya tedarikçilerden fiyatlandırma detaylarını içeren fiyat

kırılımları talep edilmelidir. Tüm bu süreçler zaman almaktadır. Tez çalışmasında önerilen fiyat tahminleme yöntemleri ile tedarikçiden alınan bir teklifin anlamlı bir seviyede olup olmadığı saniyeler içinde yorumlanabilmektedir. Benzer şekilde piyasa araştırması veya farklı firmalardan teklif almak yerine YSA ve ANFIS yöntemlerinin hangi fiyat bandında hesaplama yaptığı dikkate alınarak yorum yapılabilir.

2. KURAMSAL TEMELLER ve KAYNAK ARAŞTIRMASI

Tez çalışması iki ayrı temel konuyu kapsadığı için, kaynak araştırması da iki ayrı konu özelinde hazırlanmıştır. Bunlar, veri madenciliği ve makine öğrenmesi yöntemleridir.

2.1. Veri Madenciliği

Veri madenciliği tanınmayan, saklı veya faydalı verilerin veri tabanlarından belli kurallar çerçevesinde otomatik bir şekilde elde edilerek bilginin işlevsel hale getirildiği analiz sürecidir (Erbudak, 2022).

Kameshwaran vd. (2014) veri madenciliğini, bir veri setinden, ihtiyaç duyulan bilginin doğru şekilde çıkarılarak, anlaşılır şekle dönüştürülme süreci olarak tanımlamaktadır. Polat (2022), yaptığı çalışmada veri madenciliğinin kuvvetli ve zayıf yönlerini aşağıdaki gibi sunmaktadır.

Kuvvetli Yönler:

- Çok boyutlu ele alınabilecek sorunların çözümü için uygun bir araçtır.
- Karmaşık durumların çözümlenmesinde kullanılmak için uygundur
- Hem sayısal hem de kategorik değişkenlerin çözümünde kullanılabilir.

Zayıf Yönler:

- 0 ve 1 arasında giriş verisi olması gerekir. Yani veri setinin nümerik bir matris haline dönüştürülmesi gerekmektedir.
- Ürettikleri sonuçların açıklamasını sunamamaktadır.
- Sonuçlar farklı teknikler ile tekrarlanarak daha iyi sonuçlar ortaya konabilir.

Savaş vd. (2012) yaptıkları çalışmada, veri madenciliği çalışmalarında karşılaşılan problemleri aşağıdaki gibi belirtmektedir:

- Artık veri: Veri setindeki gereksiz ve ihtiyaç duyulmayan verilerdir.
- Belirsizlik: Hataların şiddetini belirtir.

Confidential C

- Boş değerler: Herhangi bir değere eşit olmayan ve niteliği olmayan verilerdir.
- Dinamik veri: Sürekli değişen verilerdir. Genellikle kurumsal hayattaki veriler bu gruba girmektedir.
- Eksik veri: Veri setinin içeriği gereğince veya hata ile doğan verilerdir. İstatistiksel analizlerde problem yaratma ihtimali yüksektir.
- Farklı tipteki veriler: Sadece sembolik veya kategorik veriler değil, tam sayılı, kesirli, coğrafi konum bilgileri gibi farklı tipte verilerin mevcut olmasıdır.
- Gürültülü değerler: Verinin toplanması veya aktarılması aşamasında doğan, istenmeyen ve analizleri yanlış etkileme riski bulunduran verilere gürültülü değerler denilmektedir.

Türkiye’de veri madenciliği uygulamalarına bakıldığında 2000’li yıllardan sonra popülerleşmeye başlayıp Savaş vd. (2012) yapmış olduğu çalışmada detaylı olarak yer verilmiştir. Bu çalışmaya göre, mühendislik, bankacılık, eğitim, ticaret, sağlık gibi farklı sektörlerde veri madenciliği uygulamalarından faydalandığını ortaya koymuştur.

Veri madenciliği çalışmalarında, önemli noktalardan biri veri setinin dağılımı hakkında bilgi sahibi olmaktır. Bir verinin normal dağılıma sahip olup olmadığını anlamak için geliştirilen birçok test vardır. Jeong vd. (2016) yaptıkları çalışmada, verilerin normallik testini yaparken faydalanılabilecek birçok testi ortaya koymuştur. Bunlar, Kolmogrov-Smirnov testi, Lilliefors testi, Cramer-von Mises testi, Shapiro-Francia testi, Jarque-Bera testi, D’Agostinos testi, Doornik-Hansen testi, Enerji testi , Shapiro Wilk ve bu çalışmada kullanılan Anderson Darling normallik testidir.

Veri setinin normallik testinde çıkan sonuca göre doğru makine öğrenmesi yönteminin seçilmesi önem arz etmektedir. Örneğin, veri seti normal dağılıma sahip değilse ve dönüşüm yapmadan regresyon yapılırsa sonuç olarak hatalı tahminlerin yapılması olasıdır.

Babaoğlu (2015) yaptığı tez çalışmasında, veri madenciliği uygulama aşamalarını yedi ana adımda incelemiştir. Bu adımlar;

Confidential C

1. Problemin ve çalışma sonucunda ulařılmak istenen bilginin tanımlanması ve veri setinin anlaşılması
2. Veri setinin temizlenerek, istenmeyen, amaca hizmet etmeyen, tekrarlı veya uygunsuz verilerin elimine edilmesi
3. Verilerin düzenli ve kontrollü bir şekilde bir araya getirilmesi
4. İhtiyaç duyulan verilerin belirlenmesi ve seçilmesi
5. Uygun veri madencilik tekniđi ile verilerin analizi
6. Verilerin deđerlendirilmesi
7. İhtiyaç duyulan bilgilerin, verilerden sentezlenmesidir.

Ricci vd. (2011) veri madenciliđini üç ayrı adımda incelemiřtir. Bu adımlar verinin hazırlanması, analizi ve yorumlanmasıdır. Verinin hazırlanması adımı, veriler arasındaki uzaklıkların ölçülmesi, örneklemlerin seçilmesi ve uzaklıkların azaltılmasını adımlarını kapsamaktadır. Analiz adımı, tahmin ve tanımlama olarak ikiye ayrılmaktadır. Tahmin için sınıflandırma yöntemleri kullanılmaktadır. Bu yöntemler k-en yakın komřu, karar ağaçları, Bayes yöntemi ve destek vektör makineleri olarak sayılabilir. Tanımlama için ise birliktelik analizleri kullanılmaktadır. Yorumlama adımı ise, veri madenciliđinden elde edilen çıktıların bilgiye dönüřtüđü yorum aşamasını içermektedir.

Asilkan (2008) yaptıđı çalışmasında otomotivde ikinci el araç satın almaları için veri madenciliđi yöntemlerini kullanarak fiyat tahminlemesi yapmıştır. Regresyon analizini veri madenciliđi alanında kullanarak yapay sinir ağları ile kıyaslayan bu çalışma sonucunda yapay sinir ağlarının daha yüksek performans sağladıđı ortaya koymuştur.

Hepsen (2012) yaptıđı çalışmada gayrimenkul piyasasında fiyat endekslerinin tahminlenmesi ve gayrimenkul satın almalarının sağlıklı yönetilebilmesi için regresyon yöntemi ile fiyat tahminlemesi yapmıştır.

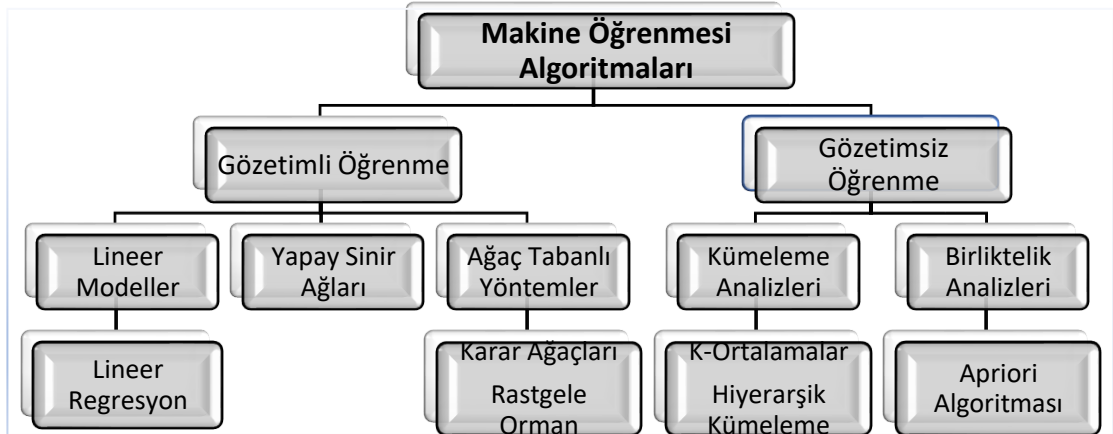
Kaur vd. (2014) çalışmasında, tarımsal ürünlere ait satın alma fiyatlarının tahminlenmesi için veri madenciliđi tekniklerinden k-en yakın komřu, k-ortalamlar, regresyon ile fiyat

tahminlemesi yapmıştır. Bu çalışmada sayısal veriler ve yöntemlerin formülleri detaylı verilmemiş olup kullanılması tavsiye edilen yöntemler belirtilmiştir.

2.2. Makine Öğrenmesi

Makine öğrenmesi kısaca, bir bilgisayarın deneyimlere dayalı öğrenmesi olarak tanımlanabilir (Al-Omary vd., 2005). Bu deneyimler Mitchell (1997) tarafından aşağıdaki birkaç örnekle açıklanmıştır;

1989 yılında geliştirilen ALVINN sistemi ile makine öğrenmesi metotlarıyla deneyimlerin veriler ile aktarıldığı bilgisayar kontrollü araçlar farklı yollarda sürüş testine çıkarılmıştır. Burada bilgisayar yol tipine göre araç hızını 70 km/h ve 90 km/h arasında hızı ayarladığı gözlemlenmiştir. Bu sayede aracın zorlu yollara yüksek hızlarla girerek zarar görmesi engellenmiştir. 1995 yılında TD-Gammon şirketi makine öğrenmesi tabanlı veriler aracılığıyla deneyim ve kuralların aktarıldığı bir satranç oyunu yaratmıştır. Bu oyunda bir milyondan fazla kez eğitilen algoritma dünya satranç şampiyonunu yenerek büyük ses getirmiştir. Benzer şekilde NASA'nın Palomar Observatory Sky Survey isimli araştırmasında makine öğrenmesi algoritmaları gök cisimlerini sınıflandırmak için kullanılmıştır. Şekil 1.de görüleceği üzere makine öğrenmesi kendi içinde, gözetimli ve gözetimsiz öğrenme olmak üzere ikiye ayrılmaktadır (Özgür, 2002):



Şekil 1. Makine öğrenmesi algoritmaları

Gözetimli öğrenme, bağımlı ve bağımsız değişken arasındaki ilişkiyi ortaya koymayı amaçlamaktadır. Bu yöntem için, bağımlı değişkenlerin değerlerinin biliniyor olması önem arz etmektedir.

Gözetimli öğrenme kendi içinde lineer modeller ve ağaç tabanlı yöntemler olarak ikiye ayrılmaktadır. Lineer modellerin alt grubunda lineer regresyon, ağaç tabanlı yöntemlerin altında ise karar ağaçları, rastgele orman yöntemi gibi örnekler yer almaktadır (Özgür, 2002).

Gözetimsiz öğrenmede tüm değişkenler ayrı ayrı ele alınmakta ve bağımlılık ilişkisi aranmamaktadır ancak gözetimli öğrenmeden farklı olarak, gözetimsiz öğrenmede ulaşılması gereken hedefler vardır. Bunlar, veri sayısının azaltılması, verilerin kümelenmesi gibi amaçlar olabilir. Değişkenlik içeren durumlarda, eğer yeterli sayıda veri varsa ve değişkenler iyi biliniyorsa gözetimsiz öğrenme tercih edilebilir (Özgür, 2002).

Gözetimsiz öğrenme, kümeleme analizleri ve birliktelik analizleri içermektedir. Kümeleme analizi altında, k-ortalamlar ve hiyerarşik kümeleme sık kullanılan makine öğrenmesi algoritmalarına örnek verilebilirken, birliktelik analizleri içerisinde apriori algoritmaları örnek verilebilir.

Gözetimli öğrenme teknikleri incelendiğinde Liu vd. (2000) çalışmasında, gözetimli öğrenme tekniklerinden biri olan karar ağaçlarını, normal dağılıma sahip olan ve olmayan iki veri seti üzerinde uygulamış ve normal dağılıma sahip veri setindeki performans ve hassasiyetin daha iyi düzeyde olduğunu ortaya koymuştur.

Arslankaya vd. (2021) çalışmasında hisse senedi fiyatlarını tahminlemek için gözetimli öğrenme tekniklerinden regresyon, rastgele orman ve yapay sinir ağları yöntemlerini kullanmıştır.

Ganesh vd. (2019) çalışmasında kullanılmış araçların satın alma fiyatlarının tahminlenmesi için regresyon ve karar ağaçları yöntemlerini kullanmıştır. Lasso regresyon ve regresyon ağaçları yöntemlerinden faydalanan bu çalışma optimal fiyatların belirlenmesi hedeflenmiştir.

Gözetimsiz öğrenme tekniklerine ait literatür çalışmaları ise aşağıdaki gibi incelenmiştir;

Erpolat (2012) çalışmasında, bir otomobil yetkili servisinde müşterilerin alışveriş verileri arasındaki bağlantıları apriori algoritması ile çalışmış ve hangi ürünlerin birlikte alındığını ortaya koymuştur.

Saputra vd. (2019) çalışmasında kümeleme yöntemleri başlığı altında, k-ortalamar yöntemini kullanarak ideal küme sayısının belirlemiş ve değerlendirmiştir. Bu çalışmada ideal küme sayısını belirlemek için dirsek yöntemi, Silhouette yöntemi ve Gap istatistik yönteminden faydalanmıştır. Bu çalışma sonucunda bu üç yöntemin aynı küme sayısını bulmasının bir zorunluluk olmadığı ve her yöntemin ayrı bir hesap mantığı ile küme sayısını belirlediğini belirtmiştir.

Patel vd. (2022) çalışmasında, bir veri setinin optimal küme sayısının bulunması için k-ortalamar yöntemi ve hiyerarşik kümeleme yöntemlerini karşılaştırmıştır. K-ortalamar yöntemi için, dirsek, Silhouette ve Gap istatistik yöntemlerini kullanarak küme sayılarını ortaya koymuştur. Küme sayılarının yöntemlere göre değişebileceğinin önemini belirtmiştir. Küme sayısının geçerliliğini test edebilmek için ayrıca bağlantısallık (Connectivity), Dunn ve Silhouette yöntemleri ile test yapmıştır. Bu çalışma sonucunda, ilgili veri seti için en uygun kümeleme yöntemi, kümelemenin geçerliliğini test eden yöntem ve küme sayısını ortaya koymuştur.

Hiyerarşik kümeleme yöntemi, literatürde dendogram ismiyle de yer almaktadır. Bu yöntemde, veriler arasındaki ilişkiler seviyeler halinde oluşturulur ve analiz edilir. İlk adımda ikiye bölme yoluyla başlayan hiyerarşik kümeleme, istenen bilgilere ulaşana

kadar kümelere ayrılarak devam eder. Bu kümeleme yöntemi büyük veri setleri için kullanıma uygundur (Zhao vd., 2005).

Satın alma alanında hiyerarşik kümelemenin kullanımı incelendiğinde Rokaha vd., (2018) yaptığı çalışmada süpermarketlerde müşterilerin satın alma davranışları üzerinde çalışmıştır. Bu analizde müşterilerin geçmişte satın aldıkları ürünlere göre, gelecekte müşterilere sunulacak teklifler için ilişkilendirme kuralları ve bu ürünlerin markette doğru şekilde yerleştirilmesi için hiyerarşik kümelemeden faydalanılmıştır.

Serin (2020) çalışmasında, piyasa değeri ilk 25 kripto para arasında yer alan paraların, 1 saatlik, 24 saatlik ve 7 günlük değişim verilerine göre hiyerarşik kümelemesi yapmıştır. Yapılan kümelemede 6 seviye oluşmuştur. Çalışma sonucuna göre fiyat değişimi parametresinde birbirine en çok benzeyen kripto paralar kümelenemiştir.

Bejaei vd. (2020) çalışmasında tüketicilerin taze elma satın alma davranışları ve tercihleri tüketici anketi kullanılarak araştırılmıştır. Ankete toplam 169 müşteri katılmıştır. Hiyerarşik kümeleme yöntemi kullanılarak müşteri profilleri belirlenmiş ve mevcut elma çeşitlerindeki değişikliğin tüketicilerin satın alma davranışları üzerindeki etkisi araştırılmıştır. Yapılan ankete göre müşteriler elma satın alırken görsel görünüm, önceki deneyimler, tat/koku ve doku kriterlerine önem vermektedir. Hiyerarşik Kümeleme ile müşteriler beş ayrı sınıfa ayrılmıştır ve bu müşteri tipleri için davranışlar ortaya konmuştur.

2.3 YSA ve ANFIS Sistemleri

Yapay zekanın popüler alanlarından biri olan yapay sinir ağları (YSA) gün geçtikçe daha önemli bir seviyeye gelmektedir. İnsan beynindeki öğrenme yeteneğinin, matematiksel modellenmesi ile ortaya çıkan YSA temelde örnekler yardımıyla öğrenen sistemlerdir. Bu sinir ağları, yapay sinir hücrelerinden oluşmaktadır. Her hücrenin bir ağırlığı vardır ve sistemde tutulan bilgi bu ağırlık değerlerinde saklanmaktadır (Öztemel, 2012).

YSA sistemleri üç katmandan oluşmaktadır. Bunlar girdi katmanı, ara katmanı ve çıktı katmanıdır. Sisteme yüklenen veriler girdi katmanına gelir, ara katmanda işlenir ve çıktı katmanında bilgiye dönüşür. Ağın doğru bilgileri çıktı katmanından sunabilmesi için, ağırlıkların doğru olması önem arz etmektedir. İlk başta rastgele atanan ağırlık değerleri, eğitim süreci ile doğru ağırlıklara ulaşmaktadır. YSA'nın hesaplama metodolojisi bilinen yöntemler ile farklılık göstermektedir. Bu sistemler, eksik, bilinmeyen ve belirsiz veriler ile çalışabilirler (Öztemel, 2012).

YSA'nın eğitilmesi için birçok eğitim algoritması söz konusudur. Bunlardan bazıları; Levenberg Marquardt, Scale Conjugate Gradient algoritmalarıdır. Literatürde en yaygın görülen algoritma Levenberg Marquardt algoritmasıdır (Ranganathan, 2004).

Levenberg Marquardt algoritmasında ağırlık değerlerini her iterasyonda güncellemektedir. Bilgisayarda bellek kullanımı yüksek bir algoritma olmasına rağmen genellikle en hızlı eğitim algoritmasıdır (The Mathworks.).

Ataseven (2013) çalışmasında YSA'da öğrenme yöntemi olarak gözetimli öğrenme kullanıldığını açıklamaktadır. Gözetimli öğrenmede YSA, kullanılmadan önce eğitilmelidir. Bu sisteme gözetimli öğrenme isminin verilme sebebi bir eğitim sürecine tabi olmasıdır. Gözetimli öğrenme kuralları; algılayıcı öğrenme, delta öğrenme, genişletilmiş delta öğrenme ve geri yayımlı öğrenmedir. Levenberg Marquardt algoritması geri yayımlı öğrenme yöntemini kullanan bir algoritmadır. Geri yayılım algoritmasında her iterasyon, ileri yayılım ve geri yayılım olmak üzere iki aşamadan oluşmaktadır. İleri yayılım aşamasında yapay sinir ağının o andaki durumunda yapay sinir ağına uygulanan giriş değerine karşı yapay sinir ağının çıkışlarında oluşan değerler bulunur. Geri yayılım aşamasında, oluşan hatalardan yola çıkılarak, nöronlardaki ağırlıkların yeniden düzenlenmesi yapılmaktadır. Ağın eğitilmesinin sağlıklı bir şekilde gerçekleştirilebilmesi için başlangıçta rastgele olarak atanan bağlantı ağırlıkları çok önemlidir. Rastgele olarak atanan ağırlıkları eğitime hangi noktadan başlanacağını belirlemektedir.

Yapay sinir ađlarında önemli noktalardan biri eğitimin ne kadar süreceđine karar verilmesidir. Eğitim sürecinin tamamlanması için iki ayrı yöntem izlenebilir. Bunlardan birincisi belirlenen hata toleransından daha düşük hata deđerine ulařana kadar eğitime devam etmektir. Diđer yöntem ise sabit bir eğitim iterasyonu sayısı seçmektir. Belirlenen iterasyona ulařtıđında eğitim durdurularak eğitim sonucunda elde edilen hata kabul edilmektedir. Eğitim esnasında hata deđerinin hep azalması beklenmemelidir, eđer hata deđeri artıyorsa bu çözümden uzaklařıldığını göstermektedir. Son zamanlarda yapay sinir ađlarında gözetimsiz öğrenme yöntemi içinde arařtırmalar yapılmaktadır. Bilgisayarların insan desteđi olmadan öğrenebilmesi temelinde çalıřan bu öğrenme yöntemi için arařtırmalar hala devam etmektedir ve günümüzde çok sınırlı kullanım alanına sahip olduđundan YSA hala gözetimli öğrenme yöntemleri arasında anılmaktadır (Ataseven, 2013).

Satın alma süreçlerinde YSA ile fiyat tahminlemesine konusunda Ertaylan vd., (2021) çalıřmasında elektrik fiyatları tahminlemesinde kullanılan piyasa takas fiyatlarını YSA ile tahminlemiřlerdir.

Erdođan vd. (2012) çalıřmasında, İstanbul menkul kıymetler borsasının ilk 100 endeksi içinde bulunan beyaz eřya firmalarının hisse senetlerinin satın alma fiyatlarını günlük olarak yapay sinir ađları ile tahminlemiřtir. Hisse senetlerinin fiyatlandırmalarını etkileyen faktörler literatürden alınmıř olup yapay sinir ađları ile tahminlemeler gerçekleştirilmiřtir.

Sakız vd. (2018) çalıřmasında kripto paraların satın alma fiyatlarını yapay sinir ađları ile tahminlemiřtir. Veri seti olarak geçmiř üç senenin satın alma fiyatlarını kullanarak gelecek satın alma fiyatlarını tahminlemiřtir.

Adaptif ađ yapısına dayalı bulanık çıkarım sistemi (ANFIS) ise Sugeno tipi bulanık sistemlere yapay sinir ađlarının sinirsel öğrenme kapasitelerinin eklenmesidir (Aksoy, (2012). ANFIS algoritmasının temelinde melez öğrenme sistemleri yer alır. Melez

öğrenme sistemleri, bulanık mantık ile yapay sinir ağlarının birleşimine dayanmaktadır. Bulanık mantık kısmı uzman görüşü alınmış kurallara dayanmaktadır (Rençber vd. 2016). ANFIS, beş katmandan oluşmaktadır. İlk katmanı girdi katmanıdır, ikinci katmanı bulanıklaştırma, üçüncü katmanı kuralların bulunduğu katman, dördüncü katmanı normalizasyon, beşinci katmanı ise durulaştırmanın yapıldığı katmandır. Durulaştırma katmanından sonra istenen bilgi kullanıcıya sunulmaktadır (Saltan vd. 2018).

Birinci katmanda girdiler sisteme verilirler ve üyelik fonksiyonlarına uygun olarak ayrılarak ikinci katmana geçmektedir. İkinci katmanda veriler bulanıklaştırılmaktadır. Üçüncü katmanda uzman görüşü ile oluşturulan bulanık mantık kuralları yer alır ve gelen sinyaller bu kurallara göre değerlendirilmektedir. Dördüncü katmanda normalizasyon yapılır, yani tüm sinyallerin normalleşme seviyesi belirlenmektedir. Beşinci katmanda veri durulaştırılmakta (berraklaştırma) ve değişkenin nihai ağırlıklandırılmış parametre değeri oluşturulmaktadır. Sonrasında çıkış noktasından nihai değer elde edilmektedir (Sel vd., 2022)

Yüksek vd. (2017) çalışmasında ANFIS modeli oluştururken adımları şu şekilde tanımlamaktadır: Bulanık sistem oluşturulur, iterasyon sayısı, hata toleransı gibi öğrenme parametreleri belirlenir. ANFIS ile eğitim başlatılır ve istenen değere ulaşıncaya kadar eğitim sona erdirilir. Bağımsız veriler ile doğrulama yapılır ve sonuçlar analiz edilir.

ANFIS yönteminin satın alma sürecine literatürdeki uygulaması olarak Giri vd. (2019) çalışmasında yapay zeka ve yapay sinir ağlarının moda ve konfeksiyonda kullanımlarını içeren farklı çalışmaları bir araya getirmiştir. Örneğin bir araya gelen çalışmalarda, tekstilde global satın alma karar destek sistemlerinin oluşturulması için adaptif ağ yapısına dayalı bulanık çıkarım sistemi ile bir çalışma yapılmıştır. Bu sayede, satın alma sorumlusunun yapabileceği manuel hataların önüne geçmiş ve sistemi daha otomatik ve akıllı hale getirmiştir. Bu çalışmanın önemli noktalarından biri de karlılık kriterine satın alma süreçlerinde ağırlık veriyor olmasıdır. Jiltoft vd. (2019) ise yaptığı çalışmasında bulanık mantık ve ANFIS sistemlerinin endüstride satın alma kararlarını etkileyecek düzeyde önemli olduğunu, yap ya da satın al kararlarının verilmesinde kullanılabileceğini

Confidential C

ortaya koymuřtur. Benzer řekilde tedarikçi seęimi gerektiren konularda ANFIS sistemlerinden faydalanılmaktadır.

Karaboga vd. (2019) ęalıřmasında ANFIS'in satın alma konularının haricinde literatürde birçok alanda kullanıldıęı ortaya koymuřtur. Bu alanlar: eęitim, saęlık, ekonomi, görüntü iřleme, tahminleme, üretim ve sistem modelleme, elektrik elektronik sistemleri olarak sayılabilir.

3. MATERYAL ve YÖNTEM

Bu çalışmada R studio paket programı 1.4.1106 versiyonunda ve Matlab paket programı 2022 b versiyonu kullanılmıştır.

3.1. Veri Madenciliği Yöntemleri

Bu tez çalışmasında, bir otomotiv işletmesinde 732 farklı sac malzeme içeren veri setinde doğru satın alma stratejilerini oluşturarak fiyat tahminlemek hedeflenmiş olup çalışma aşağıdaki adımlardan oluşmaktadır;

1. Gürültülü verilerin tanımlanması,
2. Veri madenciliğinde kümeleme analizi,
3. İdeal küme sayısının tespiti,
4. K-ortalamlar yöntemi ile maliyet tabanlı ürün grupları ve kümelerin oluşturulması
5. Kümeleme sonucundaki veriler ile satın alma stratejilerinin oluşturulması
6. Öznitelik seçim algoritmalarıyla fiyata en çok etki eden kriterin belirlenmesi
7. Malzeme fiyatlarının YSA ile tahminlenmesi,
8. Malzeme fiyatlarının ANFİS ile tahminlenmesi,
9. Fiyat tahmin modellerinin kıyaslanması
10. Fiyat tahminleme modelleriyle satın alma stratejilerinin oluşturulması

Gerçek hayat verileri hiçbir zaman mükemmel değildir, bu sebeple analizler yapılırken gürültülü verilerin de dahil edilmesi; verilerin yorumlanması, oluşturulan modelleri ve alınan kararları olumsuz yönde etkileyebilirler.

Gürültülü veriler, sınıflandırmayı da negatif yönde etkiler. Bu negatif etkiler, sınıfların sınırlarının bozulması, anlamsız küçük kümelerin oluşması veya önemli kümelerin örneklemlerinin değişmesi olarak sayılabilir (Liebchen vd., 2007).

Gürültülü verileri kolayca temizlemek için birçok yöntem mevcuttur. Bunlardan bazıları aşağıdaki gibi sıralanabilir:

- Güçlü Öğrenme (Robust Learners): Budama (pruning) yöntemi ile karar ağaçlarındaki gürültüye sebep olan olasılıkları azaltmaya çalışır. Budama işlemi karar ağaçları çalışma prensiplerine dayanmaktadır. Bu yöntem ile belirlenen hata toleransının üzerinde kalan verilere ait ağaç dalları budanır yani dikkate alınmadan devam edilir. Bu yöntem ile hata toleransı yüksek veriler sistemden çıkartılmış olur. Dolayısıyla gürültülü olasılığı da azalmış olur (Liebchen vd. 2007).
- Veri Cilalama (Data Polishing): Amacı mevcut veri setinde gürültülü verileri azaltmaya çalışmaktır. Veri cilalama yöntemine hatalı verilerin düzeltilmesi için grafiksel kümelemeler yapılarak birbiriyle benzer hataya sahip veri kümeleri oluşturulup hata düzeltilir. Ancak küçük veri setleri üzerinde çalışılması tavsiye edilir. Zira, zaman alan bir yöntemdir (Hosoe vd., 2021).
- Gürültü Filtreleri (Noise Filters): Veri setlerinde gürültüye yol açan senaryolar biliniyorsa, bu senaryoların veri setinden filtrelenerek ayrılmasıdır (Liebchen vd., 2007)

Bir satın alma veri setinde, strateji oluşturmak için çalışırken karşılaşılabilecek bazı problemler aşağıda sunulmaktadır. Burada bulunan problemler, istenmeyen sonuçlara ve kararlara yol açabilir.

- Sac satın alma veri setinin gürültülü veriler içeriyor olması,
- Gürültülü verilerin, veri setinin yanlış analiz edilmesine sebep olma riski,
- Uygun makine öğrenmesi algoritmasının seçilememesi,
- Verilerin yanlış sınıflandırılması,
- Veri setine ait ideal küme sayısının bilinmemesi,
- Projenin stratejik öneminin sadece proje bazlı yorumlanması,
- Sac üretim süreçleri ve malzeme alanında yetersiz bilgi,
- Yanlış zamanda, yanlış satın alma stratejisi oluşturulması.

Confidential C

Satın alma veri setlerinde görülecek gürültülü verilerin birçok kaynağı olabilir. Örneğin; satın alma çalışmasının fiyat girişi esnasında yapacağı hatalar, dosyaların güncel olmaması, yanlış hesaplamalar vb. gürültülü olarak tanımlanan senaryolar aşağıdaki gibidir. Bu senaryolar çoğaltılabilir.

- Referansların projesinin belli olmaması: Bir sac malzeme referansının hangi projede kullanılacağı veri setinde mutlaka bulunmalıdır. Bulunmaması halinde, proje maliyetleri belirlenirken hatalar ortaya çıkacaktır.
- Referansların seri üretimde olmaması: Seri üretim koşulları için yapılmak istenen bir analizde, seri ömrünü tamamlamış veya denemelerde kullanılan referansların yer almaması gerekir.
- Fiyatların anormal seviyelerde olması: Bir fiyat listesinde büyük sapmalar gösteren fiyatlar analiz edilmelidir. Bu fiyatlar özel bir malzemeye ait olabileceği gibi hatalı girilmiş olabilir. Örneğin çalışmaya konu veri setinde satın alma çalışmasının yanlışlıkla eksi değer girdiği fiyatlar mevcuttur.

Veri setinin normal dağılıp, dağılmadığını anlamak için yapılan Anderson Darling testinin formülü denklem 3.1’de aşağıdaki gibi gösterilmiştir:

$$AD = -n - \frac{1}{n} \sum (2i - 1) [\ln F(x_i) + \ln (1 - F(x_{n-i+1}))] \quad (3.1)$$

$X_{(i)}$: veriler artan düzende sıralandığında seçilen i. sıradaki örnek

$F(x_i)$: belirtilen dağılım için kümülatif dağılım fonksiyonu (CDF)

n : örneklem büyüklüğü

AD : Anderson Darling

Verilerin normal dağılıma uygunluğunu test etmek için aşağıda belirtilen hipotezler kurulmuştur:

H_0 : Data seti normal dağılıma sahip olması

H_1 : Data seti normal dağılıma sahip olmaması anlamına gelmektedir.

%95 güven aralığı için hesaplandığında, $P < 0.05$ ise H_0 reddedilir, veri seti normal dağılıma sahip değildir şeklinde yorumlanır. $P > 0.05$ ise H_1 reddedilir, veri seti normal dağılıma sahiptir şeklinde yorumlanır (Jeong vd., 2016) .

3.2. Makine Öğrenmesi Yöntemleri

Satın alma çalışmalarında, verinin normal dağılıp dağılmadığına göre seçilen makine öğrenmesi yöntemleri gözden geçirilmelidir. Öztuna vd. (2006) çalışmasında, Kolmogorov-Smirnov, Pearson, Jarqua-Bera ve Shapiro-Wilk testlerini dağılımlar için hatalar ve güçlük açısından incelemiş olup normal olmayan dağılıma sahip veri için düşük adetli (50 satır altı) veri setlerinde Shapiro-Wilk testinin ve Jarqua-Bera testinin en güçlü sonuçları verdiğini belirtmiştir. Daha yüksek sayıda veri içeren veri setleri için Anderson Darling testi tavsiye edilmektedir.

Bu çalışmadaki veri setinin normal dağılıma sahip olmadığı ortaya konmuştur ve gözetimsiz öğrenme yöntemi tercih edilmiştir. Gözetimsiz öğrenme tekniklerinin altında, en çok tercih edilen uygulamalardan biri olan k-ortalamlar yöntemi seçilmiştir.

Gözetimsiz öğrenme tekniklerinin altında literatürde sıklıkla kullanılan üç yöntem incelenmiştir. Bunlar, birliktelik analizleri altında yer alan Apriori algoritması ile kümeleme analizleri altında yer alan k-ortalamlar yöntemi ve hiyerarşik kümeleme yöntemidir.

3.2.1. Apriori Algoritması

Apriori algoritması, veriler arasındaki birliktelik kurallarını ortaya çıkarmayı amaçlamaktadır. Algoritma, sık karşılaştığı ilişkileri ve bağlantıları araştırmaktadır. Burada amaç, birbiri ile ilişkili ürünleri tespit etmek ve bu bilgiyi amaçlar doğrultusunda kullanmaktır. Özekes (2003) çalışmasında bir alışveriş mağazasında süt alan bir bireyin, bu ürünün yanına ekmek alması birer Apriori örneği olarak belirtmiştir. Çünkü, bu ürünler kendi içinde birbiriyle bağlantılıdır. Bu sebeple, market raflarında birbirine yakın konumlanırlar.

Confidential C

Ancak, bu çalışmada amaç sac malzeme satın alınırken hangi malzemelerle birlikte alındığı gibi bir sonuç çıkarmak değildir. Dolayısıyla, Apriori Algoritmasının çalışılması bu çalışma için tercih edilmemiştir.

3.2.2. K-ortalamlar Yöntemi

K-ortalamlar yöntemi en sık kullanılan kümeleme yöntemlerinden biridir. 1967 yılında MacQueen tarafından önerilen bu algoritma, büyük ve karmaşık verilerden basit ve anlaşılır veri kümelerine ulaşmayı amaçlamaktadır. Bu yöntemin adımları, başlangıç, tekrarlama ve iyileştirme ve durma olarak aşağıdaki gibidir (Tekin, 2018):

- Başlangıç aşamasında, her veri kendine en yakın rastgele bir küme merkezine atanır.
- Tekrarlama aşamasında ilgili küme merkezi, verilerin ortalaması olacak şekilde değiştirilir.
- İlgili verinin küme merkezine uzaklığı seçilen uzaklık metriği ile hesaplanır ve hata değeri belirlenir.
- İyileştirme aşamasında her veri yeniden kendine en yakın küme merkezine atanır.
- Hiçbir veri kümeler arası hareket yapamayacak hale geldiğinde durma işlemi gerçekleşir.

K-ortalamlar yönteminde en sık kullanılan uzaklık metrikler Öklid uzaklığı, Manhattan uzaklığı ve Minkowski uzaklığıdır.

Öklid uzaklığı, sık bir yöntem olup, iki nokta arasındaki mesafeyi ölçmenin bir yoludur. Bu denklem noktalar arasındaki farkların kareleri toplamının kare köküne eşittir (Saputra vd., 2019). Öklid uzaklığının formülasyonu denklem 3.2'de gösterilmektedir. Bu denklemde $i = 1 \dots n$, x_i değeri x_i değişkenine ait öz nitelik değerini, y_i değeri y değişkenine ait öz nitelik değerini, n ise toplam öz nitelik değerini vermektedir.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.2)$$

Bu çalışmada Öklid uzaklığı, uzaklık metriği olarak seçilmiştir.

Manhattan Uzaklığı, literatürde şehir blok uzaklığı olarak bilinmektedir. Bir verinin, diğer bir veriye uzaklığını ölçmektedir. Blokların yola olan uzaklığını hesaplama mantığına dayanmaktadır (Saputra vd., 2019). Formülü denklem 3.3'te gösterilmektedir. Bu denklemde $i = 1 \dots n$, x_i değeri x_i değişkenine ait öz nitelik değerini, y_i değeri y değişkenine ait öz nitelik değerini, n ise toplam öz nitelik değerini vermektedir.

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (3.3)$$

Minkowski uzaklığı, Öklid uzayında bir metriktir. Öklid uzaklığı ve Manhattan uzaklığının genelleştirilmesi ile oluşmuştur (Saputra vd., 2019). Minkowski uzaklığı denklem 3.4'te verilmiştir. Bu denklemde $i = 1 \dots n$, p değeri iki nokta arasındaki Minkowski uzaklığı, x_i değeri x değişkenine ait öz nitelik değerini, y_i değeri y değişkenine ait öz nitelik değerini, n ise toplam öz nitelik değerini vermektedir.

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p} \quad (3.4)$$

K-Ortalamalar yönteminde ideal küme sayısının tespiti önem arz eder. Verilerin gerekli sayıda kümelere ayrılması ve birbiriyle uyum içinde olması adına doğru sayıda kümeye ayrılması gerekmektedir. Ancak veri setine göre ideal küme sayısı değişmektedir. Dolayısıyla her çalışmada, veri setine özgü ideal küme sayısı belirlenmelidir.

Küme sayısını belirlemek için üç ayrı yöntem mevcuttur. Bunlar dirsek yöntemi, Silhouette yöntemi ve bağlantısallık yöntemleridir.

Dirsek yöntemi, veri gruplarının birbirine olan uzaklıklarının kareler toplamını hesaplama temeline dayanan bir yöntemdir. Veri kümelerinin birbirine benzerliklerini ölçmeyi ve

benzer verileri bir arada kümelemeyi hedeflemektedir. Dirsek yöntemi grafikte dirseğe benzer bir kırılmanın olduğu yerde ideal kümelemenin olduğunu belirtmektedir. İdeal küme noktasından sonra, grafikte düşüşler daha yavaş gözlemlenmektedir. (Cui, 2020; Patel vd., 2022). Dirsek yönteminin formülü denklem 3.5'te verilmiştir. P_i seçilen kümedeki i . nokta, C_N seçilen N . Küme merkezi ve N küme sayısını belirtmektedir.

$$\text{Dirsek Yöntemi} = \sum_{1.\text{Küme}} \text{uzaklık}(P_i, C_1)^2 + \dots + \sum_{n.\text{küme}} \text{uzaklık}(P_i, C_N)^N \quad (3.5)$$

Silhouette yöntemi, her değerın kümelerle arasındaki uzaklığa bakmaktadır ve kümelerdeki nesnelere bulunduğru yerlere göre kararlılıklarını göstermektedir. Silhouette değeri, 1'e yakın ise iyi kümelenemiş eğer -1'e yakın ise zayıf kümelenemiş verilerdir. Başka bir deyişle, Silhouette değeri (-1,1) olup maksimize edilmesi gereken bir değerdir (Shobha vd., 2017).

Silhouette yöntemi denklem 3.6'da gösterildiğri gibi hesaplanmaktadır. (Wang vd., 2019; Patel vd., 2022). Denklem 3.6'da a_i değeri i . verinin aynı kümedeki diğere verilere ortalama uzaklığı, b_i değeri i . verinin en yakın komşu kümelere olan uzaklığı ve $s(i)$ = Silhouette değeri olarak verilmiştir.

$$s(i) = \frac{b_i - a_i}{\max(b_i - a_i)} \quad (3.6)$$

Gap İstatistik Yöntemi, kümelerin, gerçekleşen değerlerinin karelerinin toplamının, beklenen değerlerine göre logaritmik değerce farkını karşılaştırmaktadır. Bu iki değere arasındaki Gap büyüklüğü, rastgele uniform dağılımdan ne kadar uzakta olduğunu belirtmektedir. Gap ne kadar büyükse, sonuç o kadar anlamlıdır. Başka bir deyişle, Gap büyüklükleri farklı kümelere işaret etmektedir (Patel vd., 2022). Denklem 3.7'de gösterildiğri üzere, n küme sayısı, i ve j indisler, C_r kümelere olan uzaklık, D_R ise gap istatistik değeridir.

$$D_R = \sum_{i \in C_R}^n \sum_{j \in C_R}^n d_{ij} \quad (3.7)$$

Confidential C

Yukarıda açıklanan yöntemler uygulandıktan sonra elde edilen küme sayısının kontrolü ve doğrulaması yapılması gerekmektedir. Özetle, bulunan kümelerin ne kadar doğru sonuçlar verdiği araştırılmalıdır. Bu validasyon sürecinde üç ayrı yöntem takip edilmektedir, bu yöntemler bağlantı indeksi, Silhouette ve Dunn indeksidir (Shobha vd. 2017).

Bağlantı indeksi, verilerdeki yerel yoğunlukları gözlemlemekte ve veri gruplarını en yakın komşulara göre kümelere ayırmaktadır. Bu indeks, bir verinin kendisine en yakın komşu kümeye ne düzeyde yerleştiğini göstermektedir. Bağlantı indeksinden çıkan sonuçlar sıfır ve sonsuz arasında olup, minimum değerinde elde edilmesi gerekmektedir. Burada N gözlem sayısı, m toplam veri sayısı, L veriye en yakın küme sayısı, C gözlemlere ait ayrık kümelerdir (Shobha vd., 2017). Bağlantı indeksinin formülü denklem 3.8’de gösterilmiştir.

$$Conn(C) = \sum_{i=1}^N \sum_{j=1}^L x_{i,m_i,j} \quad (3.8)$$

Silhouette yöntemi ile ideal küme sayısı hesaplanırken, denklem 3.6 ile hesaplanır. Silhouette yönteminin iyi kümeleme yapıldığı sonucu verebilmesi için çıkan sonucun 1’e yakın olması beklenir.

Dunn indeksi, aynı kümede olmayan en yakın iki verinin uzaklığının, aynı kümede olan birbirine en uzak iki veriye uzaklığına oranıdır. Dunn indeksi (0,1) aralığında olup, maksimize edilmesi gereken bir değerdir. Dunn indeksinin formülü denklem 3.9’da gösterilmiştir. Burada *diam* değeri C_m kümeleri arasındaki maksimum uzaklıktır (Shobha vd., 2017). Denklem 3.9’da c_i i kümesinin benzerlik fonksiyonunu, c_k k kümesinin benzerlik fonksiyonunu *diam* ise kümeler arası maksimum uzaklığı ifade etmektedir.

$$D(c) = \frac{\min_{k,c_i \in c_k, c_k \neq c_i} (\min_{i \in c_k, j \in c_i} dist(i,j))}{\max_{c_m} diam(c_m)} \quad (3.9)$$

3.3. Satın Alma Fiyatlarına Etki Eden Kriter Belirleme Yöntemleri

Satın alma malzemelerinin fiyatları belirlenirken hangi kriterin fiyata, diğerlerine oranla daha fazla etki ettiğinin bilinmesi, satın alma kararlarını etkilemektedir.

Öz nitelik seçimi için en çok kullanılan üç algoritma olan MRMR algoritması, karşılıklı bilgi öznitelik seçim algoritması (KBÖS) ve korelasyon algoritmaları arasından, MRMR algoritması ile çalışma yapılmıştır. Diğer algoritmaların seçilmemesinde ise, KBÖS algoritması rastgele değişkenler arasındaki düzensizliği ölçerek sonuç vermesi ve tüm verileri kontrol etmemesi, korelasyon katsayısının ise tüketim gibi değişken kriterler üzerinde 0'a yakın sonuç vererek tüketim kriteri fiyat üzerinde etkisiz gibi göstererek yanıtmasından dolayı tercih edilmemiştir.

Fiyat tahminlemesi yapılırken özniteliklerin gerçekten fiyata etki edip etmediğini anlamak adına MRMR algoritması da denen Minimum Redundancy Maximum Relevance Algoritması kullanılabilir. Bu algoritma sınıf etiketleri oluşturarak en ilişkili değişkeni bulmaya ve en ilişkisiz değişkeni ortaya çıkarmaya çalışmaktadır (Gülgezen vd., 2022).

MRMR algoritmasının formülasyonu denklem 3.10'da gösterilmiştir.

$$I(X, Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p_1(x) \cdot p_2(y)} \right) \quad (3.10)$$

$p(x, y)$ = x ve y nin birleşik olasılık fonksiyonu

$p_1(x)$ ve $p_2(y)$ = x ve y değişkenlerinin marjinal olasılık dağılım fonksiyonları

Karşılıklı bilgi öznitelik seçim algoritması (KBÖSA) ile karşılıklı bilgiye dayanan hesaplamalar yapılır ve bu sayede öznitelikler seçilir. Temeli entropi yani düzensizliktir. Veriler arasındaki düzensizliği ölçer. Denklem 3.11'de $p(x, y)$ değeri X ve Y'nin birleşik olasılık fonksiyonunu ve $p_1(x)$ ve $p_2(y)$ değerleri x ve y değişkenlerinin marjinal olasılık dağılım fonksiyonlarını belirtmektedir (Çelik vd., 2015).

Confidential C

$$I(X, Y) = \iint p(x, y) \log \left(\frac{p(x, y)}{p_1(x) \cdot p_2(y)} \right) dx dy \quad (3.11)$$

Korelasyon değeri değişkenler arasındaki ilişkinin yönünü ve gücünü belirtir. Korelasyon sayısı 1'e ne kadar yaklaşırsa veriler o kadar güçlü bir ilişkiye sahiptir denir. Ancak fiyat kriterleri üzerinde yorum yaparken tüketim gibi standart sapması yüksek veri setlerinde yanlış sonuç verdiği ve tüketimin fiyat üzerinde etkisiz olduğunu yorumladığı görülmüştür. Bu sebepten tercih edilmemiştir.

3.4. Yapay Sinir Ağları

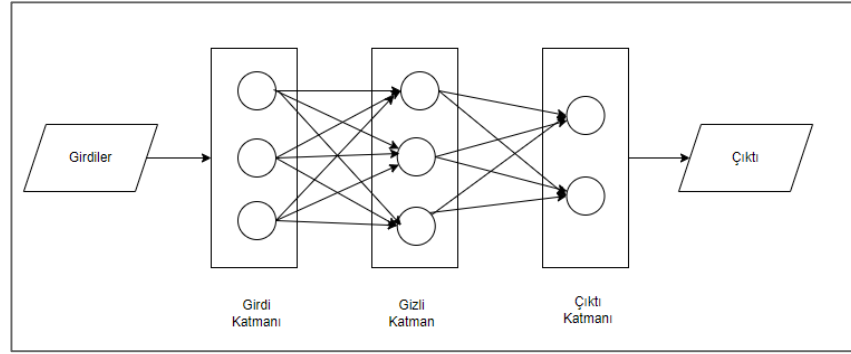
İnsanoğlu, büyük verilerin anlaşılır ve işlevsel hale getirilmesi üzerinde uzun yıllardır çalışmaktadır. Bu kapsamda yer alan yapay zekâ teknolojilerinden biri olan yapay sinir ağı (YSA) üzerindeki ilk çalışma Warren McCulloch ve Walter Pitts tarafından 1943 yılında yapılan biyolojik sinir ağlarını taklit eden çalışmaya dayanmaktadır. Bu çalışma literatürde ilk YSA çalışması olarak anılmaktadır. Burada biyolojik yapıyı, YSA elemanlarına dönüştürmek gerektiği için, karşılıkları Çizelge 1'de verilmiştir (Öztürk vd. 2018).

Çizelge 1. Biyolojik sinir sistemi elemanlarının, yapay sinir ağı karşılıkları

Biyolojik Sinir Sistemi Elemanı	Yapay Sinir Ağı Elemanı
Nöron	İşlemci Eleman
Dentrit	Toplama Fonksiyonu
Hücre Gövdesi	Transfer Fonksiyonu
Akson	Yapay Nöron Çıkışı
Sinaps	Ağırlıklar

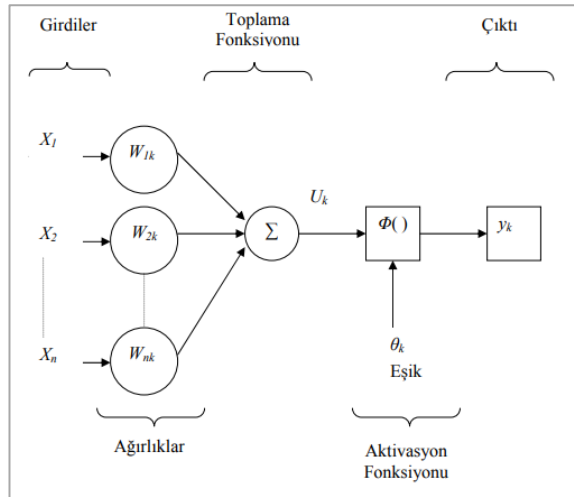
YSA insan beynindeki öğrenme sisteminin bilgisayar ortamında matematiksel modellenmesi olarak tanımlanabilir. İnsan beyninden esinlenen, temeli insan beynindeki biyolojik üniteler olan nöronlara dayanan bu sistem, zamanla bilgisayar ortamlarında geliştirilerek ve modellenerek YSA oluşturulmuştur (Sutcu, 2018).

Tolon vd. (2008) yaptıkları çalışmada YSA üç temel bölüme ayırmıştır. Bu katmanlar girdi katmanı, gizli katman ve çıktı katmanıdır. Girdi değişkenleri, girdi katmanında tanımlanır, gizli katmanlarda gerçekleşen etkileşimler sonrasında çıktı değişkenleri oluşur. Gizli katmanda dış ortamla bağlantısı olmayan sinirler mevcuttur ve bu sinirler gelen verileri işlemlerden geçirerek çıktı katmanına gönderir. Ağ yapısı Şekil 2’de gösterilmiştir.



Şekil 2. YSA yapısı

Sutcu (2018) çalışmasında yapay sinir hücrelerini beş bölümde incelemiştir. Bunlar; girdiler, ağırlıklar, toplama fonksiyonu (birleştirme fonksiyonu), aktivasyon fonksiyonu ve çıktılar olarak ayrılmaktadır. Şekil 3’de yapay sinir hücresi modeli verilmiştir.



Şekil 3.Yapay sinir hücresi modeli (Aksoy, 2012)

YSA nöron gruplarından oluşmaktadır. Nöron, diğer nöronlardan gelen bilgileri toplar, eşik değeri ile karşılaştırarak aktivasyon potansiyelini hesaplar ve bu potansiyeli aktivasyon fonksiyonundan geçirerek çıkış sinyalini üretir. Her nöron, komşularına bağlantı kuvvetlerini gösteren farklı katsayılar ile bağlıdır. Öğrenme bu katsayıların ayarlanması ile gerçekleşir (Aksoy, 2012).

Denklem 3.12’de k nöronunun matematiksel tanımı verilmiştir.

$$U_k = \sum_{j=1}^n W_{jk} \cdot X_j \quad (3.12)$$

U_k = k nöronu

W_{jk} = Ağırlık değeri

X_j = Girdi değeri

Çıktı değeri olan y_k değeri ise, denklem 3.13’deki gibi hesaplanır.

$$y_k = \varphi(U_k - \theta_k) \quad (3.13)$$

θ_k = Eşik değeri

y_k = Çıktı değeri

φ = Aktivasyon fonksiyonu

k nöronunun aktivite seviyesi (V_k) denklem 3.14’de gösterilmiştir.

$$V_k = U_k - \theta_k \quad (3.14)$$

Başka bir deyişle, çıktı eşik aktivasyon fonksiyonu ve aktivite seviyesinin çarpımı ile elde edilir. Denklem 3.15’de gösterilmiştir.

$$y_k = \varphi(V_k) \quad (3.15)$$

Aksoy (2012) çalışmasında YSA elemanlarını aşağıdaki gibi tanımlamıştır:

Confidential C

- $x_1, x_2, \dots, x_n =$ Girdilerdir. Yapay sinir hücresine dışarıdan gelen bilgilerdir. Girdiler kendinden önceki sinir hücresinden veya çevreden gelebilir.
- $w_{1k}, w_{2k}, \dots, w_{nk} = k$ nöronuna ait ağırlıklardır. Bu nörona bağlı bir j sinapsisinde yer alan x_n sinyali, w_{nk} ağırlığı ile çarpılır. Ağırlıklar değişken veya sabit değer alabilirler. Bu ağırlıkların pozitif olması arttırıcı etki, negatif olması ise azaltıcı etki anlamına gelir.
- U_k doğrusal toplama fonksiyonu: Sinir hücresine gelen net girişi hesaplayan fonksiyondur. Farklı formülasyonları olmakla birlikte, ağırlıklı toplamı en yaygın kullanılmaktadır.
- θ_k eşik değeri: Aktivasyon fonksiyonunun net girişlerinin etkilerini azaltmak için çalışan parametredir. Yani, toplanan iletinin bir sonraki seviye ile karşılaştırıp negatif aktivasyon potansiyelini aramaktadır.
- $\varphi()$ aktivasyon fonksiyonu: Aktivasyon fonksiyonunun girdi değerlerinin seviyesine bakarak nöron çıktılarını belirler. Genelde $[0,1]$ veya $[-1,1]$ sınırları içerisinde kalmasını sağlar. Aktivasyon fonksiyonları, doğrusal fonksiyon, parçalı doğrusal fonksiyon, sigmoid fonksiyon ve eşik değeri fonksiyonu olarak çeşitlendirilebilir. Yaygın olarak, doğrusal ve sigmoid fonksiyonlar kullanılmaktadır.
- y_k k . nöronun çıkış sinyali: Aktivasyon fonksiyonundan geçen iletinin çıkış değeridir.

3.4.1. Yapay Sinir Ağı Çeşitleri

YSA modelleri, tek katmanlı algılayıcılar, çok katmanlı algılayıcılar, ileri beslemeli YSA ve geri beslemeli YSA olarak dört gruba ayrılır (Öztürk vd., 2018).

- Tek katmanlı algılayıcılar: Sadece girdi ve çıktılardan oluşur. Çıktı doğrusal bir fonksiyondur ve $[-1,1]$ aralığında değer alır. Çıktı değeri 1'e eşit ise birinci kümeye, -1'e eşit ise ikinci kümeye konur.
- Çok katmanlı algılayıcılar: Doğrusal olmayan fonksiyonlarda, birçok nöronun belli bir üstünlük derecesine sahip olduğu algılayıcılarıdır.

- İleri beslemeli yapay sinir ağı: Nöronlar Şekil 2’de gösterildiği gibi girdilerden, çıktılara doğru düzenli katmanlar ile ilerlemektedir. Girdi bilgisi, gizli katmanlardan hücrelere gönderilir ve çıkış katmanında işlenerek çıktı gönderilir.
- Geri beslemeleri yapay sinir ağı: Bir nöronun çıktısı kendinden önceki katmanda veya kendi katmanındaki farklı bir nöronun girdisi olabilir. Bu yapı sebebiyle doğrusal olmayan bir davranış sergiler (Öztürk vd. 2018).

3.4.2. Matlab Programında Yapay Sinir Ağları

Matlab paket programında YSA’yı kolayca anlamak ve analiz etmek için bazı araçlar mevcuttur. Veri setinin eğitilmesi için Matlab birçok eğitim algoritmasını bünyesinde bulundurmaktadır. Bu algoritmalar Çizelge 2’de gösterilmiştir.

Çizelge 2. Matlab’te bulunan eğitim algoritmaları (Beale vd., 2018)

Matlab Fonksiyonu	Eğitim Algoritması
trainlm	Levenberg-Marquardt
trainbr	Bayes Regülerizasyonu
trainbfg	BFGS Quasi-Newton
trainrp	Esnek Geri Yayılım (Resilient Backpropagation)
trainscg	Ölçekli Eşlenik Gradyan (Scaled Conjugate Gradient)
traincgb	Powell/Beale Yeniden Başlatma ile Eşlenik Gradyan (Conjugate Gradient with Powell/Beale Restarts)
traincgf	Fletcher-Powell Eşlenik Gradyan (Fletcher-Powell Conjugate Gradient)
traincgp	Polak-Ribière Eşlenik Gradyan (Polak-Ribière Conjugate Gradient)
trainoss	Tek Adımlı Sekant (One Step Secant)
traingdx	Değişken Öğrenme Hızı Olan İnişli Gradyan (Variable Learning Rate Gradient Descent)
traingdm	Momentumlu İnişli Gradyan(Gradient Descent with Momentum)
traingd	İnişli Gradyan (Gradient Descent)

Bu kılavuzda, en hızlı eğitim fonksiyonun “*trainlm*” (Levenberg-Marquardt) olduğu ve araçların otomatik olarak bu fonksiyonu kullandığı belirtilmiştir. Ancak binlerce satırdan oluşan veri setlerinde fazla zaman ve bilgisayar belleğine ihtiyaç duyacağı belirtilmiştir. Binlerce satırdan oluşan veri setlerinde “*trainscg*” (Ölçekli Eşlenik Gradyan) ve “*traingrp*” (Esnek Geri Yayılım) algoritmaları tavsiye edilmiştir (Beale vd. 2018).

Levenberg-Marquardt algoritması Matlab ile geliştirilen bir algoritma olup, geri yayılım yapan ve momentum içeren algoritmalara göre 10 ile 100 kat daha hızlı çalıştığı belirtilmiştir (Beale vd. 2018). Bu sebepten tez kapsamında Levenberg-Marquardt algoritması kullanılmıştır. Ağın eğitimi önemli olduğu gibi, eğitimin doğru yerde durması da önem arz etmektedir. Bu işlemde eğitimi durdurma kriterleri ile yapılmaktadır (Beale vd. 2018). Matlab’da bulunan eğitim durdurma kriterleri ve açıklamaları Çizelge 3’te gösterilmiştir

Çizelge 3. Ağın eğitimini durdurma kriterleri (Beale vd. 2018)

Parametre	Durdurma Kriteri	Durdurma Kriterinin Açıklaması
Min_grad	Mimum Gradient Magnitude	Ağ, minimum gradyan büyüklüğüne ulaştıncaya, eğitim durur.
Max_fail	Maximum Number of Validation Increases	Ağ, maksimum doğrulamayı yaptığı noktada eğitim durur.
Time	Maximum Training Time	En uzun eğitim süresine ulaştıncaya eğitim durur.
Goal	Minimum Performance Value	En küçük performans değerine ulaştıncaya eğitim durur.
Epochs	Maximum Number of Training Epochs (Iterations)	Maksimum eğitim iterasyonu sayısına ulaşıldığında ve yineleme başladığında eğitim durur.
Stop training	Eğitimi durdurma butonu	Kritere bağlı olmadan eğitim durdurulur.

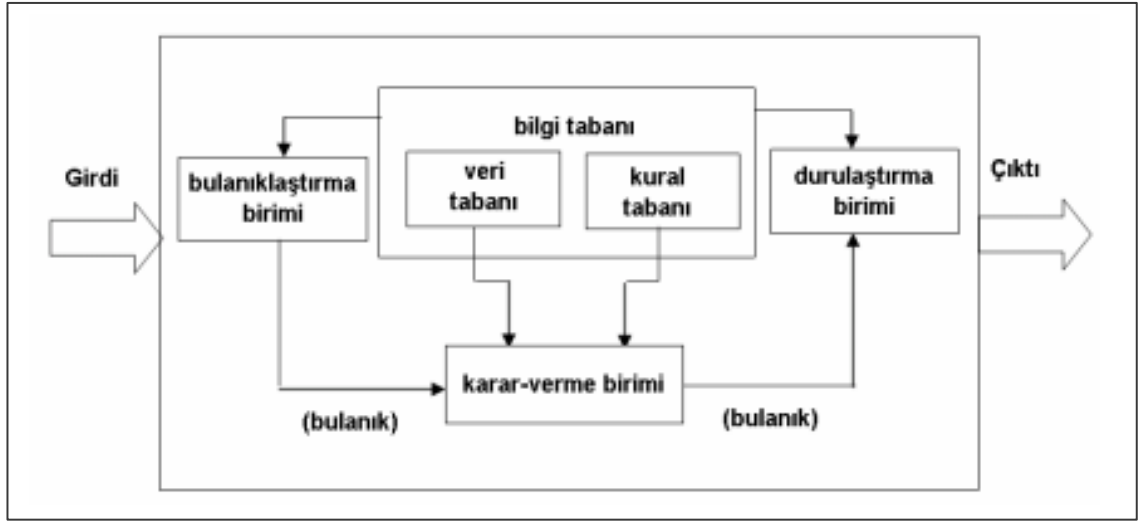
Matlab performans kriteri olarak ortalama karesel hata (MSE - Mean Square Error) dikkate almaktadır. Bu sebepten çalışmada MSE verileri takip edilmiştir.

3.5. Bütünleşik Bulanık Sistemler

Bütünleşik bulanık sistemleri anlamak için öncelikle Bulanık Mantığı (BM) anlamak gerekmektedir. 1965 yılında Zadeh tarafından ortaya atılan bulanık mantık, önemli bir yapay zeka tekniğidir. Geleneksel kümeleme teorilerinde 0 ve 1 ile ifadeler söz konusudur. Yani, bir eleman bir kümenin elemanı değilse 0, elemanıysa 1 değerini alır. Ancak bulanık mantığa göre her üye bir gruba farklı üyelik derecesi ile dahil olabilir. Bulanık mantık, günlük hayatta karşılaşılan belirsiz durumlar için belli kurallar ile karar verme mekanizmaları oluşturmayı hedefler (Karaboga vd., 2019).

Bulanık mantıkta, eğer X koşulu varsa sonucu Y'dir gibi ifade edilen kurallar vardır ve bu kurallara göre çıkarımlar oluşturulmaktadır. Örneğin "Az bulunan bir ürünün fiyatı yüksektir" şeklinde bir kuralda ürün bulunurluğu ve fiyat sözeldir ancak az ve yüksek üyelik fonksiyonlarının çıktılarıdır. Sayısal olarak yorum yapmanın zor olduğu noktalarda bu sistemler çözüm sunabilmek için geliştirilmiştir (Jang 1993).

Şekil 4'de bulanık çıkarım sisteminin şeması yer almaktadır.



Şekil 4. Bulanık çıkarım sistemi (Yücel, 2010)

Girdi kısmında girdi seti yer almaktadır. Bulanıklaştırma birimi, girdi verilerini bulanık giriş kümelerine dönüştürmektedir. Bulanık çıkarımın önemli bir parçası da bilgi tabanıdır. Yani veri tabanı ve kuralları kullanarak bir karar verme birimi yaratmaktadır.

Veri tabanı, bulanık kümelerin üyelik fonksiyonlarıdır. Kural tabanı ise veri tabanındaki

Confidential C

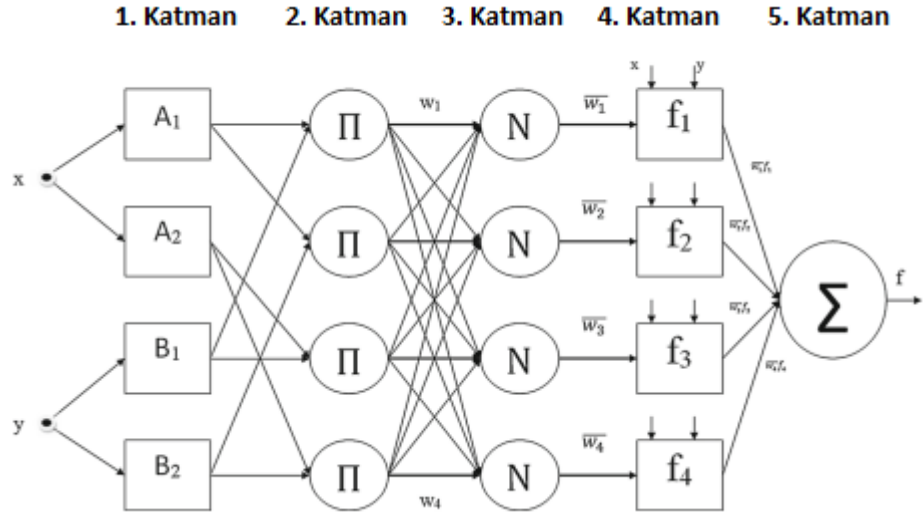
giriş çıkış değişkenlerini mantıksal kurallara bağlar. Sonrasında durulaştırma biriminde, bulanık değerler kesin sayılara dönüştürülür ve çıktı birimiyle değerler elde edilir (Yücel, 2010).

3.6. Adaptif Ağ Yapısına Dayalı Bulanık Çıkarım Sistemleri (ANFIS)

Bulanık mantık çıkarım sistemlerinin önemli üyelerinden biri de adaptif ağ yapısına dayalı bulanık çıkarım sistemleri (ANFIS)'tir. Adaptif ağ, çeşitli sayıda birbirine bağlı düğümler içeren, eldeki girdi veri setleri ile çıktı verilerini ortaya koyan ağ yapısına verilen isimdir (Yücel 2010; Aksoy 2012)

Bulanık mantıktaki adaptif ağlar ve YSA birbirini tamamlayıcı sistemlerdir. Her iki sistemin de birbirine göre üstün yanları olduğu için birlikte kullanılmaları avantajlı olabilmektedir. Bu avantajlar YSA kolay eğitimi, hesaplama gücü, hızlı çalışması iken, bulanık mantığın insan yargılarına benzer kurallar sistemini kullanması olarak sayılabilir (Aksoy 2012). ANFIS'in en önemli özelliklerinden biri de melez bir öğrenme algoritmasına sahip olmasıdır. Düğümlerin dördüncü katmana kadar soncul parametreleri en küçük kareler yöntemiyle belirlenir, geriye doğru gidişte ise öncül parametreler gradyan azaltım yöntemleri ile belirlenir (Yücel, 2010).

Şekil 5'de iki girdi, bir çıktıdan oluşan ANFIS yapısı gösterilmektedir. Katmanların açıklaması aşağıda yapılmıştır:



İki girdili bir çıktılı ANFIS yapısı (Karaboga vd., 2019)

1.Katman: Bulanıklaştırma katmanı olarak adlandırılır. Üyelik fonksiyonlarının yer aldığı katmandır. Girdi değerlerinden bulanık kümeler elde etmek için kullanılan üyelik fonksiyonları öncül parametre olarak adlandırılır (a, b, c değerleri parametre seti olarak adlandırılır. Denklem 3.16’da verilmiştir.

$$O_i^1 = \mu_{A_i}(x) = \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}} \quad (3.16)$$

A_i = i.düğümün sözel değişkeni

$O_i^1 = \mu_{A_i}(x)$ = Birinci katmana ait A_i üyelik fonksiyonu

2. Katman: Kural katmanı olarak adlandırılır. Ağırlıkları w_i 'yi bulmak için üyelik fonksiyonları birbiriyle çarpılır. $i = 1, 2$ olmak üzere denklem 3.17’de ağırlık değeri verilmiştir.

$$w_i = O_i^2 = \mu_{A_i}(x) \cdot \mu_{B_i}(y) \quad (3.17)$$

$\mu_{A_i}(x)$ = A_i ye ait üyelik fonksiyonu

$\mu_{B_i}(y)$ = B_i ye ait üyelik fonksiyonu

$w_i = O_i^2 = 2.$ katmana ait üyelik fonksiyonunun ağırlığı

3.Katman: Normalizasyon katmanı olarak isimlendirilir. Her kural için ağırlık toplam ağırlığa oranını hesaplamaktadır. $i = 1,2,3,4$ olmak üzere denklem 3.18'de verilmiştir.

$$\bar{w}_i = O_i^3 = \frac{w_i}{w_1 + w_2 + w_3 + w_4} \quad (3.18)$$

$\bar{w}_i = i.$ nöronun ağırlığı

$\bar{w}_i = O_i^3 = 3.$ katmana ait üyelik fonksiyonu ağırlığı

4.Katman: Durulaştırma katmanı olarak isimlendirilir. Belirlenen ağırlıklar ve fonksiyon değerleri çarpılarak elde edilir. Denklem 3.19'da durulaştırma katmanına ait formül verilmiştir.

$$\bar{w}_i \cdot f_i = O_i^4 = \bar{w}_i \cdot (p_i \cdot x + q_i \cdot y + r_i) \quad (3.19)$$

$\bar{w}_i = i.$ nöronun ağırlığı

$\bar{w}_i \cdot f_i = O_i^4 = 4.$ Katmana ait üyelik fonksiyonu çıktısı

$(p_i, q_i, r_i) =$ Soncul parametreler

5.Katman: Toplama katmanı olarak isimlendirilir. Bütün düğümlerden gelen sinyalleri toplar. Denklem 3.20'de gösterildiği şekilde hesaplanmaktadır.

$$O_i^5 = \sum_i \bar{w}_i \cdot f_i = \frac{\sum_i w_i \cdot f_i}{\sum_i w_i} \quad (3.20)$$

ANFIS sisteminin eğitilmesi konusunda Çizelge 2'de yer alan eğitim algoritmaları kullanılabilirle beraber genetik algoritma, karınca koloni algoritması, parçacık sürü algoritması, arı algoritmaları farklı algoritmaların da kullanılabilirdiği ortaya konmuştur. ANFIS sistemi çalıştırılmadan önce mutlaka üyelik fonksiyonları ve ağ tabanlı bulanık çıkarım sistemi (FIS) tanımlanmalıdır (Karaboga, 2019).

Confidential C

4. BULGULAR VE TARTIŞMA

Günlük hayatta satın alma çalışanının karşılaştığı birçok problem vardır. Örneğin bir ürün grubuna ait fiyat veri setinde gürültülü veriler mevcut ise, bunların tespiti ve doğru şekilde temizlenmesi gerekmektedir. Bunun için, satın alma çalışanının hatasız bir şekilde temizlemesi gereken verileri bulması ve müdahale etmesi gerekmektedir. Fakat günlük hayatta, yanlış verilere müdahale edilmesi, silinmemesi gereken verilerin silinmesi veya verilerin yanlış yorumlanması gibi problemlerle karşılaşmaktadır. Tüm bu hatalarla karşılaşmamak için bu çalışmada veri madenciliği yöntemleri ile önışleme yaparak gürültülü veriler temizlenmiştir.

Veri madenciliği sonucunda temiz bir veri seti elde edilmiştir. Burada yapılan analiz, cirosu en yüksek referansları gruplayarak bu referanslar özelinde kümeleme ile karlılık getirmektir.

Kümeler belirlendikten sonra bir sonraki adım fiyat tahminlemektir. Günlük hayatta satın alma çalışanları bir malzeme için tedarikçilerden teklif almaktadır. Bu teklifler değerlendirilirken, teklifin gerçekten mantıklı bir seviyede olup olmadığını anlamak için farklı tedarikçilerden teklif almak, pazarı araştırma gibi yöntemler izlenebilmektedir. Fakat bu yöntemler satın alma çalışanının zaman kaybetmesine veya malzemeyi satın alma kararını geç verip daha yüksek fiyattan satın alma yapmasına sebep olabilir. Bu gibi sorunlara karşılık fiyat tahminlemesi yapabilen bir sistemden faydalanılması satın alma departmanında yaşanabilecek muhtemel gecikmeler ve yanlış karar vermeyi engellemektedir.

Sac malzeme fiyatlarını oluşturan kriterler; kesim tipi, boy, kalınlık, tüketim ve ürün grubudur. Başka bir deyişle, yeni bir sac malzeme fiyatlanırken bu kriterler ile fiyat belirlenmektedir. Burada hangi kriterin diğer kriterlere göre daha önemli olduğunu ayırt edilmesi gerekmektedir. Tecrübesiz bir satın alma çalışanı ürünü tanımadığı için hangi kritere dikkat edilmesi gerektiğini bilemeyebilir. En önemli kriterlerin belirlenmesinde öznitelik seçim yöntemlerinden MRMR algoritması kullanılmıştır. Bu sayede tecrübesiz

ve satın almasını yapacağı ürünü tanımayan bir çalışan dahi hangi kritere dikkat etmesi gerektiğini bilmektedir.

Fiyat tahminleme sisteminin daha kullanışlı olması için satın alma departmanındaki çalışanın tecrübesi mutlaka göz önüne alınmalıdır. Yapay sinir ağları ile elde edilen fiyat tahminleme sistemi, tecrübesiz satın alma çalışanları için kullanışlı iken, tecrübeli satın alma çalışanlarının yapay sinir ağı kullanması tecrübelerini ve bilgilerini kullanamamalarına yol açabilir. Bu sebepten tecrübeli satın alma çalışanları için ANFIS kullanılmış ve tecrübeleriyle kural oluşturularak fiyat tahminlemeleri gerçekleştirilmiştir.

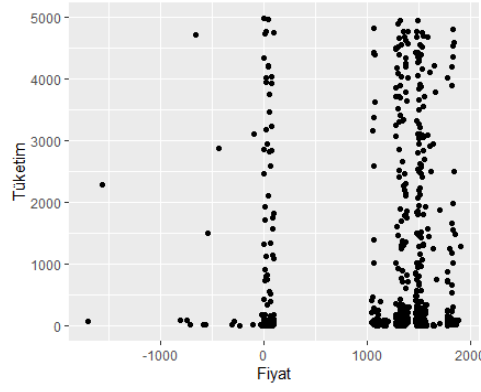
Bu sayede tecrübesiz satın alma çalışanı, yapay sinir ağları ile fiyat tahminlerken tecrübe kazanacak, tecrübeli satın alma çalışanları ise, yılların birikimi ile elde ettiği tecrübelerini makine öğrenmesi yöntemleri ile birleştirecektir.

Satın alma çalışanları günlük hayatta karşılaştığı problemler bu sayede çözülecek ve doğru fiyat tahminleri ile güçlü stratejiler kurulacaktır.

4.1. Veri Madenciliği Yöntemleri ile Yapılan Çalışmalar

Çalışmanın yapıldığı veri setindeki gürültülü verileri doğru yorumlamak ve temizleyebilmek için, ilk adımda veri setinin iyi tanınması ve analiz edilmesi gerekmektedir. Mevcut durumun anlaşılması adına mevcut veri setinin grafiği oluşturulmuştur.

Şekil 6'ya bakıldığında fiyatı sıfır ve negatifte bulunan birçok değer olduğu gözlemlenmektedir. Ancak bir satın alma operasyonunda, bir ürünün fiyatı sıfır veya negatif olmamalıdır. Eğer, veri seti bu şekilde gürültülü verileri barındırırken bir analiz yapılacak olursa, analiz sonuçlarının sağlıklı olmaması veya yanlış yönlendirmeye açık olması gibi sorunlarla karşılaşılabilir.

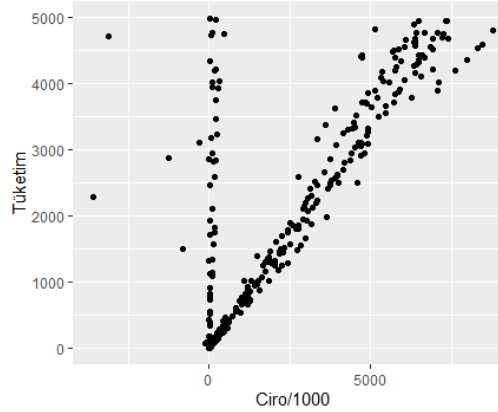


Şekil 5. Gürültülü veriler barındıran veri setinin fiyat-tüketim trafiği

Şekil 7’de ise grafik ciro ve tüketim parametreleri ile oluşturulmuştur. Burada ciro miktarları çok yüksek olduğu için 1000’e bölünerek grafiğe eklenmiştir. Ciro, denklem 4.1’de gösterildiği gibi bulunmaktadır.

$$\text{Ciro} = \text{Yıllık Tüketim} * \text{Birim Fiyat} \quad (4.1)$$

Şekil 6’da görülen sıfır ve negatif değer problemi Şekil7’deki verilerde de gözlemlenmektedir. Bu gruptaki verilerin de aynı şekilde gürültülü veri olarak etiketlenmesi gerekmektedir.

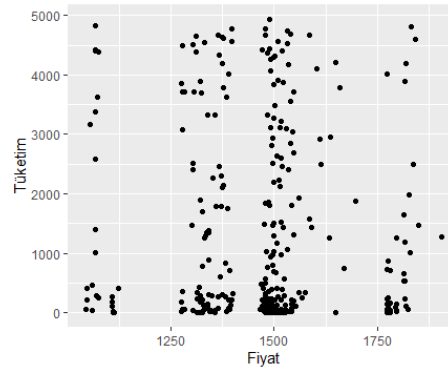


Şekil 6. Gürültülü veriler barındıran veri setinin ciro-tüketim grafiği

Gürültülü verilerin temizlenmesi üç adımda R studio paket programında yapılmıştır:

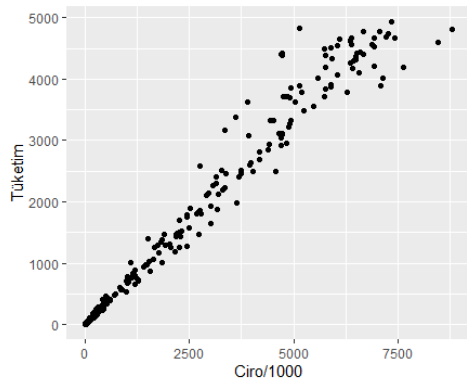
- Adım 1: Fiyatı negatif olmayan ve 500 €/ton'dan yüksek verilerin seçilmesi (500 €/ton temsili bir rakam olarak alınmıştır.)
- Adım 2: Sadece seri üretimde devam eden referansların seçilmesi
- Adım 3: Proje sütunu boş olan tüm satırların silinmesi

R studio'da yazılan kodlar EK 2'de yer almaktadır. Bu işlemler sonucunda 732 satır veri, 414 satır veriye düşürülmüştür. Şekil 8'de görüldüğü üzere, gürültülü veri olarak adlandırılan fiyatı sıfır veya negatif gelen veriler temizlenmiştir.



Şekil 7. Gürültülü verilerin temizlenmesinden sonra fiyat-tüketim grafiği

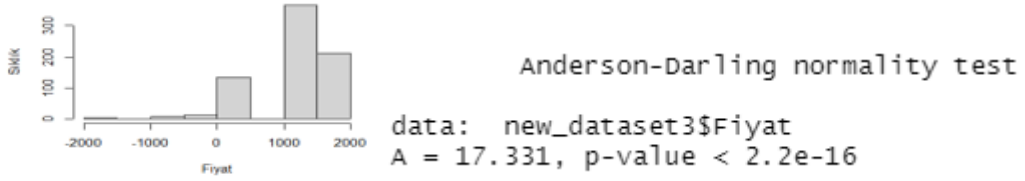
Şekil 9'da görüleceği üzere, Şekil 8'de yer alan sıfır ve negatif değerler ortadan kaldırılmıştır.



Şekil 8. Gürültülü veriler temizlenmesinden sonra ciro-tüketim grafiği

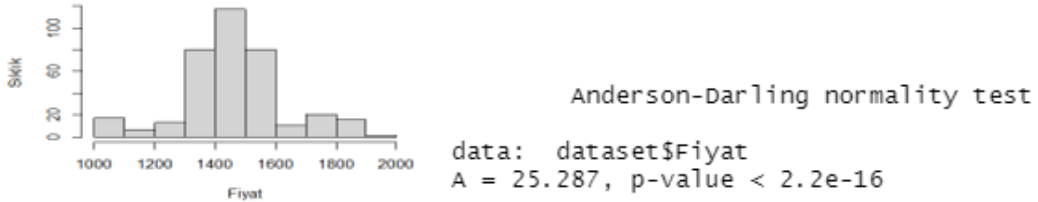
Veri temizliği tamamlandıktan sonra veri setinin dağılımının normal olup olmadığının tespit edilmesi gerekmektedir. Bunun için Anderson Darling testi yapılmıştır.

Şekil 10'da gürültülü veriler temizlenmeden önceki durumun histogram grafiği ve Anderson Darling testi sonuçları sunulmuştur.



Şekil 9. Gürültülü veriler temizlenmeden önceki duruma ait grafik

Test sonuçlarına bakıldığında A değeri 17,331 bulunmuş olup $p < 0.05$ 'tir (p -değeri = 2.2×10^{-16}). Dolayısıyla, H_0 hipotezi reddedilir. Anderson Darling testi, gürültülü veriler temizlenmeden önce verilerin normal dağılıma sahip olmadığını belirtmektedir.



Şekil 10. Gürültülü veriler temizlendikten sonraki duruma ait grafik

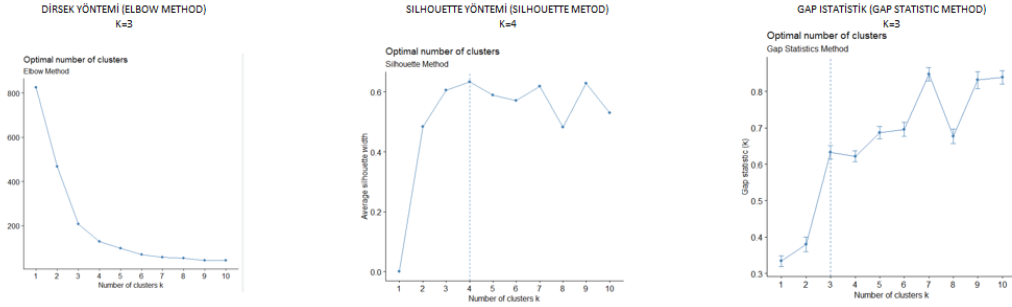
Şekil 11'de görüldüğü üzere, gürültülü verileri temizledikten sonraki durumda grafik normal dağılım grafiğine benzese de sadece grafik ile yorum yapılmamalıdır.

Anderson Darling normallik testi, gürültülü veriler temizlendikten sonra A değerinde 17,331'den 25,287'ye yükselerek iyileşme sergilemiştir. Ancak $p < 0.05$ (p -değeri = 2.2×10^{-16}) olduğu için H_0 reddedilir, veri seti normal dağılıma sahip değildir.

4.2 Makine Öğrenmesi Yöntemleri ile Sınıflandırma

Veriler temizlendikten sonra, bir sonraki adımda makine öğrenmesi yöntemlerinden, kümeleme yapılarak stratejik önemi en fazla kümenin bulunması gerekmektedir. Optimal küme sayısının bulunması için üç farklı yöntem uygulanmıştır. Bunlar; dirsek yöntemi, Sillhoutte yöntemi ve Gap istatistik yöntemidir.

Şekil 14'te yer alan grafiklere göre, veri setinin, dirsek yöntemine göre üç kümeyle, Sillhoutte yöntemine göre dört kümeyle, Gap istatistik yöntemine göre ise üç kümeyle ayrılması önerilmiştir. İdeal küme grafikleri R Studio paket programında çizilmiş olup EK-2'de kodları verilmiştir.



Şekil 11. Kümeleme uygulamaları

Belirlenen küme sayısının doğrulaması için bağlantı indeksi, Silhouette indeksi ve Dunn indeksi kullanılmıştır. Çizelge 4'de kümeleme yöntemleri, optimal küme sayısının bulunması için farklı küme sayılarına göre hesaplanan indeks değerleri verilmiştir. Çizelge 5'de görüleceği üzere hiyerarşik kümeleme yapılması durumunda bağlantı indeksi ve Dunn indeksi üç kümenin tercih edilmesi gerektiğini göstermektedir. Bu sonuç, Şekil 12'de gösterilen yöntemlerin grafik sonuçları ile eşleşmektedir.

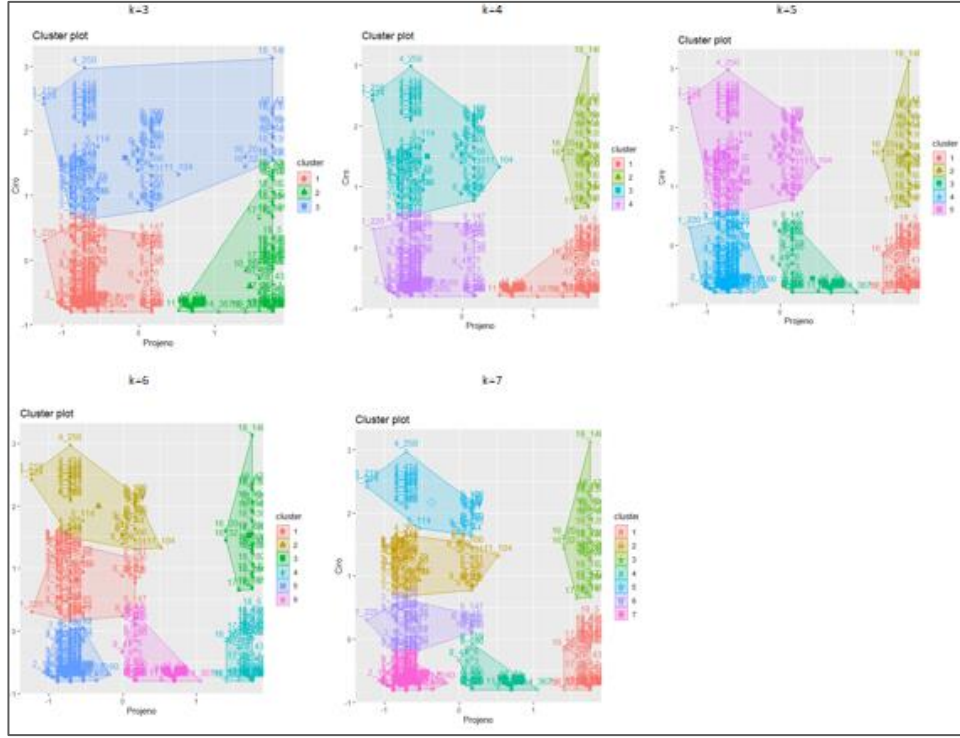
Çizelge 4. Farklı küme sayıları için validasyon değerleri

Kümeleme Yöntemi	Validasyon Ölçüm Yöntemi	3 Küme	4 Küme	5 Küme	6 Küme	7 Küme
Hiyerarşik Kümeleme	Bağlantı İndeksi	3,3778	8,3361	11,3901	12,0929	12,8119
	Dunn İndeksi	0,0773	0,0481	0,0481	0,0578	0,0578
	Silhouette İndeksi	0,4714	0,5836	0,5724	0,5449	0,6064
K-Ortalamlar ile Kümeleme	Bağlantı İndeksi	12,756	9,6663	15,9483	27,8802	34,0944
	Dunn İndeksi	0,0252	0,0391	0,0398	0,0479	0,0479
	Silhouette İndeksi	0,6063	0,6342	0,6039	0,5762	0,6215

Çizelge 5.Optimal küme sayıları

Optimal Değerler	Değer	Yöntem	İdeal Küme Sayısı
Bağlantı İndeksi	3,3778	Hiyerarşik Kümeleme	3
Dunn İndeksi	0,0773	Hiyerarşik Kümeleme	3
Silhouette İndeksi	0,6342	K-Ortalamlar ile Kümeleme	4

Şekil 13’de veri setinin k-ortalamlar yöntemiyle farklı küme sayıları (3 kümeden 7 kümeye kadar) için kümeleme analizleri yapılmış ve belirlenen kümeler Matlab programında oluşturulmuştur. Analiz sonucunda incelenen veri setinin farklı kümelere bölünebildiği, ancak Çizelge 5’de yer alan Silhouette değerine göre en uygun küme sayısı olan dört kümeyle bölünmesinin daha uygun olduğuna karar verilmiştir. İdeal küme grafikleri R Studio paket programında çizilmiş olup EK-2’de kodları verilmiştir.



Şekil 12.Farklı küme sayıları için k-ortalamalar yöntemi ile kümeler

Çizelge 6’da satırlarda kümeler, sütunlarda projeler yer almaktadır. Buradaki referanslara bakıldığında Küme 1’de toplam 91 referans, Küme 2’de toplam 25 referans, Küme 3’te 85 referans ve Küme 4’te 213 referans bulunmaktadır. Başka bir deyişle, Küme 4 en çok veriye sahip ve farklı projeleri barındıran kümedir. Projeler arasında yer alan 10, 13 ve 15 numaralı projelerde herhangi seride olan bir referans olmadığı için bu projedeki veriler gürültülü veri olarak işaretlenerek silinmiştir. Bu sebepten Çizelge 6’da bu üç proje yer almamaktadır.

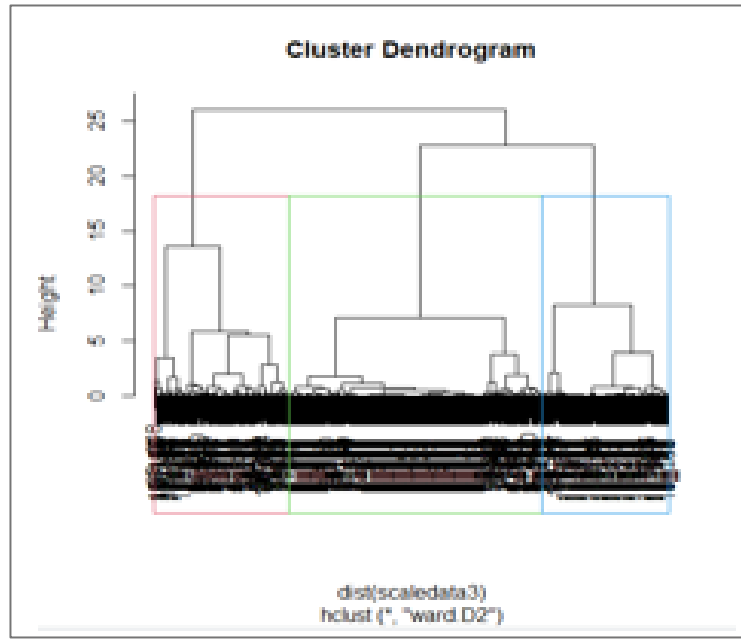
Çizelge 6. K-ortalamalar yöntemi ile belirlenen kümelerin proje bazlı ayrılması

	PR1	PR2	PR3	PR4	PR5	PR6	PR7	PR8	PR9	PR11	PR12	PR14	PR16	PR17	PR18
Küme_1	0	0	0	0	0	0	0	0	0	1	25	1	3	7	54
Küme_2	0	0	0	0	0	0	0	0	0	0	0	0	2	1	22
Küme_3	2	0	17	34	5	0	0	6	20	1	0	0	0	0	0
Küme_4	1	1	19	149	16	2	1	2	22	0	0	0	0	0	0

İncelenen veri seti, hiyerarşik kümeleme yöntemi ile de kümelendirilmiştir. Hiyerarşik kümeleme yöntemine göre Çizelge 5’te yer alan bağlantı ve Dunn indeks değerlerine göre en uygun küme sayısı üç olarak önerildiği için, veri seti üç kümeye ayrılmış, incelenen 18 projeye ait referans numaralarının kümelere göre dağılımları incelenmiştir. Hiyerarşik kümeleme yöntemine göre stratejik açıdan en önemli kümenin ikinci küme olduğu belirlenmiştir. Hiyerarşik kümeleme yöntemine göre kümelere ait dendrogram Şekil 14’de ve proje bazlı dağılım Çizelge7’de verilmiştir.

Çizelge 7. Hiyerarşik kümeleme ile belirlenen kümelerin proje bazlı ayrılması

Küme Sayısı	P1	P2	P3	P4	P5	P6	P7	P8	P9	P11	P12	P14	P16	P17	P18
Küme 1	2	0	16	33	5	0	0	6	20	1	0	0	2	1	22
Küme 2	1	1	20	150	16	2	1	2	11	0	0	0	0	0	0
Küme 3	0	0	0	0	0	0	0	0	11	1	25	1	3	7	54



Şekil 13. Hiyerarşik kümeleme dendogramı

Çalışmada k-ortalamlar yöntemi seçildiği için, en önemli küme olan dördüncü küme verileri seçilerek fiyat tahminlemesi için girdi olarak kullanılmıştır.

4.2.1 Kümeleme Çalışmaları Sonrasında Geliştirilen Satın Alma Stratejileri

Satın alma stratejilerinin oluşturulmasında makine öğrenmesinin faydalı olduğu görülmüştür. Stratejik açıdan en büyük öneme sahip dördüncü küme ayrıntılı incelendiğinde ürün grubu, ortalama kalınlık ve tüketim verileri ortaya konduğunda farklı satın alma stratejileri aşağıdaki gibi oluşturulmuştur:

Dört numaralı küme çelik malzemelerin içerisinde üç ayrı ürün grubu söz konusudur,

- HR (Hot Rolled) -> Sıcak Haddelenmiş Çelik Malzemeler
- HDG (Hot Dip Galvanized) -> Sıcak Galvanizlenmiş Çelik Malzemeler
- CR (Cold Rolled) -> Soğuk Haddelenmiş Çelik Malzemeler

Sıcak haddelenmiş çelik malzemeler sıcak haddeleme prosesinden geçmektedir. Soğuk haddelenmiş malzemeler ise, sıcak haddelendikten sonra, soğuk haddeleme prosesine girmektedir. Sıcak galvanizli malzemeler ise, haddeleme operasyonlarının haricinde galvaniz havuzlarına girerek galvanizleme prosesine girmektedir.

Üç ayrı ürün grubuna göre önerilen satın alma stratejileri aşağıda açıklanmıştır:

Strateji 1: HR ürün grubunda ortalama kalınlık projedeki çeliklerin ortalama kalınlığının yaklaşık %60 üzerindedir. Bu projede ortalama kalınlığın üzerinde kalan çelik malzemeler için gerekli teknik kontroller yapılarak kalınlık azaltma çalışması yapılabilir. Bu çalışma yapılırken, mevcut çeliklerin hangi fonksiyonda kullanılacağı ve dayanım, mukavemet gibi teknik detayları göz önünde bulundurulmalıdır.

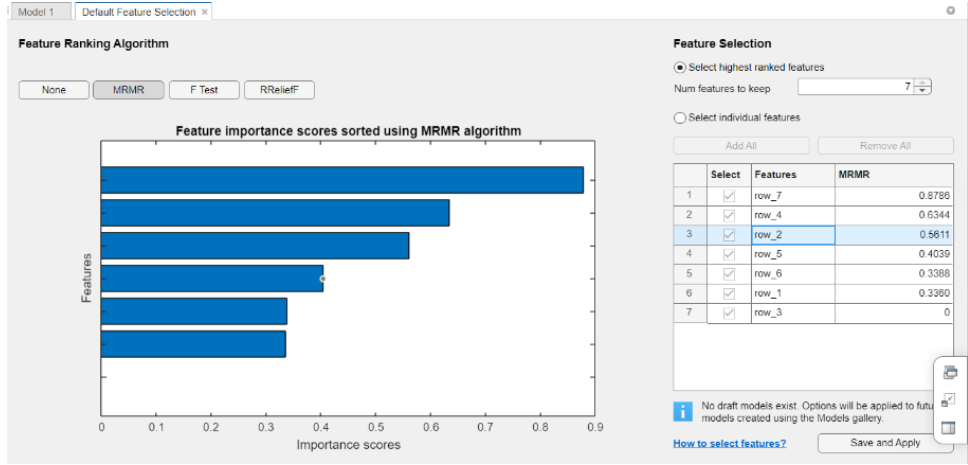
Strateji 2: HDG malzemeler tüketim tonajı en yüksek ürün grupları arasında yer almaktadır. Galvanizleme, sıcak haddeleme haricinde ekstra bir operasyon olduğu için, ekstra maliyet getirdiği bilinmektedir. Bu grupta yer alan malzemelerde galvanizleme operasyonu kaldırılarak, HR olarak beklentileri yerine getirip getirmediği araştırılmalıdır. Bunun haricinde HDG grubunda kalitelerde bir ortaklaştırma ile maliyet avantajı sağlanabilir.

Strateji 3: CR, HR'ye göre daha iyi yüzey kalitesine sahip olup hassas uygulamalar için uygundur. Ancak CR'de kesilme, kırılma veya kaynak operasyonu malzemedeki gerilimler serbest kalabilir ve malzemede kırılma veya çatlak gözlemlenebilir. HR'ye göre ekstra bir operasyon daha içerdiği için piyasadaki ekonomik dalgalanmalardan hızlı etkilenme ve fiyat artışına uğrama riski vardır. Malzemenin kullanıldığı yer, CR hassasiyeti gerektirmiyor ise CR ürün grubu tercih edilmemelidir.

4.3 Fiyat Tahmin Modelinin Oluşturulması

Fiyat tahmin modelinin oluşturulması için önce fiyatı etkileyen kriterler belirlenmiştir. Bu kriterlerin gerçekten fiyatı belirleyip belirlemediğinin ortaya konması için öznitelik seçimi yapılmıştır. Sonrasında YSA Tabanlı ve ANFIS tabanlı iki ayrı tahmin modeli oluşturulmuştur. Bu iki tahmin modeli birbiriyle kıyaslanarak sonuçlar ortaya konmuştur.

Fiyatı belirleyen kriterleri YSA ve ANFIS girdisi olarak kullanılmadan önce ilk adımda gerçekten fiyata etki edip etmediği anlaşılmalıdır. Fiyata etki etmeyen bir öznitelik varsa, bu özneliğin YSA veya ANFIS sistemlerine girdi olarak verilmemesi gerekmektedir. Burada MRMR algoritmasının seçilme sebebi, KBÖS algoritmasının tüm fiyat belirleyici kriterleri kontrol etmeme riskinin olması ve korelasyon katsayısının ise tüketim gibi değişken kriterler üzerinde sıfıra yakın sonuç vererek tüketim kriteri fiyat üzerinde etkisiz gibi göstererek yanıltmasıdır. Şekil 15'de Matlab'de MRMR algoritmasının sonuçları belirtilmiştir.



Şekil 14. MRMR algoritması sonuçları

Çıktıları ve özniteliklerin anlamının daha iyi anlaşılabilmesi için Çizelge 8’de değerler bir araya getirilmiştir. Değerler incelendiğinde Sakar vd. (2012) çalışmasında belirttiği üzere 0,20 üzerindeki MRMR değerlerinin özniteliklere dahil edilmiştir. 0,20’nin altında kalan genişlik özniteliği ise fiyata etki etmemektedir. Dolayısıyla bu öznitelik çalışmada dikkate alınmamıştır.

Çizelge 8. MRMR algoritma değerleri

Öznitelik	Öznitelik Numarası	MRMR Değeri
Kesim Tipi	Row 7	0,8786
Boy	Row 4	0,6344
Kalınlık	Row 2	0,5611
Tüketim	Row 5	0,4039
Ürün Grubu	Row 6	0,3388
Kalite	Row 1	0,3360
Genişlik	Row 3	0,0000

4.3.1 Yapay Sinir Ağları ile Fiyat Tahmin Modeli

YSA ağ tasarımı için altı girdi ve bir çıktıdan oluşan, girdileri kesim tipi, boy, kalınlık, tüketim, ürün grubu, kalite ve genişlik olan, çıktısı ise fiyat olan bir ağ tasarımı yapılmıştır.

Ağ tasarımının yapılmasında eğitim algoritması, eğitim setinin veri bölünme oranları, gizli katman sayıları, gizli katmanlardaki nöron sayısı gibi farklı parametreler eğitim performanslarını etkilemektedir. Bu kriterler üzerinde farklı denemeler yapılarak en iyi ağ tasarımı belirlenmeye çalışılmıştır.

YSA ile ağ tasarımı parametrelerinde eğitim algoritması olarak Levenberg-Marquardt algoritması kullanılmıştır. Eğitim setinin bölünme oranları oranı %70'i eğitim, %15'i validasyon ve %15'i test ve %80 eğitim, %10 validasyon ve %10 Test olarak iki ayrı oranda çalışılmıştır. Gizli katman sayısı bir ve iki gizli katman olarak seçilmiştir. Nöron sayıları ise dört nörondan on iki nörona kadar denenmiştir.

Ağ tasarım parametrelerinde, eğitim setinin veri bölünme oranı için Çizelge 6'dan gelen küme 4'e ait toplamda 213 adet verinin %70'i eğitim, %15'i validasyon ve %15'i test için kullanılmıştır. Bu durumda 149 adet veri eğitim için, 32 adet veri validasyon için, 32 adet veri de test için kullanılmıştır. Eğitim setinin veri bölünme oranı %70'i eğitim, %15'i validasyon ve %15'i test olduğu durum VBO1 olarak anılacaktır.

İkinci veri bölünme oranında ise küme 4'e ait 213 adet verinin %80 eğitim, %10 validasyon ve %10 test için kullanılmıştır. Bu durumda 171 adet veri eğitim, 21 adet veri validasyon ve kalan 21 adet veride test için kullanılmıştır. Burada amaç, farklı karar verme kriterlerinin etkilerini görmektir. Eğitim setinin veri bölünme oranı %80 eğitim, %10 validasyon ve %10 test olduğu durum VBO2 olarak anılacaktır.

Ayrıca, yapay sinir ağında tek katman her zaman iyi sonuçlar vermeme olasılığına karşı denemeler hem tek katman hem çift katmanda yapılmıştır ve R değerleri incelenmiştir. R değeri bir değişken veya değişkenler ile açıklanabilir bağımlı değişken için istatistiksel açıdan fark veya varyans oranını temsil eden bir istatistiksel ölçüdür. Tek gizli katman için, gizli katmandaki nöron sayısı 4'ten 12'ye kadar değiştirilmiş, her nöron seviyesi için en iyi beş adet R değeri seçilerek Çizelge 9'da gösterilmiştir.

Çizelge 9. VBO1 ile en iyi R değerine sahip 5 Deneme

Nöron Sayısı	Deneme 1	Deneme 2	Deneme 3	Deneme 4	Deneme 5
4	0,95	0,95	0,95	0,95	0,91
5	0,89	0,90	0,91	0,92	0,93
6	0,94	0,90	0,95	0,96	0,95
7	0,95	0,91	0,95	0,91	0,90
8	0,96	0,96	0,92	0,95	0,95
9	0,93	0,96	0,96	0,93	0,92
10	0,89	0,94	0,91	0,95	0,96
11	0,95	0,97	0,89	0,95	0,92
12	0,92	0,93	0,91	0,97	0,94

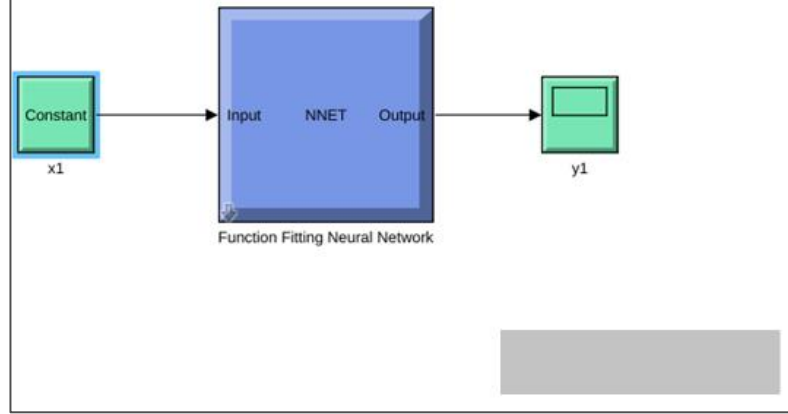
Çizelge 10. VBO2 ile en iyi R değerine sahip 5 Deneme

Nöron Sayısı	Deneme 1	Deneme 2	Deneme 3	Deneme 4	Deneme 5
4	0,90	0,94	0,94	0,95	0,96
5	0,94	0,93	0,92	0,95	0,96
6	0,95	0,96	0,96	0,96	0,94
7	0,95	0,92	0,94	0,95	0,94
8	0,96	0,94	0,95	0,94	0,90
9	0,94	0,92	0,95	0,93	0,96
10	0,93	0,94	0,94	0,94	0,94
11	0,94	0,93	0,90	0,92	0,91
12	0,96	0,94	0,96	0,96	0,94

Çizelge 9 ve Çizelge 10'da %70 Eğitim ve %80 Eğitim verileri için en iyi beş denemeye ait tek gizli katmana sahip ağ yapısı için sonuçlar sıralanmıştır. Bu çalışmada R değerlerine eğitim, validasyon ve test olarak ayrı ayrı bakıldığı için bu üç değerlerin ortalaması alınarak ortalama R değerine göre yorum yapılmıştır.

En iyi denemelere ait Simulink modelleri oluşturulmuş ve fiyat tahminleri Simulink modelleri üzerinde de çalıştırılmıştır. Simulink modeline, girdi olarak kesim tipi, boy, kalınlık, tüketim, ürün grubu ve kalite değerleri girilmektedir. Model çalıştırıldığında tahminleme gerçekleşmektedir. Bu sayede, fiyat tahmini için fabrikalarda kolayca tahmin yapan bir model oluşturulmuştur.

Şekil 16’da görülen Simulink model örneğinde temsili bir model verilmiştir. Sağ altta yer alan gri kutucukta model çalıştırıldığında, YSA tahminlediği fiyatlar verilmektedir.



Şekil 15. Simulink model örneği

İlk adımda Çizelge 11’de üç adet veri setinin hiç görmediği, kümeleme çalışmasında farklı kümelerde kalan ve YSA veri setine girmeyen verilere ait üç tane rastgele veri seçilmiştir ve bu veri setleri R değerine göre iyi performans gösterildiği düşünülen modellerde denenmiştir.

Burada ortalama mutlak sapma (MAE) dikkate alınmıştır. MAE’nin formülü denklem 4.2’de verilmiştir. Burada n değer sayısı, y_i tahminlenen fiyat, x_i ise ürünün gerçek fiyatıdır.

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (4.2)$$

Gerçek fiyat ve tahminlenen fiyat arasındaki mutlak sapmalar (MAE) ve üç test için ortalama mutlak sapmalar yüzdesi (OMS) Çizelge 11’de verilmiştir.

Sapma Değerleri denklem 4.3’deki formül ile hesaplanmaktadır.

$$\text{Teste Ait Sapma Değeri} = \left| \frac{(\text{Tahminlenen Fiyat} - \text{Gerçek Fiyat})}{\text{Gerçek Fiyat}} \right| \times 100 \quad (4.3)$$

Çizelge 11’de sapmalar negatif ya da pozitif olabildiğinden birbirini götüreceği için mutlak değer olarak hesaplanmıştır. Çizelge 11 tek gizli katman için oluşturulmuştur.

Çizelge 11. VBO1 ve VBO2 için üç test sonucunda değerler

VBO	Nöron Sayısı	Deneme No	R Değeri	Test 1 Sapma	Test 2 Sapma	Test 3 Sapma	OMS	MAE
VBO1	6	2	0,96	0,51	0,01	0,68	0,4	0,51
	12	1	0,96	0,63	0,28	0,28	0,4	0,52
	7	4	0,95	0,33	0,71	0,09	0,38	0,55
	10	4	0,94	0,81	0,91	0,31	0,68	0,85
	8	3	0,95	0,39	1,61	0,28	0,76	0,94
	4	2	0,94	0,47	0,02	1,89	0,79	1,14
	8	1	0,96	0,09	2,45	0,69	1,08	1,2
	5	5	0,96	1,70	0,16	1,50	1,12	1,41
	9	3	0,95	1,29	1,3	0,76	1,14	1,44
	5	4	0,95	0,51	2,88	0,33	1,24	1,51
	4	5	0,96	0,39	0,50	0,83	0,58	1,72
	9	5	0,96	0,45	3,64	1,58	1,89	2,31
VBO2	6	1	0,95	0,75	8,90	0,95	3,53	4,08
	9	2	0,96	0,51	0,06	0,17	0,25	0,31
	6	3	0,95	0,57	0,34	0,02	0,31	0,39
	9	3	0,96	0,15	0,51	0,63	0,43	0,53
	8	2	0,96	0,21	0,34	0,27	0,28	0,62
	6	4	0,96	0,39	0,52	0,43	0,45	0,67
	11	2	0,97	0,45	0,02	0,73	0,4	0,7
	12	4	0,97	0,24	3,39	0,07	1,23	1,58
	8	1	0,96	0,21	4,95	0,03	1,73	2,05
	4	1	0,95	1,76	2,78	3,58	2,71	3,31

Çizelge 12’de tek katmanda en iyi R değerleri, Çizelge 13’te ise çift katmandaki en iyi R değerleri hesaplanmış olup üç teste ait OMS ve MAE değerleri bulunmuştur.

Çizelge 12’de yer alan tanım, ağa verilen isimdir. Karar kriteri, veri setinin % kaçını eğitime ayırdığını göstermektedir (VBO1 veya VBO2). Katman sayısı kaç gizli katman olduğunu göstermektedir. Nöron sayısı bu katmanda kaç nöron bulunduğunu belirtmektedir. Deneme numarası yapılan beş adet denemenin kaçınıcı denemesinde bu değer tespit edildiğini belirtilmektedir. R değeri, yapılan üç test için mutlak sapma değerleri, OMS ve MAE değerleri eklenmiştir.

Çizelge 12. Tek gizli katman için en iyi R değerine sahip ağlar

Tanım	VBO	Katman Sayısı	Nöron Sayısı	Deneme No	R	Test 1 Sapma	Test 2 Sapma	Test 3 Sapma	OMS	MAE
AI_1	VBO2	1	11	2	0,97	0,45	0,02	0,73	0,40	0,70
AI_2	VBO2	1	9	2	0,96	0,51	0,06	0,17	0,25	0,31
AI_3	VBO2	1	6	4	0,96	0,39	0,52	0,43	0,45	0,67
AI_4	VBO1	1	10	4	0,94	0,81	0,91	0,31	0,68	0,85
AI_5	VBO1	1	6	2	0,96	0,51	0,01	0,68	0,40	0,51
AI_6	VBO1	1	5	5	0,96	1,70	0,16	1,50	1,12	1,41

Çizelge 13'te ilen uygun ağ yapısını belirlerken farklı versiyonlar denenerek en iyi çıkan sonuçlar dikkate alınmıştır. Toplam nöron sayısı ile iki katmandaki toplam nöronların sayısı verilmiştir.

Çizelge 13. Çift gizli katman için en iyi R değerine sahip ağlar

Tanım	VBO	Katman Sayısı	Toplam Nöron Sayısı	Deneme No	R	Test 1 Sapma	Test 2 Sapma	Test 3 Sapma	OMS	MAE
AI_7	VBO2	2	11	2	0,96	0,75	0,42	2,85	1,34	2,02
AI_8	VBO2	2	9	2	0,97	0,05	0,28	2,16	0,83	0,65
AI_9	VBO2	2	6	4	0,97	0,59	0,21	2,34	1,05	1,24
AI_10	VBO1	2	10	4	0,93	0,31	1,26	1,83	1,14	1,49
AI_11	VBO1	2	6	2	0,95	0,02	0,37	1,64	0,68	0,46
AI_12	VBO1	2	5	5	0,97	0,63	0,53	1,64	0,93	1,16

Bu çizelgeler incelendiğinde YSA ile yapılan tahminlerde MAE AI_2, AI_5, AI_8 ve AI_11 numaralı ağlarda en küçük bulunmuştur. Detaylı değerleri Çizelge 14'te gösterilmiştir. Dolayısıyla final adımda YSA'nın hiç görmediği 28 adet veriyi tahminleyeceği yapılar bu sayede toplamda 22 farklı ağdan 4 adet ağa düşürülmüştür ve hesaplama kolaylığı sağlanmıştır.

Final adımda, veri setinin hiç görmediği 28 adet veri, YSA'ya verilmiştir. Dört ayrı ağın gerçek fiyatlar ve yapay sinir ağının tahminlediği fiyatlar arasındaki OMS ve MAE değerleri Çizelge 14'te hesaplanmıştır.

Korelasyon katsayısı 1'e ne kadar yakınsa iki veri arasında o kadar kuvvetli bir ilişki vardır denir. Bu sebepten 1'e en yakın değer 0,8255 değeri ile AI_11 ile isimlendirilen VBO1 için iki gizli katmanında toplamda 6 nöronlu çalışan YSA'ya aittir. Çizelge 14'te korelasyon değerleri de verilmiştir.

Çizelge 14. 28 Adet verinin Tahmin Sonuçları

Tanım	VBO	Katman Sayısı	Toplam Nöron Sayısı	Deneme No	R	OMS	MAE	Korelasyon
AI_11	VBO1	2	6	2	0,95	1,0	29,49	0,8255
AI_8	VBO2	2	9	2	0,97	1,0	30,27	0,8245
AI_2	VBO2	1	9	2	0,96	1,2	32,43	0,8148
AI_5	VBO1	1	6	2	0,96	1,6	38,42	0,7935

Yapay sinir ağları ile tecrübesiz satın alma çalışanları fiyat tahminleyebilir, çünkü herhangi bir tecrübe istemeden veri setleri üzerinden ilerleyebilmektedir. Ancak bu çalışmada tecrübeli satın alma çalışanlarının tecrübesini de yapay sinir ağlarının üzerine koymak hedeflenmektedir. Bu sebepten ANFIS yöntemi tecrübeli satın alma çalışanları için oluşturularak bir sonraki bölümde anlatılmıştır.

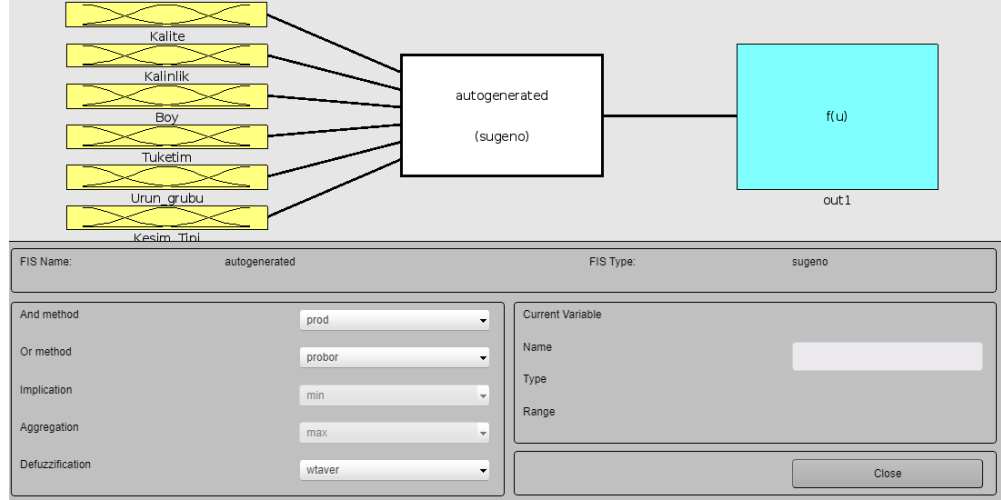
Tahminlemeyi Matlab ortamında yapan YSA kodu EK 3'te dört nöronu, bir gizli, katmanda, Levenberg Marquardt eğitim algoritması ile VBO2 için verilmiştir.

4.3.2. Adaptif Ağ Yapısına Dayalı Bulanık Çıkarım Sistemleri ile Fiyatların Tahminlenmesi

Çalışmada Adaptif ağ yapısına dayalı bulanık çıkarım sistemi (ANFIS) için Matlab 2022b paket programında "Neuro-Fuzzy Designer" uygulaması kullanılmıştır.

Bu ekran ile "Load data" kısmından Matlab'ın workspace alanından veya doğrudan dosyalarından eğitim, test, kontrol ve deneme(demo) olmak üzere data setleri yüklenmiştir. Toplamda 213 adet olan verinin 179 adeti eğitim, 34 adeti test verisi olarak bu sistemden yüklenmiştir.

Generate FIS kısmı, bulanık ağ tabanlı çıkarımların (FIS) yapıldığı kısımdır. Sisteme değişkenlerin ne olduğu, aralıkları ve tipleri verilir. Şekil 17’de fiyatı oluşturan öznitelikler bulanık ağ tabanlı çıkarım sistemine aktarılmıştır. Burada üçgen üyelik fonksiyonu tercih edilmiştir.



Şekil 17. FIS ekran örneği

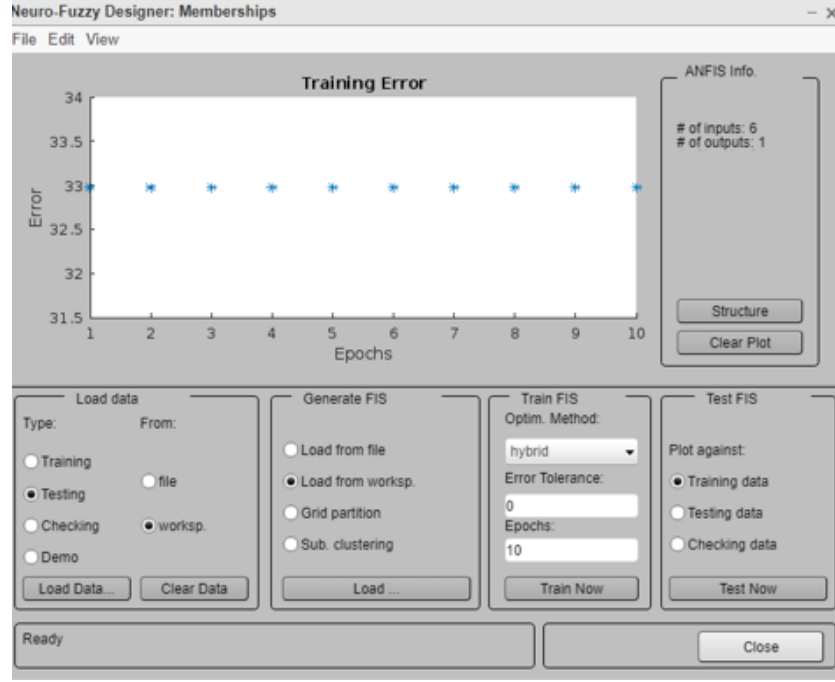
FIS’e girilecek veriler girişi esnasında üyelik fonksiyonlarının girilmesi gerekmektedir. Çizelge 15’de hangi üyelik fonksiyonunun nasıl adlandırıldığı gösterilmiştir. Söz konusu üyelik fonksiyonları sisteme tanımlanmıştır.

Bir sonraki adımda ANFIS’in takip edeceği kurallar satın alma alanındaki tecrübeye dayanarak hazırlanmıştır. Burada gerçek hayat koşullarında tecrübeye dayanarak, özniteliklerin değişmesinin fiyat üzerindeki etkilerine dair kurallar oluşturulmuş ve ANFIS’e eklenmiştir. EK-1’de kurallar Çizelgesi verilmiştir. Parametrelerin kriterleri fiyat dikkate alınarak belirlenmiştir.

Çizelge 15. Üyelik fonksiyonları

Parametre	Detay	Kriter
Kalite	1-9 numaralı kaliteler	Ortalama
	10-18 numaralı kaliteler	Yüksek
	19-25 numaralı kaliteler	Düşük
Kalınlık	0,5 - 1,15 mm arası	Yüksek
	1,17 - 2 mm arası	Orta
	2,1 - 8 mm arası	Düşük
Boy	540-1000 mm arası	Düşük
	1030-1350 mm arası	Orta
	1474-2320 mm arası	Yüksek
Tüketim	1,4-87,2 ton arası	Düşük
	97,4 - 336,6 ton arası	Yüksek
	337-2584 ton arası	Orta
Ürün Grubu	1 (Sıcak Haddeli)	Düşük
	2 (Sıcak Galvanizli)	Orta
	3 (Soğuk Haddeli)	Yüksek
Kesim Tipi	5 nolu tip	Düşük
	1 ve 2 nolu tip	Orta
	3 ve 4 nolu tip	Yüksek

Kurallar, üyelik fonksiyonları ve bulanık ağ tabanlı çıkarımlar (FIS) girildikten sonra sistem çalıştırılmıştır. Sistem çalıştığında hata değerinin (MSE) sıfıra yakın küçük bir sayı olması beklenmektedir. Çalışmada 32,94 bulunmuştur. MAE değeri ise 29,12 olarak bulunmuştur. Bu değerler düşük olduğu için kabul edilebilir düzeydedir. Şekil 17’de ANFIS hata değerleri verilmiştir.



Şekil 16.ANFIS hata değeri

Bu çalışma sonucunda elde edilen model ile, ANFIS modeline hiç tanıtılmamış 28 adet malzemenin fiyatı YSA’da olduğu gibi tahminlenmiştir. Çizelge 16’da 28 adet verinin ANFIS test sonuçları verilmiştir.

Çizelge 16. 28 adet verinin ANFIS ile test sonuçları

Tanım	MAE	OMS	Korelasyon
ANFIS	29,12	1,05	0,8257

Kullanılan ANFIS ve YSA’nın korelasyonları dikkate alındığında ANFIS’in korelasyon değeri 0,8257, AI_11’in korelasyon değeri 0,8255, AI_8’in korelasyon değeri 0,8245, AI_2’nin korelasyon değeri 0,8148 ve AI_5’in korelasyon değeri 0,7935 olarak bulunmuştur. Bu çıktılara göre gerçek veriler ile arasında en büyük korelasyon bulunan yöntem ANFIS, korelasyon değeri ise 0,8257’dir.

4.4 YSA ve ANFIS ile Satın Alma Stratejilerinin Oluşturulması

Elde edilen YSA ve ANFIS tahminleri ile karlı satın alma stratejileri oluşturmak hedeflenmiştir. Söz konusu 28 malzeme için hem tedarikçilerden teklif alınmıştır hem de YSA ve ANFIS ile tahminlemeler gerçekleştirip mevcut durum analiz edilmiştir.

Strateji 1: Sac malzemelerin kalitelerine ait ortalama tahminlenen fiyatlar ve tedarikçiden gelen fiyatlar karşılaştırıldığında ANFIS'in 12. Kaliteye tabi olan malzeme grubunda tedarikçiden gelen teklife göre %6 daha düşük fiyat tahmini yaptığı görülmüştür. Tedarikçi ile pazarlık yapılması sonucunda, tedarikçiden indirim alınarak ANFIS'in tahminlediği fiyata getirilmiştir.

Strateji 2: Daha önce teknik detayları çalışılmamış ve tanınmayan bir malzeme için ANFIS ve YSA'ya malzeme fiyatı tahminletilmiştir. Tedarikçinin ise bu teklifi vermek için çalışması gereken süre minimum iki haftadır. İki hafta sonunda tedarikçiden gelen teklif ve tahminler karşılaştırıldığında iki makine öğrenmesi yönteminde de %1'in altında sapma gözlenmiştir. Dolayısıyla, tanınmayan malzemelerin fiyatları için makine öğrenmesi yöntemleri fikir verici olarak değerlendirilmelidir. Bu sistemler insanlardan daha hızlı hareket edebildiği unutulmamalıdır, zira iki haftada gelen teklif YSA veya ANFIS kullanarak saniyeler içinde hesaplanabilir.

Strateji 3: Malzeme fiyatını düşürmek adına yapılan çalışmalarda, fiyatı en çok etkileyen faktörü bulmak için tüm faktörler teker teker denenmektedir. Ancak MRMR Algoritması ile belirlenen kesim tipi faktörü fiyat üzerinde en yüksek etkiye sahip olduğundan bu faktör üzerinde yoğunlaşılarak maliyetler düşürülebilir.

Strateji 4: Tecrübeli satın alma çalışanlarının kullandığı ANFIS yönteminde oluşan kurallar ile, yıllardır süregelen tecrübeler, nasihat olmak yerine bilgisayarlarda kullanılan aktif bir sisteme dönüşecektir. Bu sayede tecrübe bilgisayarda depolanabilir konuma gelmiş olacaktır.

5. SONUÇ

Bu tez çalışmasında, bir satın alma departmanı çalışanın kritik malzemelerden olan sac malzemeler için doğru kararları kısa sürede, doğru ve efektif şekilde verebilmesi amaçlanmıştır.

İlk adımda, fiyatlar ve malzeme detaylarını içeren veri setinde, gerçek hayatta yaşanan problemlerden yola çıkarak gürültülü verilerin neler olabileceği tanımlanmıştır. R studio paket programında bu veriler temizlenmiştir. Detaylı ve doğru analiz yapabilmenin ilk kuralının temiz veri setleri ile çalışmaktır. Aksi halde hatalı sonuçlar çıkabilmektedir. Örneğin, çalışmanın başında seri üretimi devam etmeyen referanslar, gürültülü veri olarak tanımlanmıştır. Çünkü, bu malzemeler günümüzde kullanılmamaktadır. Dolayısıyla, fiyat analizlerinin içerisine girmemelidir. İkinci adımda kümeleme analizi yaparken ise k-ortalamar yönteminde onuncu, on üçüncü ve on beşinci projelere ait kümelerin boş olduğu görülmüştür. Bu bir hata gibi görünebilir ancak bu projeler tekrar kontrol edildiğinde, yaklaşık otuz sene önce sonlanmış projelerdir. Bu projelerin referansları seri üretimde kullanılmadığı için veri madenciliği adımıyla temizlenerek gelecek süreçlerde kullanılmamışlardır. Ancak veri madenciliği teknikleri uygulanmamış olsaydı, hiç kullanılmayan referansların fiyatları kümeleme analizine dahil edilmiş ve değerlendirmeye girmiş olacaktı. Bu örnekten yola çıkarak, veri madenciliği tekniklerinin, satın alma süreçlerinde yanlış kararların alınmaması için kullanılması tavsiye edilmektedir.

İkinci adımda makine öğrenmesi tekniklerinden biri olan kümeleme analizi kullanılmıştır. Burada veri setinin nasıl dağıldığını tespit etmenin önemi vurgulanmıştır. Çünkü normal dağılan bir veri seti için regresyon gibi yöntemler uygulanabileceği gibi, normal dağılmayan veri setleri için ya gerekli dönüşüm teknikleri uygulandıktan sonra regresyon yapılmalıdır veya farklı yöntemler seçilmelidir. Burada amaç cirosu yüksek referansları bir kümede toplamak olduğu için kümeleme analizinden faydalanmıştır. K-ortalamar yöntemi ile kümeleme analizi yapılmış olup geçerliliği ve doğruluğu bağlantı indeksi, Dunn indeksi ve Silhouette indeksi ile yapılmıştır. K-ortalamar yöntemi için bu

Confidential C

çalışmada ideal küme sayısı dördür. Ancak k-ortalamalar yönteminin yanında, hiyerarşik kümeleme de yapılabileceğini göstermek için ikinci kümeleme yöntemi olarak hiyerarşik kümeleme seçilmiştir. Hiyerarşik kümeleme için üç kümenin ideal olduğu tespit edilmiş olup k-ortalamalar yöntemi için yapılan geçerlilik ve doğruluğu ölçen indeksler ile paralel olduğu ortaya konmuştur. Dolayısıyla iki yöntemde birbirini doğrulamaktadır.

Üçüncü adımda k-ortalamalar yöntemiyle seçilen kümedeki veriler YSA ile fiyat tahminlemesine girdi olarak kullanılmıştır. Burada Levenberg-Marquardt algoritması ile veriler eğitilmiş olup hem tek katmanda hem çift katmanda farklı nöron değerleri için testler yapılmıştır. En iyi performansa sahip ağ yapısını bulana kadar farklı denemeler yapılmıştır ve dört tane ağ kalana kadar test edilmiştir. En sona kalan bu dört YSA için veri setinin hiç görmediği 28 adet referans özelliği için fiyat tahminlemesi yapılmıştır. YSA ile gerçek hayat koşullarında satın alma çalışanın tecrübesine dayanmadan, veri setleri ile eğitim gerçekleştirerek fiyat tahminlenebilmektedir. Bu durum, tecrübesi az olan satın alma çalışanın çalıştığı satın alma departmanları için avantaj sağlamaktadır.

YSA sistemine alternatif olacak ikinci bir sistem araştırılmış olup, araştırmalar sonucunda adaptif ağ yapısına dayalı bulanık çıkarım sistemleri (ANFIS) uygun bulunmuştur. ANFIS'in seçilme sebeplerinden biri de YSA'ya, satın alma çalışanlarının tecrübelerine dayanan kurallar eklendiğinde tecrübeyi kullanabilir hale getirmesidir. Örneğin satın alma departmanları için fiyatın yüksek çıkması için hangi faktörün değişmesi gerektiğini bilmek tecrübe gerektirmektedir. Tecrübeli satın alma çalışanları ANFIS'e bu tecrübelerini liste halinde aktarmıştır.

Son adımda, ağ yapısına dayalı bulanık çıkarım sistemler için veri setinin daha önce hiç görmediği ve YSA'nın test edildiği 28 adet veri ile test edilmiştir.

Korelasyon değeri bir sayısına ne kadar yakınsa, iki veri arasında o kadar güçlü bir ilişki vardır denir. Çizelgedeki en yüksek korelasyon değeri ANFIS'e aittir. Ancak ANFIS'ten gelen korelasyon değeri 0,8257 ile AI_11 isimli YSA 0,8255 değerleri birbirine çok yakın

bulunmuştur. AI_11 isimli YSA'ya yakından bakıldığında, VBO1 ile oluşturulmuş, toplamda 6 adet nöron içeren ve R değeri 0,95 olan iki gizli katmanlı yapay sinir ağıdır.

YSA ve ANFIS karşılaştırıldığında, bu çalışmada ANFIS tahminlerinin, gerçek fiyat verileri ile daha yüksek korelasyona sahip olduğu gözlemlenmiştir. Ancak ANFIS ve YSA arasındaki korelasyon farkının 0,0002 olduğu göz önünde bulundurulmalıdır. Küçük bir fark ile ANFIS'in YSA'ya göre daha iyi sonuç verdiği ortaya konmuştur.

Önemli olan noktalardan biri de satın alma sistemleri için ANFIS tabanlı fiyat tahmin modelleri oluşturulurken, kurallar kısmının manuel hataya açık bir noktası olmasıdır. Burada yanlış girilebilecek bir kural, sonuçların korelasyonunu düşürebilir. Bu sebeple dikkat edilmesi gereken bir husustur.

Sonuç olarak bu çalışmalar ışığında YSA ve ANFIS modellerinin doğru şartlarda satın alma fonksiyonlarında kullanılabileceği ortaya konmuştur ve birtakım stratejiler oluşturulmuştur. Bu çalışmalar ile tahminlenen fiyatlar, tedarikçilerin fiyatlarıyla karşılaştırma veya tanınmayan ürünler için tahmini fiyat gereken noktalarda tedarikçinin teklif hazırlamasının bekleyerek zaman kaybedilmemesi gibi hususlarda satın alma çalışanlarına destek sağlamaktadır. Bu çalışma sonucunda, satın alma çalışanlarına en az manuel hata ile en doğru fiyatları tahminlemesi sağlanmıştır.

Tecrübe satın alma departmanlarında çok önemli bir kriterdir. YSA sistemleri, tecrübesi az çalışanların bulunduğu satın alma departmanları için, ANFIS sistemleri ise tecrübesi fazla satın alma çalışanların bulunduğu satın alma departmanları için önerilmektedir.

Bu çalışma ile tecrübenin makine öğrenmesi teknikleriyle öğrenilebildiği ve bilgisayarda kullanılabilir hale geldiği gösterilmiştir. Aynı zamanda mesleğe yeni giriş yapan ve tecrübesiz satın alma çalışanlarının hata yapmasını önlemekte ve doğru kararları vermesi için yol açmaktadır.

KAYNAKLAR

- Aksoy, A. (2012). Global Dış Satın Alma İçin Akıllı Karar Destek Sistemi Tasarımı .*Bursa Uludağ Üniversitesi/ Fen Bilimleri Enstitüsü, Doktora Tezi* <https://www.proquest.com/docview/2593562010/abstract/1EC38F4F4037449CPQ/1>
- Al-Omary, A. Y., & Jamil, M. S. (2005). A new approach of clustering based machine-learning algorithm. *Knowledge Based Systems, 19*, 248-258. <https://doi.org/10.1016/j.knosys.2005.10.011>
- Arslankaya, S., & Toprak, Ş. (2021). Using Machine Learning and Deep Learning Algorithms for Stock Price Prediction. *Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi, 13*(1), 178-192. <https://doi.org/10.29137/umagd.771671>
- Asilkan, Ö. (2008). Veri Madenciliği Kullanılarak İkinci El Otomobil Pazarında Fiyat Tahmini. *Akdeniz Üniversitesi/Sosyal Bilimler Enstitüsü, Doktora Tezi*. <http://acikerisim.akdeniz.edu.tr/xmlui/bitstream/handle/123456789/5004/T02237.pdf?sequence=1&isAllowed=y>
- Ataseven, B. (2013). Yapay Sinir Ağları İle Öngörü Modellemesi. *Öneri Dergisi Makale Basvuru, 39*(10), 101-115.
- Babaoğlu, A. (2015). Veri Madenciliği Yöntemleri ve Bir Uygulama. *Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi*.
- Beale, M. H., Hagan, M., & Demuth, H. (2018). Neural Network Toolbox User Guide. *The MathWorks Inc*.
- Bejaei, M., Cliff, M. A., & Singh, A. (2020). Multiple Correspondence and Hierarchical Cluster Analyses for the Profiling of Fresh Apple Customers Using Data from Two Marketplaces. *Foods, 9*(7), Art. 7. <https://doi.org/10.3390/foods9070873>
- Çelik, C., & Bilge, H. Ş. (2015). Ağırlıklandırılmış Koşullu Karşılıklı Bilgi İle Öznitelik Seçimi. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 30*(4). <https://doi.org/10.17341/gummfd.81654>
- D'Agostino, R. & Stephens, M. (1986). Goodness-of-fit-techniques (Statistics: a Series of Textbooks and Monographs), Vol. 68, 1st Edition. Dekker.
- Erbudak, A. E. (2022). Veri Madenciliği Ve Makine Öğrenimi İle Döviz Kuru Tahmini Uygulaması. *Altınbaş Üniversitesi/Lisansüstü Eğitim Enstitüsü, Yüksek Lisans Tezi*(97)
- Erdoğan, E., & Özyürek, H. (2012). Yapay Sinir Ağları İle Fiyat Tahminlemesi. *Sosyal ve Beşeri Bilimler Dergisi, 4*(1), 85-92.
- Erpolat, S. (2012). Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması. *Anadolu Üniversitesi Sosyal Bilimler Dergisi, 12*(1),(151-166)). <http://earsiv.anadolu.edu.tr/xmlui/bitstream/handle/11421/163/846470.pdf?sequence=1&isAllowed=y>
- Ertaylan, A., Aktaş, Ö., & Doğan, Y. (2021). Yapay Sinir Ağları ile Piyasa Takas Fiyatı Tahminlemesi. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi 23*(67), 93-105
- Ganesh, M., & Venkatasubbu, P. (2019). Used Cars Price Prediction using Supervised Learning Techniques. *International Journal of Engineering and Advanced Technology, 9*, 216-223. <https://doi.org/10.35940/ijeat.A1042.1291S319>

- Giri, C., Jain, S., Zeng, X., Bruniaux, P. (2019). A Detailed Review of Artificial Intelligence Applied in the Fashion and Apparel Industry. *IEEE Access*, 7, 95376-95396. <https://doi.org/10.1109/ACCESS.2019.2928979>
- Gülgezen, G., Cataltepe, Z., Yu, L. (2022). MRMR Algoritması Kullanılarak Kararlı Öznitelik Seçimi (5). https://web.itu.edu.tr/~cataltepe/pdf/2009_SIU_Gulgezen.pdf
- Henderson, A. R. (2006). Testing experimental data for univariate normality. *Clinica Chimica Acta*, 366(1-2), 112-129. <https://doi.org/10.1016/j.cca.2005.11.007>
- Hepsen, A. (2012). The Importance of Real Estate Price Indices in Financial Crisis and the Approaches for Constructing Them. *Finansal Mühendislik ve Risk Yönetimi Özel Sayısı*, 19.
- Hosoe, M., Kuwano, M., & Moriyama, T. (2021). A method for extracting travel patterns using data polishing. *Journal of Big Data*, 8(1), 13. <https://doi.org/10.1186/s40537-020-00402-w>
- Jang, J.-S. R. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 665-685. <https://doi.org/10.1109/21.256541>
- Jeong, J., Park, S., & Lee, C. (2016). Comprehensive comparison of normality tests: Empirical study using many different types of data. *Journal of the Korean Data and Information Science Society*, 27(5), 1399-1412. <https://doi.org/10.7465/jkdi.2016.27.5.1399>
- Jilletefth, P., Sequeira, M., & Adlemo, A. (2019). Three novel fuzzy logic concepts applied to reshoring decision-making. *Expert Systems with Applications*, 126, 133-143. <https://doi.org/10.1016/j.eswa.2019.02.018>
- Kameshwaran K., Malarvizhi K. (2014). Survey on Clustering Techniques in Data Mining. *International Journal of Computer Science and Information Technologies*, 5 (2) , 2272-2276.
- Karaboga, D., & Kaya, E. (2019). Adaptive network based fuzzy inference system (ANFIS) training approaches: A comprehensive survey. *Artificial Intelligence Review*, 52(4), 2263-2293. <https://doi.org/10.1007/s10462-017-9610-2>
- Kaur, M., Gulati, H., & Kundra, H. (2014). Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications. *International Journal of Computer Applications*, 99(12), 1-3. <https://doi.org/10.5120/17422-8273>
- Kirgiz, A. (2021). Lüks Otomotiv Sektöründe Satın Alma Kararını Etkileyen Faktörler. *R&S - Research Studies Anatolia Journal*. <https://doi.org/10.33723/rs.860512>
- Köylüoğlu, A. S., Acar, Ö. E., & İnan, Ü. S. E. (2018). Tüketicilerin Otomobil Satın Alma Davranışlarına Etki Eden Faktörlerin Belirlenmesi: Akademisyenlere Yönelik Bir Uygulama. *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, 21(2), 251-273. <https://doi.org/10.29249/selcuksbmyd.448302>
- Liebchen, G., Twala, B., Shepperd, M., Cartwright, M., & Stephens, M. (2007). Filtering, Robust Filtering, Polishing: Techniques for Addressing Quality in Software Data. *First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)*, 99-106. <https://doi.org/10.1109/ESEM.2007.70>
- Liu, B., Xia, Y., & Yu, P. S. (2000). Clustering Through Decision Tree Construction. *Conference Of Information And Knowledge Management*, 10.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill Inc., New York, (414)

- Özekes, S. (2003). Veri Madenciliği Modelleri ve Uygulama Alanları. *İstanbul Ticaret Üniversitesi Dergisi*. (65-82)
- Özgür, A. (2002). *Supervised And Unsupervised Machine Learning Techniques For Text Document Categorization*. Boğaziçi Üniversitesi/Bilgisayar Mühendisliği, Lisans Tezi.(98)<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=09c158245241cef51bdca6a4cb68f28926d64eff>
- Öztemel, E. (2012). Yapay Sinir Ağları. Papatya Yayıncılık. (3. bs).İstanbul.
- Öztürk, K., & Şahin, M. E. (2018). Yapay Sinir Ağları ve Yapay Zekâ'ya Genel Bir Bakış. *Takvim-i Vekayı*, 6(2), 25-36.
- Patel, P., Sivaiah, B., & Patel, R. (2022). Approaches for finding Optimal Number of Clusters using K-Means and Agglomerative Hierarchical Clustering Techniques. *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCCSP)*, 1-6. <https://doi.org/10.1109/ICICCCSP53532.2022.9862439>
- Polat, E. D. (2022). Beyaz eşya sektöründe satış tahmini: Bir veri madenciliği uygulaması. *Bursa Uludağ Üniversitesi/Sosyal Bilimler Enstitüsü.Yüksek Lisans Tezi (79)* (<http://acikerisim.uludag.edu.tr/jspui/handle/11452/26450>)
- Ranganathan, A. (2004). The Levenberg-Marquardt Algorithm. (1-5).
- Rençber, O., & Akbulut, R. (2016). ANFIS Yöntemi ile Finansal Başarısızlıkların Tahmin Edilmesi: BIST'te Ticaret ve Hizmet Sektörü Üzerine Bir Uygulama. *17.Uluslararası Ekonometri, Yöneylem ve İstatistik Kongresi*,2-4 Haziran 2016, Sivas.
- Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (Ed.). (2011). Recommender Systems Handbook. Springer US. <https://doi.org/10.1007/978-0-387-85820-3>
- Rokaha, B., Ghale, D. P., & Gautam, B. P. (2018). Enhancement of Supermarket Business and Market Plan by Using Hierarchical Clustering and Association Mining Technique. *2018 International Conference on Networking and Network Applications (NaNA)*, 384-389. <https://doi.org/10.1109/NANA.2018.8648716>
- Sakar, C. O., Kursun, O., & Gurgun, F. (2012). A feature selection method based on kernel canonical correlation analysis and the minimum Redundancy–Maximum Relevance filter method. *Expert Systems with Applications*, 39(3), 3432-3437. <https://doi.org/10.1016/j.eswa.2011.09.031>
- Sakız, B., & Gencer, A. H. (2018). Forecasting the Bitcoin Price via Artificial Neural Networks. *International Conference on Eurasian Economies* 438-444. <https://doi.org/10.36880/C10.02070>
- Saltan, M., & Alaefary, F. (2018). Anfis Yönteminin Esnek Yol Üstyapılarının Yapısal Analizinde Kullanılabilirliği. *Mühendislik Bilimleri ve Tasarım Dergisi*, 6(3), 32-41. <https://doi.org/10.21923/jesd.313427>
- Saputra, D. M., Saputra, D., & Oswari, L. D. (2019). Effect of Distance Metrics in Determining K-Value in KMeans Clustering Using Elbow and Silhouette Method. *Advances in Intelligent Systems Research, Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*(172). <https://doi.org/10.2991/aisr.k.200424.051>
- Savaş, S., Topaloğlu, N., & Yılmaz, M. (2012). Veri Madenciliği Ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*. 11(21) s. 1-23.

- Sel, A., & Tekgün, B. (2022). Anfis Yöntemi İle Türkiye Karbondioksit Salinimi Tahmini. *Visionary E-Journal / Vizyoner Dergisi*, 13(34), 486-504. <https://doi.org/10.21076/vizyoner.990380>
- Serin, D. Z. V. (2020). Seçili Krito Paralarda Kümeleme Analizi. *Türkiye Sosyal Bilimler Araştırma Dergisi*, 5, 41-52.
- Shobha, N., & Asha, T. (2017). Monitoring weather based meteorological data: Clustering approach for analysis. *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 75-81. <https://doi.org/10.1109/ICIMIA.2017.7975575>
- Sutcu, E. (2018). Estimation of Welded Manufacturing Operation Times with Matlab Artificial Neural Networks. *Gazi Journal of Engineering Sciences*, 4(1), 16-26. <https://doi.org/10.30855/gmbd.2018.04.01.003>
- Tekin, B. (2018). Ward, K-Ortalamlar Ve İki Adımlı Kümeleme Analizi Yöntemleri İle Finansal Göstergeler Temelinde Hisse Senedi Tercih. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 1-1. <https://doi.org/10.31795/baunsobed.492464>
- The Mathworks. (t.y.). *Neural Net Fitting*. The Mathworks. https://ch.mathworks.com/help/deeplearning/ref/neuralnetfitting-app.html?searchHighlight=Levenberg-Marquardt&s_tid=srchtitle_Levenberg-Marquardt_8
- Tolon, M., & Güner, N. (2008). Tüketici Tatmini Verilerinin Analizi: Yapay Sinir Ağları Ve Regresyon Analizi Karşılaştırması. *Gazi Üniversitesi İktisadi Ve İdari Bilimler Dergisi*, 2(20), 247-259.
- Wang, X., & Xu, Y. (2019). An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. *IOP Conference Series: Materials Science and Engineering*, 569(5), 052024. <https://doi.org/10.1088/1757-899X/569/5/052024>
- Yücel, A. (2010). Tedarikçi Seçimi Probleminde Bütünleşik Sinirsel Bulanık Mantık Yaklaşımı. *Yıldız Teknik Üniversitesi/Fen Bilimleri Enstitüsü. Doktora Tezi*. (135)
- Yüksek, A. G., Arslan, H., & Kaynar, O. (2017). Comparison of the effects dimensionality methods in the training of neuro-fuzzy (ANFIS) classifications. *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 1-9. <https://doi.org/10.1109/IDAP.2017.8090204>
- Zhao, Y., Karypis, G., & Fayyad, U. (2005). Hierarchical Clustering Algorithms for Document Datasets. *Data Mining and Knowledge Discovery*, 10(2), 141-168. <https://doi.org/10.1007/s10618-005-0361-3>

EKLER

- EK 1** ANFIS Kurallar Çizelgesi
- EK 2** R Studio Veri Madenciliđi ve Kümeleme Kodları
- EK 3** YSA Matlab Kodları

EK-1**ANFIS KURALLAR ÇİZELGESİ**

Kalite	Kalınlık	Boy	Tüketim	Ürün grubu	Kesim Tipi	Fiyat
Düşük	Yüksek	Düşük	Orta	Orta	Orta	Düşük
Düşük	Yüksek	Düşük	Orta	Düşük	Yüksek	Düşük
Düşük	Orta	Düşük	Orta	Düşük	Yüksek	Düşük
Düşük	Orta	Düşük	Orta	Orta	Orta	Orta
Orta	Orta	Düşük	Düşük	Düşük	Yüksek	Düşük
Orta	Orta	Düşük	Orta	Düşük	Yüksek	Düşük
Yüksek	Orta	Düşük	Düşük	Yüksek	Düşük	Orta
Yüksek	Yüksek	Düşük	Düşük	Yüksek	Düşük	Yüksek
Orta	Yüksek	Düşük	Düşük	Düşük	Yüksek	Düşük
Düşük	Orta	Düşük	Orta	Orta	Orta	Orta
Orta	Orta	Düşük	Düşük	Düşük	Yüksek	Düşük
Yüksek	Orta	Düşük	Düşük	Yüksek	Düşük	Yüksek
Orta	Yüksek	Düşük	Düşük	Düşük	Yüksek	Düşük
Yüksek	Orta	Düşük	Orta	Orta	Orta	Düşük
Yüksek	Orta	Düşük	Düşük	Yüksek	Düşük	Düşük
Yüksek	Orta	Düşük	Orta	Yüksek	Düşük	Orta
Düşük	Orta	Düşük	Düşük	Yüksek	Düşük	Orta
Yüksek	Orta	Düşük	Yüksek	Orta	Orta	Düşük
Yüksek	Orta	Düşük	Yüksek	Yüksek	Düşük	Orta
Düşük	Orta	Düşük	Düşük	Orta	Orta	Düşük
Yüksek	Orta	Orta	Orta	Yüksek	Düşük	Yüksek
Orta	Orta	Düşük	Düşük	Orta	Orta	Orta
Yüksek	Orta	Düşük	Orta	Orta	Orta	Orta
Yüksek	Orta	Düşük	Yüksek	Orta	Orta	Yüksek
Yüksek	Orta	Düşük	Orta	Yüksek	Düşük	Yüksek
Orta	Orta	Düşük	Orta	Orta	Orta	Yüksek

Confidential C

Yüksek	Orta	Düşük	Düşük	Yüksek	Düşük	Yüksek
Yüksek	Yüksek	Düşük	Düşük	Yüksek	Düşük	Yüksek
Yüksek	Düşük	Yüksek	Yüksek	Orta	Orta	Düşük
Yüksek	Düşük	Orta	Yüksek	Yüksek	Düşük	Orta
Düşük	Orta	Düşük	Düşük	Orta	Orta	Düşük
Yüksek	Düşük	Orta	Orta	Yüksek	Düşük	Yüksek
Orta	Orta	Orta	Düşük	Orta	Orta	Orta
Düşük	Düşük	Yüksek	Orta	Orta	Yüksek	Orta
Yüksek	Orta	Düşük	Yüksek	Orta	Yüksek	Yüksek
Düşük	Düşük	Düşük	Orta	Orta	Orta	Orta
Orta	Yüksek	Orta	Düşük	Düşük	Yüksek	Düşük
Yüksek	Orta	Orta	Düşük	Yüksek	Düşük	Yüksek

EK 2

R Studio Veri Madenciliği ve Kümeleme Kodları

```
##seçilen kütüphaneler
```

```
library(dplyr)
```

```
library(tidyverse)
```

```
library(ggplot2)
```

```
library(factoextra)
```

```
library(class)
```

```
library(cluster)
```

```
library(cIValid)
```

```
# Fiyat Tüketim Grafiği
```

```
ggplot(data=dataset,aes(x=Fiyat,y=Tüketim)) + geom_point()
```

```
#Ciro ve Tüketim Grafiği
```

```
ggplot(data=dataset,aes(x=Ciro/1000,y=Tüketim)) + geom_point()
```

```
#Gürültülü veri temizliği (Negatif ve -500 €/Ton fiyatların temizlenmesi).
```

```
new_dataset <- filter(dataset, Fiyat > 500)
```

```
#only serial references preferred
```

```
new_dataset2 <- filter(new_dataset, Durum == "EVET")
```

```
#unknown projects has cleared
```

```
new_dataset3 <- na.omit(new_dataset2)
```

```
#Ciro Analizi
```

```
ciroanaliz <- select(new_dataset3, Projeno, Ciro)
```

```
ciroanaliz1 <- group_by(ciroanaliz, Projeno)
```

```
ciroanaliz2 <- summarise(ciroanaliz1, sum(Ciro))
```

```
# En yüksek turnover analizi
```

```
arrange(ciroanaliz2, desc(`sum(Ciro)`))
```

```

boxplot(ciroanaliz2$Projeno)
boxplot(ciroanaliz2$`sum(Ciro)` )
hist(ciroanaliz2$`sum(Ciro)`/1000)
hist(dataset$Fiyat,xlab = "Fiyat",ylab = "Sıklık")
hist(new_dataset3$Fiyat,xlab = "Fiyat",ylab = "Sıklık")

#Dağılımı anlamak için Anderson Darling uygulaması
library(nortest)
ad.test(new_dataset3$Fiyat)
ad.test(dataset$Fiyat)

#Yeni durumun grafikleri
ggplot(data=new_dataset3,aes(x=Fiyat,y=Tüketim)) + geom_point()
ggplot(data=new_dataset3,aes(x=Ciro/1000,y=Tüketim)) + geom_point()

#Datayı indirme kodları
#library(openxlsx)
#write.xlsx(new_dataset3, file="newdataset3.xlsx")
#write.xlsx(scaledata, file="scaledata.xlsx")

#Kümeleme adımları
library(factoextra)
mydata.labes <- new_dataset3$Projeno
clustering1=select(new_dataset3,c(3,9,10,11))
clustering2=select(clustering1,c(1,4))
scaledata3 <- scale(clustering2)

#Uzaklık hesapları
clustering2<- dist(scaledata3,method = "euclidian")

```



```

#K-Ortalamlar ile ideal küme sayısı
fviz_nbclust(scaledata3, kmeans, method = "wss") + labs(subtitle="Elbow Method")
#fviz_nbclust(scaledata, kmeans, method = "silhouette")+ labs(subtitle="Silhouette
Method")
fviz_nbclust(scaledata3, kmeans, method = "silhouette") + labs(subtitle="Silhouette
Method")
fviz_nbclust(scaledata3, kmeans, method = "gap_stat")+ labs(subtitle="Gap Statistics
Method")
km.out <- kmeans(scaledata3, centers = 4, nstart=300)
print(km.out)

#Kümeleri görselleştirme
km.clusters <- km.out$cluster
rownames(scaledata3) <- paste(new_dataset3$Projeno, 1:dim(scaledata3)[1], sep = "_")
fviz_cluster(list(data=scaledata3, cluster=km.clusters))
table(km.clusters, new_dataset3$Projeno)

table(km.clusters, new_dataset3$Durum)

clmethods <- c("hierarchical", "kmeans", "pam")
intern <- clValid(scaledata3, nClust = 3:7, clMethods = clmethods,
validation = "internal")
summary(intern)

#Hiyerarşik kümeleme
hierarch <- hclust(dist(scaledata3), method = "ward.D2")
plot(hierarch)
rect.hclust(hierarch, k = 3, border = 2:5)
sub_grp <- cutree(hierarch, k = 3)
fviz_cluster(list(data = scaledata3, cluster = sub_grp))
table(sub_grp, new_dataset3$Projeno)

```

Confidential C

EK 3

YSA MATLAB KODU

```
x = input;
```

```
t = output;
```

```
trainFcn = 'trainlm'; % Levenberg-Marquardt
```

```
hiddenLayerSize = 4;
```

```
net = fitnet(hiddenLayerSize,trainFcn);
```

```
net.input.processFcns = {'removeconstantrows','mapminmax'};
```

```
net.output.processFcns = {'removeconstantrows','mapminmax'};
```

```
net.divideFcn = 'dividerand'; % Divide data randomly
```

```
net.divideMode = 'sample'; % Divide up every sample
```

```
net.divideParam.trainRatio = 80/100;
```

```
net.divideParam.valRatio = 10/100;
```

```
net.divideParam.testRatio = 10/100;
```

```
net.performFcn = 'mse'; % Mean Squared Error
```

```
net.plotFcns = {'plotperform','plottrainstate','ploterrhist', ... 'plotregression', 'plotfit'};
```

```
[net,tr] = train(net,x,t);
```

```
y = net(x);
```

```
e = gsubtract(t,y);
```

```
performance = perform(net,t,y)
```

```
trainTargets = t .* tr.trainMask{1};
```

```
valTargets = t .* tr.valMask{1};
```

Confidential C

```

testTargets = t .* tr.testMask{1};
trainPerformance = perform(net,trainTargets,y)
valPerformance = perform(net,valTargets,y)
testPerformance = perform(net,testTargets,y)

view(net)
figure, plotperform(tr)
figure, plottrainstate(tr)
figure, ploterrhist(e)
figure, plotregression(t,y)
figure, plotfit(net,x,t)

if (false)
    genFunction(net,'myNeuralNetworkFunction');
    y = myNeuralNetworkFunction(x);
end
if (false)
    genFunction(net,'myNeuralNetworkFunction','MatrixOnly','yes');
    y = myNeuralNetworkFunction(x);
end
if (false)
    gensim(net);
end

```

ÖZGEÇMİŞ

Adı Soyadı : Seray Mirasçı
Doğum Yeri ve Tarihi : Bursa – 24.11.1996
Yabancı Dil : İngilizce, İbranice, Almanca

Eğitim Durumu
Lise : Bursa Anadolu Kız Lisesi
Lisans : Pamukkale Üniversitesi – Endüstri Mühendisliği
: Pamukkale Üniversitesi – Makine Mühendisliği (Yandal)
Yüksek Lisans : Uludağ Üniversitesi – Endüstri Mühendisliği

Çalıştığı Kurum/Kurumlar : Maysan Mando Otomotiv Par. San. Ve Tic. A.Ş – Satın alma Mühendisi (2018-2021)
Valeo Otomotiv Sanayi ve Ticaret A.S. – Satın alma Mühendisi (2021-2022)
Oyak Renault Otomobil Fabrikaları A.Ş. – Satın alma Mühendisi (2022-Halen)

İletişim (e-posta) : seraymir@hotmail.com

Yayımları : Mirasçı, S., Aksoy, A. (2022). Sac Malzeme Satın Alma Süreçlerinde Veri Madenciliği Ve Makine Öğrenmesi Uygulamaları. Yöneylem Araştırması/Endüstri Mühendisliği 41. Ulusal Kongresi (YA/EM 2022), 26-28 Ekim 2022, Pamukkale Üniversitesi, Denizli (145)