

**TAHMİN PROBLEMLERİ İÇİN REGRESYON AĞACI  
VE KOMŞULUK TABANLI YÖNTEMLER  
GELİŞTİRİLMESİ: KALIPÇILIK SEKTÖRÜNDE BİR  
UYGULAMA**

**Gözde ESER**



T.C.  
BURSA ULUDAĞ ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

**TAHMİN PROBLEMLERİ İÇİN REGRESYON AĞACI VE KOMŞULUK  
TABANLI YÖNTEMLER GELİŞTİRİLMESİ: KALIPÇILIK SEKTÖRÜNDE  
BİR UYGULAMA**

**Gözde ESER**  
0000-0002-5417-5442

Doç. Dr. Tülin İNKAYA  
(Danışman)

YÜKSEK LİSANS TEZİ  
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI

BURSA – 2021  
Her Hakkı Saklıdır

## TEZ ONAYI

Gözde ESER tarafından hazırlanan “TAHMİN PROBLEMLERİ İÇİN REGRESYON AĞACI VE KOMŞULUK TABANLI YÖNTEMLER GELİŞTİRİLMESİ: KALIPÇILIK SEKTÖRÜNDE BİR UYGULAMA” 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. Tülin İNKAYA

**Başkan** : Doç. Dr. Tülin İNKAYA  
0000-0002-6260-0162  
Bursa Uludağ Üniversitesi,  
Mühendislik Fakültesi,  
Endüstri Mühendisliği Anabilim Dalı

İmza

**Üye** : Doç. Dr. Aytaç YILDIZ  
0000-0002-0729-633X  
Bursa Teknik Üniversitesi,  
Mühendislik ve Doğa Bilimleri Fakültesi,  
Endüstri Mühendisliği Anabilim Dalı

İmza

**Üye** : Doç. Dr. Duygu YILMAZ EROĞLU  
0000-0002-7730-2707  
Bursa Uludağ 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 çalışmada;**

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

**beyan ederim.**

**03/02/2021**

  
**Gözde ESER**

## ÖZET

Yüksek Lisans Tezi

TAHMİN PROBLEMLERİ İÇİN REGRESYON AĞACI VE KOMŞULUK TABANLI YÖNTEMLER GELİŞTİRİLMESİ: KALIPÇILIK SEKTÖRÜNDE BİR UYGULAMA

**Gözde ESER**

Bursa Uludağ Üniversitesi  
Fen Bilimleri Enstitüsü  
Endüstri Mühendisliği Anabilim Dalı

**Danışman:** Doç. Dr. Tülin İNKAYA

Üretim ile hizmet sistemlerinde gerçekçi ve hızlı karar almak firmalara rekabet avantajı kazandırmaktadır. Bilgi teknolojilerindeki gelişmeler firmaların büyük miktarda veriye kolay erişimini sağlamaktadır. Ancak sayısal değerlerin tahmininin yapılması firmaların karşılaştığı büyük zorluklardandır. Bu çalışmada tahmin problemlerinin çözümü için veri madenciliğine dayalı bir metodoloji önerilmiştir. Önerilen metodolojide ağaç tabanlı yöntemler ve komşuluk tabanlı yöntemler kullanılmıştır. Ağaç tabanlı yöntemler Regresyon Ağacı, Torbalama Regresyon Ağacı ve Güçlendirme Regresyon Ağacıdır. Komşuluk tabanlı yöntemler, K-En Yakın Komşuluk ve Torbalama K- En Yakın Komşuluktur. Tahmin modelleri oluşturulurken veri kümelerindeki nesnelerin yerel aykırı değer faktörlerini, uzaklıklarını ve en yakın komşuluk sıralamasını dikkate alan ağırlıklı tahmin fonksiyonları kullanılmıştır. Aykırı değer analizi çalışması yapılarak tahmin modellerinin doğruluğunun artırılması hedeflenmiştir. Önerilen yaklaşımların performansı dokuz adet karşılaştırmalı değerlendirme veri kümesi üzerinde test edilmiştir. Yapılan karşılaştırmalarda aykırı değer analizi ile veri önileme yapıldıktan sonra ağırlıklı tahmin fonksiyonları kullanılarak geliştirilen topluluk yöntemlerin doğruluğu arttırdığı görülmüştür. Ayrıca sac metal kalıp imalatı yapan bir firmada kalıp üretim sürelerinin tahmini için bir vaka analizi çalışması yapılmıştır. Firmanın 2015-2018 yılları arasında üretimini tamamladığı 85 kalıba ait veriler kullanılarak geliştirilen modellerin performansları değerlendirilmiştir. İstatistiksel sonuçlar önerilen yaklaşım ile tahmin doğruluğunun arttığını göstermiştir.

**Anahtar Kelimeler:** Veri madenciliği, ağaç tabanlı yöntemler, komşuluk tabanlı yöntemler, topluluk yöntemler, ağırlıklı tahmin fonksiyonu

**2021, x + 99 sayfa.**

## **ABSTRACT**

MSc Thesis

### **DEVELOPMENT OF REGRESSION TREE AND NEIGHBORHOOD-BASED METHODS FOR PREDICTION PROBLEMS: AN APPLICATION IN THE DIE INDUSTRY**

**Gözde ESER**

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

**Supervisor:** Assoc. Prof. Dr. Tülin İNKAYA

Making realistic and fast decisions in production and service systems gives companies a competitive advantage. Developments in information technologies provide companies with easy access to large amounts of data. However, estimating numerical values is one of the major challenges faced by companies. In this study, a methodology based on data mining is proposed for the solution of prediction problems. Tree-based and neighborhood-based methods are used in the proposed methodology. Tree-based methods are Regression Tree, Bagging Regression Tree, and Boosting Regression Tree. Neighborhood-based methods,  $K$ -The Nearest Neighborhood and Bagging  $K$ -The Nearest Neighborhood. Weighted estimation functions that take into account the local outlier factors, distances and the nearest neighborhood order of the objects in the data sets were used while creating the prediction models. It was aimed to increase the accuracy of the prediction models by performing an outlier analysis study. The performance of the proposed approaches was tested on nine comparative evaluation datasets. In the comparisons, it was observed that the ensemble methods developed by using weighted estimation functions after data preprocessing with outlier analysis increased the accuracy. In addition, a case study was conducted to estimate the die production times in a company that manufactures sheet metal dies. The performances of the developed models were evaluated using the data of 85 dies produced by the company between 2015-2018. Statistical results showed that the accuracy of the prediction increased with the proposed approach.

**Key words:** Data mining, tree-based methods, neighborhood-based methods, community methods, weighted prediction function

**2021, x + 99 pages**

## TEŐEKKÜR

Yüksek lisans eğitimim ve tez çalışmam süresince her zaman yanımda olan, büyük bir özveri ile bana her konuda yol gösteren ve motive eden, değerli fikirleriyle tez çalışmamı yönlendiren, desteğini hiçbir zaman esirgemeyen danışman hocam sayın Doç. Dr. Tülin İnkaya'ya en başta sonsuz teşekkür etmek istiyorum.

Bugünlere gelmemin tek sebebi, annem Eda Öngelen ve babam Nafi Öngelen'e her zaman yanımda olduğunu bildiğim sevgili ablam Sevda Öngelen ve canım yeğenim Elif Beril Akaret'e çok teşekkür ederim.

Hayat yolculuğumda bana eşlik eden, her zaman yanımda olan ve tüm akademik kariyerim boyunca sonsuz desteğiyle beni yalnız bırakmayan eşim, Emrah Eser'e teşekkür ederim.

Akademik kariyerim boyunca bana güvenen, motive eden, bilgisini ve desteğini esirgemeyen sayın yöneticim Onur Ekdik'e teşekkür ederim.

Benimle birlikte yorulan, heyecanlanan, gülen, ağlayan, varlığıyla kendimi şanslı hissettiren tüm arkadaşlarıma ve benden desteğini esirgemeyen başta Begüm Erol olmak üzere tüm dostlarıma teşekkür ederim.

Çalışmalarımın tamamını tüm aileme ithaf ediyorum.

Gözde ESER  
03/02/2021

## İÇİNDEKİLER

	Sayfa
ÖZET.....	i
ABSTRACT.....	ii
TEŞEKKÜR.....	iii
SİMGELER ve KISALTMALAR DİZİNİ.....	vi
ŞEKİLLER DİZİNİ.....	ix
ÇİZELGELER DİZİNİ.....	x
1. GİRİŞ.....	1
2. KAYNAK ARAŞTIRMASI.....	3
2.1. Regresyon Ağaçları.....	3
2.2. En Yakın Komşuluk Tabanlı Yöntemler.....	6
2.3. Topluluk Yöntemler.....	9
2.4. Aykırı Veri Analizi.....	10
3. MATERYAL ve YÖNTEM.....	12
3.1. Materyal.....	12
3.2. Yöntem.....	13
3.2.1. Regresyon ağacı.....	13
3.2.2. K-en yakın komşu yöntemi.....	14
3.2.3. Benzemezlik (dissimilarity) ölçüsü.....	15
3.2.4. Performans değerlendirme yöntemleri.....	16
3.2.5. Torbalama ve güçlendirme yöntemleri.....	18
3.2.6. Yerel aykırı değer faktörü.....	21
3.2.7. Önerilen yaklaşımlar.....	22
4. BULGULAR ve TARTIŞMA.....	30
4.1. Karşılaştırmalı Değerlendirme Veri Kümeleri.....	30
4.2. Performans Metrikleri.....	32
4.3. Karşılaştırmalı Değerlendirme Veri Kümelerinin Sayısal Sonuçları.....	33
4.4. Tahmin Yöntemlerinin Karşılaştırılması.....	38
4.4.1. Friedman testi.....	39
4.4.2. Wilcoxon eşleştirilmiş işaretli sıra testi.....	42
4.5. Sac Metal Kalıpcılık Sektöründe Üretim Sürelerinin Tahmini için Vaka Analizi...47	
4.5.1. Üretim ortamı.....	48
4.5.2. Veri kümesi tanımlama.....	51
4.5.3. Vaka analizi bulguları.....	52
5. SONUÇ.....	55
KAYNAKLAR.....	57
EKLER.....	65
EK 1 Önerilen Regresyon Ağacı ile Tahmin Modeli Algoritması.....	66
EK 2 Önerilen Torbalama Regresyon Ağacı ile Tahmin Modeli Algoritması.....	67
EK 3 Önerilen Güçlendirme Regresyon Ağacı ile Tahmin Modeli Algoritması.....	68
EK 4 Önerilen KNN Algoritması.....	69



EK 5 Önerilen Torbalama KNN Algoritması .....	70
EK 6 Orijinal Veri Kümelerinde KNN ve TKNN için MAPE Değerleri (%) .....	71
EK 7 Orijinal Veri Kümelerinde KNN ve TKNN için RMSE Değerleri .....	76
EK 8 Aykırı Değer Analizi Yapılan Veri Kümelerinde KNN ve TKNN için MAPE (%) Değerleri.....	81
EK 9 Aykırı Değer Analizi Yapılan Veri Kümelerinde KNN ve TKNN için RMSE Değerleri.....	86
EK 10 Orijinal Veri Kümelerine Komşuluk Tabanlı Yöntemlerin Uygulanması ile Elde Edilen MAPE Sonuçlarına Göre $R_j$ Değerleri .....	91
EK 11 Orijinal Veri Kümelerine Komşuluk Tabanlı Yöntemlerin Uygulanması ile Elde Edilen RMSE Sonuçlarına Göre $R_j$ Değerleri .....	92
EK 12 Aykırı Değer Analizi Yapılan Veri Kümelerine Komşuluk Tabanlı Yöntemlerin Uygulanması ile Elde Edilen MAPE Sonuçlarına Göre $R_j$ Değerleri.....	93
EK 13 Aykırı Değer Analizi Yapılan Veri Kümelerine Komşuluk Tabanlı Yöntemlerin Uygulanması ile Elde Edilen RMSE Sonuçlarına Göre $R_j$ Değerleri.....	94
EK 14 <i>Kalıp</i> Veri Kümesinin Komşuluk Tabanlı Yöntemler için MAPE Sonuçları (%) .....	95
EK 15 <i>Kalıp</i> Veri Kümesinin Komşuluk Tabanlı Yöntemler için RMSE Sonuçları.....	96
EK 16 Öznitelik Seçimi Sonrası <i>Kalıp 2</i> Veri Kümesinin Komşuluk Tabanlı Yöntemler İçin MAPE Sonuçları (%) .....	97
EK 17 Öznitelik Seçimi Sonrası <i>Kalıp 2</i> Veri Kümesinin Komşuluk Tabanlı Yöntemler için RMSE Sonuçları .....	98
ÖZGEÇMİŞ .....	99

## SİMGELER ve KISALTMALAR DİZİNİ

Simgeler	Açıklama
$a$	Güç katsayısı
$Acc(M)$	Modelin doğruluğu (Accuracy of the model)
$B$	Topluluk yaklaşımında oluşturulacak model sayısı
$\beta_b$	$b$ . modelin ağırlık güçlendirme değeri
$D$	Veri kümesi
$D_i^*$	$i$ . nesnenin düştüğü yaprak düğümündeki veri kümesi
$D_b$	$b$ . modelin eğitim veri kümesi
$D_b^T$	$b$ . modelin test veri kümesi $D_b^T = D \setminus \{D_b\}$
$d(x_p, x_q)$	$x_p$ nesnesinin $x_q$ nesnesine olan uzaklığı
$derinlik_j$	$j$ . düğüm derinliği
$derinlik_{max}$	Maksimum derinliği
$\epsilon$	Beklenen iyileşme miktarı
$\varepsilon$	Hata miktarı
$f_{fin}(x)$	Topluluk model çıktısı
$k$	Çapraz geçiş sayısı
$K$	En yakın komşuluk sayısı
$Kdistance$	Bir nesnenin $K$ . en yakın komşusu ile arasındaki mesafe
$k(x_p, x_q)$	$x_q$ nesnesinin $x_p$ nesnesine olan en yakın komşuluk sırası
$lof$	Yerel aykırı değer faktörü
$lrd$	Yerel ulaşılabilirlik yoğunluğu
$m$	Veri kümesindeki öznitelik sayısı
$M_b$	Topluluktaki $b$ . tahmin modeli
$M^*$	Birleşik model
$MinPts$	Minimum nesne sayısı
$n$	Veri kümesindeki nesne sayısı
$n_{min}$	Bir düğümün bölünebilmesi için gerekli olan minimum nesne sayısı
$n^T$	Test veri kümesinin nesne sayısı
$n_i^*$	$i$ . nesnenin düştüğü yaprak düğümdeki nesne sayısı
$nesneSayısı_j$	$j$ . düğümündeki nesne sayısı
$N_{Kdistance}$	Bir nesnenin $K$ komşuluk mesafesi
$reachdist_K$	Ulaşılabilirlik mesafesi
$s(x_p, x_q)$	$x_p$ nesnesinin $x_q$ nesnesine olan benzerliği
$\varphi$	Doğru ve yanlış tahminleri ayırmak için eşik değeri
$w_i$	$i$ . nesnenin ağırlığı
$W$	Güçlendirme regresyon ağaçlarının ağırlık vektörü
$x_i$	$i$ . nesne
$x^T$	Test nesnesi
$x_{ij}$	Veri kümesindeki $i$ . nesnenin $j$ . öznitelikte aldığı değer
$X_j$	Veri kümesindeki $j$ . öznitelik
$X^*$	Modelin hata miktarını en aza indiren öznitelik

**Simgeler****Açıklama**

$y_i$	<i>i.</i> nesnenin çıktı değişkeninin gerçek değeri
$y_i^T$	<i>i.</i> nesnenin çıktı değişkeninin tahmin değeri
$\bar{Y}_j$	<i>j.</i> düğümün tahmin değeri
Z	Normalleştirme faktörü

**Kısaltmalar****Açıklama**

AID	Otomatik Etkileşim Algılama (Automatic Interaction Detection)
ANOVA	Varyans Analizi (Analysis of Variance)
ARE	Mutlak Bağıl Hata (Absolute Relative Error)
AR-LOF	Boyut Derecelendirme Yerel Aykırı Faktörü (Aspect-Rating Local Outlier Factor)
ATF	Ağırlıklı Tahmin Fonksiyonu
BEM	Büyük Hata Marjlı (Big Error Margin)
CAD	Bilgisayar Destekli Tasarım (Computer Aided Design)
CAM	Bilgisayar Destekli İmalat (Computer Aided Manufacturing)
CART	Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Trees)
CHAID	Ki-kare Otomatik Etkileşim Algılama (CHi-squared Automatic Interaction Detector)
CNC 2D	İki Boyutlu Bilgisayarlı Sayısal Kontrol (Two Dimensional Computer Numerical Control)
CNC 3D	Üç Boyutlu Bilgisayarlı Sayısal Kontrol (Three Dimensional Computer Numerical Control)
CRUISE	Tarafsız Etkileşim Seçimi ve Tahmini ile Sınıflandırma Kuralı (Classification Rule with Unbiased Interaction Selection and Estimation)
CTREE	Koşullu Çıkarım Ağaçları (Conditional Inference Trees)
FACT	Hızlı ve Doğru Sınıflandırma Ağacı (Fast and Accurate Classification Tree)
GBART	Değişken Gruplama Tabanlı Bayes Katkı Regresyon Ağacı (Variable Grouping Based Bayesian Additive Regression Tree)
GI-KNN	Küresel Bilgilendirici K-en Yakın Komşu (Globally Informative K-Nearest Neighbor)
GRA	Güçlendirme Regresyon Ağacı
GUIDE	Genelleştirilmiş, Tarafsız, Etkileşim Algılama ve Tahmin (Generalized, Unbiased, Interaction Detection and Estimation)
KMRA	K-en Yakın Komşu Algoritması ve Çoklu Regresyon Algoritması (K-Nearest Neighbor and Multiple Regression Algorithm)
KNCN	K-En Yakın Ağırlık Merkezi Komşuları (K-Nearest Centroid Neighbors)
KNN	K-En Yakın Komşu (K-Nearest Neighbor)
KNN-KFGP	Gauss Süreç Regresyonu için K-En Yakın Komşu tabanlı Kalman Filtresi (K-Nearest Neighbor Based Kalman Filter for Gaussian Process Regression)

<b>Kısaltmalar</b>	<b>Açıklama</b>
LI-KNN	Yerel Bilgilendirici $K$ -en Yakın Komşu (Locally Informative $K$ -Nearest Neighbor)
LOF	Yerel Aykırı Değer Faktörü (Local Outlier Factor)
LOOCV	Birini Dışarıda Bırak Çapraz Geçerleme (Leave One Out Cross Validation)
MAPE	Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error)
MFS	Özellik Alt Kümeleri (Multiple Feature Subsets)
NN	En Yakın Komşu (Nearest Neighbor)
OCR	Optik Karakter Tanıma (Optical Character Recognition)
QUEST	Hızlı, Tarafsız ve Etkili İstatistik Ağacı (Quick, Unbiased and Efficient Statistical Tree)
PRP	Yayınlanan Göreceli Performans (Published Relative Performance)
RA	Regresyon Ağacı
RMSE	Hata Kareler Ortalamasının Karekökü (Root Mean Square Error)
RSS	Artık Kareler Toplamı (Residual Sum of Squares)
RTA	Regresyon Gövde Yaklaşımı (Regression Trunk Approach)
RTC	Güvenilirlik Regresyon Ağacı (Regression Tree Credibility)
STIMA	Eşzamanlı Eşik Etkileşim Modelleme Algoritması (Simultaneous Threshold Interaction Modeling Algorithm)
SUPPORT	Düzleştirilmiş ve Düzleştirilmemiş Parçalı Polinom Regresyon Ağaçları (Smoothed and Unsmoothed Piecewise Polynomial Regression Trees)
THAID	THeta Otomatik Etkileşim Algılama (THeta Automatic Interaction Detection)
TKNN	Torbalama $K$ -En Yakın Komşuluk
TRA	Torbalama Regresyon Ağacı

## ŞEKİLLER DİZİNİ

	Sayfa
Şekil 3.1. Regresyon ağacı algoritmasının temel adımları.....	14
Şekil 3.2. $K$ -en yakın komşu algoritmasının temel adımları.....	15
Şekil 3.3. $k$ -kat çapraz geçirme yönteminin şematik gösterimi.....	16
Şekil 3.4. Önyükleme yönteminin şematik gösterimi .....	18
Şekil 3.5. Topluluk yöntemlerin şematik gösterimi .....	18
Şekil 3.6. Torbalama algoritmasının temel adımları.....	19
Şekil 3.7. Adaboost.RT algoritmasının temel adımları.....	20
Şekil 3.8. Yerel aykırı değer faktörü algoritmasının temel adımları.....	22
Şekil 3.9. Veri ön işleme akış şeması .....	23
Şekil 3.10. Uzaklık matrisinin ve sıralama matrisinin oluşturulması .....	23
Şekil 3.11. Ağaç tabanlı yöntemlerin akış şeması .....	24
Şekil 3.12. Komşuluk tabanlı yöntemlerin akış şeması .....	25
Şekil 3.13. Ağaç tabanlı yöntemlerde $M_b$ eğitim modelinin oluşturulması .....	27
Şekil 4.1. Orijinal veri kümelerine ağaç tabanlı yöntemlerin uygulanması ile elde edilen MAPE sonuçlarına göre $r_i^j$ değerlerinin kutu bilyık diyagramı .....	39
Şekil 4.2. Orijinal veri kümelerine ağaç tabanlı yöntemlerin uygulanması ile elde edilen RMSE sonuçlarına göre $r_i^j$ değerlerinin kutu bilyık diyagramı .....	40
Şekil 4.3. Aykırı değer analizi yapılan veri kümelerine ağaç tabanlı yöntemlerin uygulanması ile elde edilen MAPE sonuçlarına göre $r_i^j$ değerlerinin kutu bilyık diyagramı .....	40
Şekil 4.4. Aykırı değer analizi yapılan veri kümelerine ağaç tabanlı yöntemlerin uygulanması ile elde edilen RMSE sonuçlarına göre $r_i^j$ değerlerinin kutu bilyık diyagramı .....	41
Şekil 4.5. Orijinal veri kümelerinin MAPE sonuçlarına göre belirlenen karşılaştırma modellerinin $r_i^j$ değerlerinin kutu bilyık diyagramı.....	45
Şekil 4.6. Orijinal veri kümelerinin RMSE sonuçlarına göre belirlenen karşılaştırma modellerinin $r_i^j$ değerlerinin kutu bilyık diyagramı.....	45
Şekil 4.7. Aykırı değer analizi yapılan veri kümelerinin MAPE sonuçlarına göre belirlenen karşılaştırma modellerinin $r_i^j$ değerlerinin kutu bilyık diyagramı .....	45
Şekil 4.8. Aykırı değer analizi yapılan veri kümelerinin RMSE sonuçlarına göre belirlenen karşılaştırma modellerinin $r_i^j$ değerlerinin kutu bilyık diyagramı .....	46
Şekil 4.9. Firmada üretilen sac şekillendirme kalıplarına ait parça örnekleri .....	48
Şekil 4.10. Sac şekillendirme kalıp örnekleri .....	49
Şekil 4.11. Temel kalıp imalat süreci .....	49
Şekil 4.12. Kalıp üretim süresine etki eden faktörlerin kategorilere ayrılması .....	52

## ÇİZELGELER DİZİNİ

### Sayfa

Çizelge 4.1. Karşılaştırmalı değerlendirme veri kümelerinin özellikleri.....	32
Çizelge 4.2. Orijinal veri kümelerine ağacı tabanlı yöntemlerin uygulanması ile elde edilen MAPE sonuçları (%).....	34
Çizelge 4.3. Orijinal veri kümelerine ağacı tabanlı yöntemlerin uygulanması ile elde edilen RMSE sonuçları .....	35
Çizelge 4.4. Aykırı değer analizi yapılan veri kümelerine ağacı tabanlı yöntemlerin uygulanması ile elde edilen MAPE sonuçları (%).....	36
Çizelge 4.5. Aykırı değer analizi yapılan kümelerine ağacı tabanlı yöntemlerin uygulanması ile elde edilen RMSE sonuçları .....	37
Çizelge 4.6. Orijinal veri kümeleri için Friedman test sonuçları .....	42
Çizelge 4.7. Aykırı değer analizi yapılan veri kümeleri için Friedman test sonuçları....	42
Çizelge 4.8. Orijinal veri kümelerinde ağaç ve komşuluk tabanlı yöntemler için MAPE ve RMSE sonuçlarına göre seçilen modeller .....	43
Çizelge 4.9. Aykırı değer analizi yapılan veri kümelerinde ağaç ve komşuluk tabanlı yöntemler için MAPE ve RMSE sonuçlarına göre seçilen modeller.....	44
Çizelge 4.10. Model ağacı, ağaçlı tabanlı yöntemler ve komşuluk tabanlı yöntemlerin birlikte değerlendirilmesi ile en iyi performansla sahip yöntemlerin listesi .....	46
Çizelge 4.11. <i>Kalıp</i> veri kümesinin ağaç tabanlı yöntemler için MAPE sonuçları (%)..	53
Çizelge 4.12. <i>Kalıp</i> veri kümesinin ağaç tabanlı yöntemler için RMSE sonuçları.....	53
Çizelge 4.13. Öznitelik seçimi sonrası <i>Kalıp-2</i> veri kümesinin ağaç tabanlı yöntemler için MAPE sonuçları (%).....	54
Çizelge 4.14. Öznitelik seçimi sonrası <i>Kalıp-2</i> veri kümesinin ağaç tabanlı yöntemler için RMSE sonuçları .....	54
Çizelge 4.15. Model ağacı, ağaç tabanlı yöntemler ve komşuluk tabanlı yöntemlerin vaka çalışması için karşılaştırılması .....	54

## 1. GİRİŞ

Endüstri 4.0 ve bilgi teknolojilerindeki gelişmeler büyük miktarda verinin üretim veya hizmet ortamlarından toplanması, depolanması ve kolay erişimi için şirketlere avantaj sağlamaktadır. Toplanan veriler işlenmediğinde verinin gerçek önemi ve değeri bilinmemektedir. Veri madenciliği, büyük miktarda verinin içerisinden anlamlı, gizli kalmış bilgileri ortaya çıkartmak için çeşitli yöntemlerle verinin işlenmesidir. Bu süreçte veri madenciliği istatistik, veri tabanı teknolojileri, makine öğrenmesi, yapay zekâ ve görselleştirme gibi birçok farklı disiplini bir araya getirmektedir. Veri madenciliği ile elde edilen bilgiler iş yöntemi, pazarlama, bankacılık, tıp, mühendislik, satış tahmini gibi birçok alanda yaygın olarak kullanılmaktadır.

Talep tahmini, rekabet analizi, fiyatlandırma, kapasite tahmini, üretim süresi tahmini, müşteri kredi limitlerinin belirlenmesi gibi birçok alanda numerik çıktı değerleri tahmin edilmek istenmektedir. Üretim ve hizmet sistemlerinde numerik çıktı değerlerinin doğru ve hızlı tahmin edilmesi büyük bir problemdir. Literatüre bakıldığında bu problem regresyon veya öngörü problemi olarak adlandırılmaktadır. Regresyon problemlerinin çözümünde yaygın olarak kullanılan yöntemler regresyon ağaçları, destek vektör regresyonları, yapay sinir ağları ve en yakın komşuluk tabanlı yöntemlerdir.

Bu tezde tahmin problemlerinin çözümü için veri madenciliği tabanlı bir tahmin metodolojisi geliştirmek amaçlanmıştır. Veri madenciliği doğrusal olmayan, aykırı değerler içeren büyük miktarda verilerle başa çıkabildiği için tahmin problemlerinin çözümünde tercih edilmektedir.

Bu tezde veri madenciliği yöntemlerinden regresyon ağaçları ve komşuluk tabanlı yöntemler esas alınarak çalışmalar yapılmıştır. Regresyon ağaçları anlaşılması ve yorumlanması kolay, veri ön işleme ihtiyacı az, sayısal ve kategorik veri işleyebilen ve gürültüden etkilenmeyen bir yöntemdir. Regresyon ağaçları bir kural mekanizması oluşturduğu için tahmin değerlerinin hangi koşullar altında üretildiği kolaylıkla açıklanmaktadır. Ancak regresyon ağaçlarının bazı dezavantajları da vardır.

Regresyon ağaçlarının yaprak düğümünde kullanılan tahmin yönteminin seçimi önemli bir problemdir. Karmaşıklığı yüksek tahmin fonksiyonlarının kullanılması hesaplama maliyetini arttırırken, ortalamaların alınması gibi basit bir yöntem seçilmesi tahmin doğruluğunun düşük olmasına neden olmaktadır. Bu problemin çözümü için  $K$ -en yakın komşu yönteminin basit ama etkili yapısını regresyon ağaçlarına entegre eden bir yaklaşım önerilmiştir. Önerilen yaklaşımda en yakın komşulara daha fazla ağırlık verilirken, nesnelerin yerel aykırı değer faktörleri de bir diğer ağırlık kriteri olarak dikkate alınmaktadır. Yerel aykırı değer faktörü ile komşuluk tabanlı algoritmalarda veri kümesindeki aykırı değerlerin tahmin sonuçlarına olan olumsuz etkisinin azaltılması amaçlanmıştır. Bununla birlikte bir diğer zorluk en yakın komşu algoritmasında  $K$  değerinin belirlenmesidir. Bunun için önerilen yaklaşımda farklı  $K$  değerleri kullanılmıştır.

Literatüre bakıldığında önerilen yaklaşıma benzer çalışmalar bulunmaktadır, ancak regresyon ağaçlarının yaprak düğümlerine nesnelerin komşuluk ve aykırı değer faktörlerini entegre eden bir çalışmaya rastlanmamıştır. Aynı zamanda yapılan çalışmalardan farklı olarak torbalama ve güçlendirme algoritmaları önerilen komşuluk tabanlı regresyon ağaçları ile birlikte kullanılmıştır.

Geliştirilen yaklaşımlar karşılaştırmalı değerlendirme için kullanılan veri kümelerine uygulanarak elde edilen sonuçlara göre model performansları karşılaştırılmıştır. Ayrıca siparişe dayalı imalat yapan bir kalıp firmasının gerçek verileri üzerinde bir vaka çalışması yapılmıştır. Yapılan vaka çalışmasında önerilen yöntemler ile kalıp imalat süresi tahmin edilmiştir.



## 2. KAYNAK ARAŞTIRMASI

Bu bölümde, bu tezde kullanılan veri madenciliği yöntemleri ile ilgili çalışmalar incelenmiştir. Bölüm 2.1’de regresyon ağaçlarındaki gelişmeler açıklanmıştır. Bölüm 2.2’de *K*-en yakın komşu algoritması ile ilgili yapılan çalışmalara yer verilmiştir. Bölüm 2.3’te literatürde yer alan temel topluluk yöntemler açıklanmıştır. Son olarak Bölüm 2.4’te aykırı veri analizi ile ilgili çalışmalar yer almaktadır.

### 2.1. Regresyon Ağaçları

Regresyon ağaçları, bir tahmin modeli oluşturmak için verileri yinelemeli olarak bölümlere ayıran ve her bölüme basit bir tahmin modeli yerleştiren bir makine öğrenimi yöntemidir (Loh 2014). Regresyon ağaçları bölünme sayısı, ağaç derinliği, budama koşulu, hata hesabı ve her bölüme uygulanan regresyon modeli açısından birbirinden farklılaşmaktadır.

Ağaç tabanlı yaklaşımların temelini oluşturan ve tarihte bilinen ilk regresyon ağacı 1960'larda Morgan ve Sonquist (1963) tarafından geliştirilen Otomatik Etkileşim Algılama (Automatic Interaction Detection – AID) algoritmasıdır. AID algoritmasında ağaç ikili bölünmelerle oluşturulmaktadır. Bir düğümün bölünme sonrası hata miktarındaki iyileşme 0,01’den küçükse bölünmeyi durdurmaktadır. Daha sonra 1970’lerde Morgan ve Messenger (1973), sınıflandırma ağacı üretmek için AID algoritması kullanarak THeta Otomatik Etkileşim Algılama (THeta Automatic Interaction Detection – THAID) algoritmasını geliştirmiştir. THAID algoritmasında bölünme düğüm safsızlığına göre yapılmaktadır ve durdurma kriteri AID algoritmasında olduğu gibi sabit bir değer değildir. Başlangıçta sınıflandırma için tasarlanmış ve daha sonra regresyona genişletilmiş bir diğer algoritma Ki-kare Otomatik Etkileşim Algılama (CHi-squared Automatic Interaction Detector – CHAID) algoritmasıdır (Kass 1980). CHAID algoritması ki-kare analizi yaparak, değişkenin türüne bağlı olarak bir düğümü iki veya daha fazla çocuk düğüme bölmektedir. 1980’lerde Breiman ve ark. (1984), AID ve THAID ile aynı ağaçlı arama yaklaşımını izleyerek maksimum ağacı oluşturduktan sonra ağacı en düşük çapraz doğrulama hata tahminine göre budayan Sınıflandırma ve

Regresyon Ağaçları (Classification and Regression Trees – CART) algoritmasını geliştirmiştir. Quinlan (1992) CART gibi ağaç temelli bir model olan M5 algoritmasını geliştirmiştir. CART regresyon ağaçlarının yapraklarında değerler varken, M5 tarafından oluşturulan ağaçların yapraklarında çok değişkenli doğrusal modeller olması aralarındaki en büyük farktır. M5 ağacının dezavantajı parçalı sabit modeller oldukları için başlangıçta oldukça büyük olma eğiliminde olmasıdır. Wang ve Witten (1996), M5 algoritmasını daha az ayrıntılı ama daha hızlı bir yaklaşım kullanarak parçalı çoklu doğrusal ağaç modeli algoritmasına genişletmiştir.

Chaudhuri ve ark. (1994) regresyon ağacı oluştururken her düğümde verilere parametrik olmayan doğrusal bir model uygulayan Düzleştirilmiş ve Düzleştirilmemiş Parçalı Polinom Regresyon Ağaçları (Smoothed and Unsmoothed Piecewise Polynomial Regression Trees – SUPPORT) algoritmasını geliştirmiştir. SUPPORT algoritmasının ağaç boyutunu belirlemek için çapraz doğrulayıcı çok adımlı ileriye dönük durma kriteri kullanılması CART algoritmasından farklılaştığı noktalardan biridir. Ayrıca CART, artık kareler toplamındaki (Residual Sum of Squares – RSS) azalma derecesine göre bölünmeleri seçerken, SUPPORT artık dağılımlarının analizi ile bölünmeleri seçmektedir. Brown ve ark. (1996), CART algoritmanın öznitelikleri doğrusal bölmesi ile ilgili sorunları incelemiştir ve ardından her düğümde optimum çok değişkenli bölünmeyi kullanan bir doğrusal programlama çözümü önermiştir. CART algoritmasını parçalı doğrusal modellere genişletilmesinin pratikte çok zaman alıcı olmasından dolayı Alexander ve Grimshaw (1996) hesaplama miktarını azaltmak için her bir düğümde basit bir doğrusal regresyon modelinin yerleştirilmesini önermiştir. Torgo (1997) ise yaprak düğümlerdeki doğrusal modellere ek olarak çekirdek regresyonu ve en yakın komşu modelini kullanan bir regresyon ağacı önermiştir.

Loh ve Shih (1997), Hızlı ve Doğru Sınıflandırma Ağacı (Fast and Accurate Classification Tree – FACT) (Loh ve Vanichsetakul 1988) algoritmasından yola çıkarak her düğümü bölmek için anlamlılık testi kullanan iki adımlı bir yaklaşım olan Hızlı, Tarafsız ve Etkili İstatistik Ağacı (Quick, Unbiased and Efficient Statistical Tree – QUEST) algoritmasını geliştirmiştir. QUEST, sırasız değişkenler için ki kare testlerini ve sıralı değişkenler için varyans analizi (Analysis of Variance – ANOVA) testlerini

kullanmaktadır. Hothorn ve ark. (2004) regresyon ağaçlarını, koşullu çıkarım prosedürünün iyi tanımlanmış bir teorisine yerleştiren özyinelemeli bölünme için Koşullu Çıkarım Ağaçları (Conditional Inference Trees – CTREE) olarak adlandırılan bir yaklaşım geliştirmiştir. CTREE tarafsız parçalı sabit regresyon ağaçları oluşturmak için permütasyon testlerini kullanmaktadır. Su ve ark. (2004), parçalı sabit regresyon modelinde bölünme noktalarını seçmek için maksimum olasılık kullanarak CART algoritmasını genişletmiştir. Dusseldorp ve Meulman (2004), regresyon ağaçlarının ve çoklu doğrusal regresyon analizinin bir entegrasyonu olan Regresyon Gövde Yaklaşımını (Regression Trunk Approach – RTA) önermiştir. RTA'dan daha genel ve daha verimli bir regresyon gövde yaklaşımını tahmin etmek için Eşzamanlı Eşik Etkileşim Modelleme Algoritması (Simultaneous Threshold Interaction Modeling Algorithm – STIMA) geliştirilmiştir (Dusseldorp ve ark. 2010).

QUEST (Loh ve Shih 1997) ve Tarafsız Etkileşim Seçimi ve Tahmini ile Sınıflandırma Kuralı (Classification Rule with Unbiased Interaction Selection and Estimation – CRUISE) (Kim ve Loh 2001) algoritmalarının güçlü yanları dikkate alınarak ve zayıf yönleri iyileştirilerek Loh (2009) tarafından Genelleştirilmiş, Tarafsız, Etkileşim Algılama ve Tahmin (Generalized, Unbiased, Interaction Detection and Estimation – GUIDE) yaklaşımı geliştirilmiştir. GUIDE algoritmasında dört teknik kullanılarak ağaçlar oluşturulmaktadır. Bu teknikler parçalı sabit (piecewise constant), en iyi basit doğrusal (best simple linear), en iyi basit ikinci dereceden regresyon (best simple quadratic regression) ve aşamalı doğrusaldır (stepwise linear).

Loh ve ark. (2015), GUIDE algoritmasını üç temel fikir kullanarak genişletmiştir. Bu fikirler doğrusal bir tahmin edici, ki-kare testi ve poisson regresyonudur. Ayrıca her düğüme güven aralığı oluşturmak için önyükleme tekniği kullanmıştır. Czajkowski ve Kretowski (2016) analiz edilen verilere kendi kendine adapte olabilen bir yapıya sahip karar ağacı indüksiyonu için yeni bir evrimsel algoritma önermiştir. Painsky ve Rosset (2016) CART algoritmasında her bir bölme değişkenini seçmek için Birini Dışarıda Bırak Çapraz Geçerleme (Leave One Out Cross Validation – LOOCV) kullanmıştır, ardından seçilen değişken için mevcut bölünme kurallarını uygulamıştır. Johansson ve ark. (2018), regresyon ağaçlarında iki yeni yaklaşım ele almıştır. İlk yaklaşım, her yapraktaki hedef

değerlerin standart sapması ile zorluğu tahmin etmektedir. İkinci yaklaşım, her kuralın (kök düğümünden yaprak düğümüne giden yol) bağımsız olarak geçerli olduğu regresyon ağaçları ile sonuçlanan Mondrian konformal tahminini kullanmaktadır. Diao ve Weng (2019) klasik Bühlmann-Straub güvenilirlik formülünü CART algoritmasına uygulayarak, Güvenilirlik Regresyon Ağacı (Regression Tree Credibility – RTC) algoritmasını geliştirmiştir. Su ve Ding (2019), farklı değişken grupları arasında doğrusal olmayan bir etkileşim olmayacak şekilde değişkenleri potansiyel gruplarına ayırarak Değişken Gruplama Tabanlı Bayes Katkı Regresyon Ağacı (Variable Grouping Based Bayesian Additive Regression Tree – GBART) algoritmasını önermiştir. Nourali ve Osanloo (2020), bir maliyet tahmin modeli geliştirmek için regresyon ağacından yararlanmıştır.

Sutton (2005), sınıflandırma ve regresyon ağaçları ile ilgili genel bilgiler verdikten sonra, yöntemin kökeni, gelişimini ve yöntem adımlarını açıklamaktadır. Ayrıca yaptığı çalışmada torbalama ve güçlendirmeye yer vermektedir. Loh (2011) sınıflandırma ve regresyon problemlerinin çözümü için yaygın olarak kullanılan C4.5, CART, CHAID, CRUISE, GUIDE, QUEST ve M5 algoritmalarını gözden geçirmiştir ve bu algoritmaların yeteneklerini, güçlü zayıf yönlerini iki örnek üzerinde karşılaştırmıştır. Alberg ve ark. (2012) veri akış madenciliği için regresyon ağacı yöntemlerinin gelişmiş bir incelemesini yapmıştır. Loh (2014) sınıflandırma ve regresyon ağaçları ile ilgili son elli yılda yapılan çalışmaları incelemiştir. Çalışmasında AID ve CART algoritmalarının parçalı sabit regresyon ağaçları olduğunu belirtmiş ve sonraki gelişmeleri iki kola ayırmıştır. Bu iki koldan biri parçalı doğrusal (piecewise linear) veya daha yüksek dereceden en küçük kareler (higher order least squares) modellerini içermektedir. İkinci kolda ise diğer kayıp fonksiyonları ile parçalı sabit veya doğrusal (piecewise constant or linear) modeller yer almaktadır.

## **2.2. En Yakın Komşuluk Tabanlı Yöntemler**

*K*-en yakın komşu (*K*-nearest neighbor – KNN), örüntü sınıflandırma ve regresyon modelleri için en eski, en basit ve doğru algoritmalarından biridir. KNN, ilk olarak Evelyn ve Hodges (1951) tarafından önerilmiştir. Daha sonra Cover ve Hart (1967) tarafından

değiştirilen KNN, veri madenciliğinde ilk 10 yöntemden biri olarak belirlenmiştir (Kumar ve ark. 2008).

KNN algoritması etkili bir yöntem olsa da yapılan çalışmalarda birçok zayıf yönü olduğu görülmüştür. Wettschereck ve Dietterich (1994) sınıflandırma problemleri için en uygun  $K$  değerini seçen yerel uyarlanabilir  $K$  en yakın komşuluk (Locally adaptive KNN) algoritması geliştirmiştir. Han ve ark. (2001) geliştirdikleri ağırlık ayarlı KNN algoritmasında en yakın komşuları belirlemek için hesaplanan mesafelerin, ağırlık olarak kullanılabilir benzer ölçülere dönüştürülmesi gerektiğini önermiştir. Song ve ark. (2007) nesnelere arasındaki yakınlığı ölçmek için sorgu tabanlı bir mesafe ölçütü kullanarak Yerel Bilgilendirici-KNN (Locally Informative-KNN – LI-KNN) ve Küresel Bilgilendirici-KNN (Globally Informative-KNN – GI-KNN) olarak isimlendirilen iki yeni KNN yaklaşımı önermiştir.

KNN, sürekli çıktı değişkenine sahip veri kümeleri için tahmin problemlerini çözmek için de kullanılmaktadır. Dudani (1976) bir tahmin nesnesinin komşuları ile arasındaki mesafeleri, komşuluğundaki her bir nesnenin ağırlığı olarak kullanmıştır. Böylece tahmin nesnesinin uzak komşularının çıktı değerine katkısı az olurken, yakın komşuların katkısı daha fazla olmaktadır. Biau ve ark. (2012), Guyader ve Hengartner (2013) ile Lee ve ark. (2014) KNN regresyon algoritması ile tahmin yaparken  $K$  komşuluğun çıktı değerlerinin ortalamasını kullanmaktadır.

Wang ve Chaib-draa (2016), Gauss süreci regresyonu için yeni bir KNN tabanlı Kalman filtresi ( $K$ -Nearest Neighbor based Kalman Filter for Gaussian Process Regression – KNN-KFGP) önermiştir. Chen ve Han (2017) optimizasyon ve özellik seçme yöntemlerinden elde edilen ağırlıkları kullanarak ağırlıklı KNN yaklaşımı geliştirerek regresyon problemlerindeki performans ve verimliliğine bakmıştır. García ve ark. (2018) regresyon problemlerinin çözümü için  $K$ -en yakın ağırlık merkezi komşuları ( $K$ -Nearest Centroid Neighbors – KNCR) yaklaşımını geliştirmiştir. Barrash ve ark. (2019) KNN regresyonunun  $K$  parametresini uyarlanabilir olarak seçilme problemini ele almıştır. Kosasih ve ark. (2020) bir tahmin modeli oluşturmak için Tensorflow ile  $K$ -En Yakın Komşu Regresyon yöntemini birlikte kullanmıştır. Li ve Guan (2020)  $K$ -en yakın komşu

algoritması ve çoklu regresyon algoritmasını birleştirerek yeni bir yaklaşım (KNN and Multiple Regression Algorithm – KMRA) geliştirmiştir.

Batista ve Silva (2009) en yakın komşu sayısı, mesafe fonksiyonu ve ağırlıklandırma fonksiyonlarının KNN algoritmasının performansı üzerindeki etkilerini incelemiştir. Bhatia (2010) içinde ağırlıklı KNN (weighted KNN) ve model tabanlı KNN (Model based KNN) algoritmalarının da yer aldığı birçok en yakın komşu (Nearest Neighbor – NN) algoritmasının avantajlarını ve dezavantajlarını incelemiştir. Parametrik olmayan bir regresyon yöntemi olan  $K$ -en yakın komşu (KNN) regresyonun dezavantajı yüksek boyutlu ve seyrek verilerle başa çıkamamasıdır (Hastie ve ark. 2001, Kramer 2011). KNN yönteminin zayıf yönlerinin ortadan kaldırılması için geliştirilen KNN versiyonları Taunk ve ark. (2019) tarafından yapılan bir çalışmada 11 başlık altında incelenmiştir. Alfeilat ve ark. (2019) KNN algoritmasının performansını etkileyen iki önemli kriterden biri olan mesafe ölçütü seçimini araştıran bir çalışma yapmıştır. Alfeilat ve ark. (2019) çalışmalarında KNN sınıflandırıcısının performansını çok sayıda gerçek dünya veri kümesini farklı gürültü seviyelerinde ve birçok farklı mesafe ölçüsü kullanarak değerlendirmiştir.

En yakın komşuluk algoritması ile topluluk yöntemlerin birlikte de kullanıldığı çalışmalar vardır. Bay (1998) NN sınıfının doğruluğunu artırmak için tasarlanmış bir birleştirme algoritması olan çoklu özellik alt kümeleri (Multiple Feature Subsets – MFS) kullanmıştır. Domeniconi ve Yan (2004) verilerin yüksek boyutlu olmasından yararlanarak bir NN sınıflandırıcı seti oluşturmak için farklı özellik seçeneklerini kullanan en yakın komşu topluluğu oluşturmuştur. Haixiang ve ark. (2016) özellik seçimi, güçlendirme ve yeni bir değerlendirme metriğini birleştirerek BPSO-Adaboost-KNN adlı bir topluluk algoritması geliştirmiştir. Farrelly (2017) KNN regresyon yöntemlerini iyileştirmek için değişen  $K$  değerine ve torbalamaya dayalı bir dizi KNN regresyon topluluğu geliştirmeyi amaçlamıştır.

### 2.3. Topluluk Yöntemler

Torbalama (bagging) (Breiman 1996a,b), güçlendirme (boosting) (Schapire 1990, Freund ve Schapire 1996, 1997) ve rastgele orman (random forest) (Breiman 2001), tahmin edicinin genel doğruluğunu artırmak için farklı tahmin edicilerden gelen çıktıları birleştiren popüler topluluk yöntemidir.

Breiman (1996a,b) tahmin varyansını azaltmak ve tahmin sürecini iyileştirmek için birçok sınıflandırma ve regresyon yöntemi ile kullanılabilen bir topluluk yöntem olan torbalama yaklaşımını ortaya çıkartmıştır.

Orijinal güçlendirme yaklaşımı ise Schapire (1990) tarafından geliştirilen filtreleme yoluyla güçlendirilmiştir. Filtreleme yoluyla güçlendirme yaklaşımı gerçek hayat verilerinde mümkün olmayacak kadar çok sayıda eğitim örneği gerektirmektedir. Bu dezavantaj bir başka güçlendirme algoritması olan AdaBoost algoritmasının Freund ve Schapire (1996) tarafından geliştirilmesiyle giderilmiştir. Freund ve Schapire (1996) yaptıkları çalışmada AdaBoost.M1 ve AdaBoost.M2 olarak adlandırılan Adaboost algoritmasının iki versiyonunu açıklamıştır. Geliştirdikleri bu iki algoritmayı torbalama (Breiman 1994) ve bir Optik Karakter Tanıma (Optical Character Recognition – OCR) probleminde en yakın komşu sınıflandırıcısını kullanarak karşılaştırmıştır. Freund ve Schapire (1997) yeni bir güçlendirme algoritması elde etmek için çarpımsal ağırlık güncelleme tekniğini (multiplicative weight-update technique) kullanılmıştır. Yapılan bu çalışmada Adaboost.M2 algoritması regresyon problemlerine güçlendirme uygulamak için genişletilerek Adaboost.R algoritması ortaya çıkmıştır. Adaboost.R algoritması incelendiğinde regresyon problemlerinin sınıflandırma problemlerine indirgeyerek çözüm elde edildiği görülmüştür. Breiman (1997), regresyon problemleri için ark oyun değeri (arcing game value) algoritması önermiştir. Drucker (1997), AdaBoost.R'nin geçici bir modifikasyonu olan AdaBoost.R2 algoritmasını geliştirmiştir. Avnimelech ve Intrator (1999), zayıf ve güçlü öğrenme kavramını ve aralarındaki uygun denkliği tanıtarak güçlendirme algoritmasını regresyon problemlerine genişletmiştir. Bu çalışmadan yararlanan Feely (2000) büyük hata marjlı (Big Error Margin – BEM) güçlendirme tekniğini geliştirmiştir. Namee ve ark. (2000) basit bir yapay sinir ağı tahmin

ediciğine AdaBoost.R2 ve BEM algoritmalarını entegre ederek bu iki algoritmayı karşılaştırmıştır. Friedman ve ark. (2000) AdaBoost algoritmasının belirli bir üstel kayıp fonksiyonunu optimize etmek için bir Newton yöntemi olduğunu göstererek Adaboost algoritmasını istatistiksel olarak ele almıştır. Zemel ve Elman (2001) ise sadece eğitim örneklerinin dağılımını değiştirerek yeni hipotezler oluşturan gradyan tabanlı bir güçlendirme algoritması geliştirmiştir. Shrestha ve Solomatine (2006) önceden ayarlanmış eşik değerinden daha yüksek olan göreceli tahmin hatası ile örnekleri filtreleyerek ve ardından AdaBoost algoritmasını uygulayarak regresyon problemleri için yeni bir güçlendirme algoritması geliştirmiştir.

Quinlan (1996) torbalama ve güçlendirme algoritmalarını C4.5 algoritmasına uygulamıştır ve her iki yaklaşımın tahmin doğruluğunu iyileştirdiği ancak güçlendirmenin, torbalamaya göre daha fazla iyileşme sağladığını tespit etmiştir.

Opitz ve Maclin (1999) hem sinir ağlarını hem de karar ağaçlarını kullanarak yirmi üç veri kümesini torbalama ve güçlendirme algoritmalarını sınıflandırmak için kullanmıştır. İlk olarak torbalama algoritmasının neredeyse her zaman tek sınıflandırıcıya göre daha doğru, ancak bazen güçlendirme algoritmasına göre daha az doğru olduğunu belirlemiştir. Opitz ve Maclin (1999) elde ettiği bir diğer sonuç ise güçlendirme algoritmasının, özellikle sinir ağı kullanıldığında, tek sınıflandırıcıdan daha az doğru olan topluluklar oluşturabilmesidir.

#### **2.4. Aykırı Veri Analizi**

Aykırı değer, farklı bir mekanizma tarafından üretildiğine dair şüpheler uyandıracak kadar diğer gözlemlerden çok sapan bir gözlemdir (Hawkins 1980). Aggarwal (2015) aykırı değeri, kalan verilerden önemli ölçüde farklı olan bir veri noktası olarak tanımlamıştır. Günümüze kadar aykırı değer analizi detaylı olarak araştırılmış bir problemdir. Aykırı değer analizi algoritmaları mesafe tabanlı algoritmalar ve yoğunluk tabanlı algoritmalar olarak iki gruba ayrılabilir.



Knorr ve Ng (1998) büyük ve çok boyutlu veri kümelerinde aykırı değerleri belirlemek için mesafeye dayalı aykırı değer analizi üzerine çalışma yapmıştır. Mesafeye dayalı aykırı değer algılama algoritmalarının, yoğunluk tabanlı algoritmalarla karşılaştırıldığında doğal bir sınırlaması vardır. Birkaç küme içeren veri kümeleri üzerinde çalışamazlar. Breuning ve ark. (2000) her nesneye bir aykırılık derecesinin atandığı bir çalışma yapmıştır. Bu derece, bir nesnenin yerel aykırı değer faktörü (Local Outlier Factor – LOF) olarak adlandırmıştır.

Bakar ve ark. (2006), veri madenciliğinde aykırı değer tespiti için kontrol şeması, doğrusal regresyon ve Manhattan mesafe tekniklerinin performansını karşılaştırmıştır. Rana ve ark. (2012) mevcut aykırı değer tekniklerini ve gürültülü verilerin mevcut olduğu uygulamaları içeren bir çalışma yapmıştır. Chauhan ve Shukla (2015) veri kümesini kümelemek için K-Ortalamlar algoritmasında kullanılan farklı aykırı değer tespit yaklaşımlarını gözden geçirmiştir ve aykırı değer tespitinin farklı uygulama alanlarını tartışmıştır. Domingues ve ark. (2015) aykırı değer analizi için denetimsiz makine öğrenimi algoritmalarını araştırmıştır. Yao ve ark. (2018) veri akışındaki yerel aykırı değeri dinamik olarak değerlendirmek için yeni bir artan yerel aykırı değer algılama yaklaşımı geliştirmiştir. You ve ark. (2020) istenmeyen e-posta incelemelerini tanımlamak için bir Boyut Derecelendirme Yerel Aykırı Faktör Modeli (Aspect-Rating Local Outlier Factor – AR-LOF) önermiştir.

### 3. MATERYAL ve YÖNTEM

Bu bölümde, tezde kullanılan veri kümeleri ve veri madenciliği yöntemleri açıklanmaktadır. Bölüm 3.1’de tezde kullanılan materyal hakkında bilgi verilmektedir. Bölüm 3.2 yüksek doğrulukta bir tahmin metodolojisi geliştirmek için kullanılan yöntemleri açıklamaktadır.

#### 3.1. Materyal

Veri kümesi, nesnelere ve her bir nesnenin özelliğini temsil eden özniteliklerden oluşan bir kümedir.  $D = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\} \in (x \times y)^n$  bağımsız ve aynı şekilde dağıtılan rastgele çiftlerden  $(\mathbf{x}_i, y_i)$  oluşan bir veri kümesi olsun, burada  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{im}]$  vektörü  $m$  boyutlu bir öznitelik uzayındaki bir örneği temsil eder ve  $y_i$  kendisiyle ilişkili sürekli çıktı değerini belirtir. Regresyonun amacı  $f : \mathbf{X} \rightarrow y$  fonksiyonunu öğrenerek yeni bir örnek için  $y$  değerini belirlemektir  $\mathbf{X} = [X_1, X_2, \dots, X_m]$ .

Bir veri kümesinde yer alan öznitelikler sınıfsal (nominal), ikili (binary), tamsayı (integer), sıralı (ordinal), aralık ölçekli (interval-scaled) ve oran ölçekli (ratio-scaled) olarak türlere ayrılır (Bramer 2007). Sınıfsal, ikili ve sıralı öznitelikler kategorik özniteliklerdir. Tamsayı, aralık ölçekli ve oran ölçekli öznitelikler ise sürekli özniteliklerdir. Veri kümesinde kategorik veya sürekli özniteliklerin olması kullanılacak yöntemin seçiminde oldukça önemlidir. Çıktı değişkeni, kategorik değişken olan veri kümelerinde sınıflandırma yöntemleri kullanılırken, çıktı değişkeni sürekli değişken olan veri kümelerinde sayısal öngörü yöntemleri kullanılmaktadır (Han ve Kamber 2011).

Bu tezde çıktı değişkeni sürekli değişken değerine sahip, karma öznitelik türlerini içeren veri kümeleri üzerinde çalışmalar yapılmıştır. Bu sebeple veri madenciliği ile tahmin modellerinin oluşturulmasında en yaygın ve etkin olarak kullanılan yöntemlerden ağaç tabanlı yöntemler ve komşuluk tabanlı yöntemler üzerinde çalışmalar yapılmıştır.

## 3.2. Yöntem

Bu tezde kullanılan yöntemler aşağıdaki alt bölümlerde açıklanmıştır. Bölüm 3.2.1’de veri madenciliği tahmin yöntemlerinden regresyon ağacı açıklanmaktadır. Bölüm 3.2.2, sürekli çıktı değişkeni tahmini için kullanılan  $K$ -en yakın komşu algoritması sunulmaktadır. Bölüm 3.2.3’te benzeme ölçütü tanımlanmaktadır. Bölüm 3.2.4, tahmin algoritmalarının doğruluğunu ölçmek için kullanılan performans değerlendirme yöntemlerini açıklamaktadır. Bölüm 3.2.5, topluluk yöntemlerden torbalama ve güçlendirme algoritmalarını tanıtmaktadır. Bölüm 3.2.6, aykırı değer analizi için kullanılan yaklaşımı açıklamaktadır. Son olarak bölüm 3.2.7, tez kapsamında geliştirilen yaklaşımı açıklamaktadır.

### 3.2.1. Regresyon ağacı

Regresyon ağacı parametrik olmayan kural tabanlı bir yaklaşımdır. Regresyon ağaçları düğümlerin dallanma sayısı, öznitelik seçim yöntemi, durdurma kriteri veya yaprak düğümlerde kullanılan tahmin modellerine göre birbirinden farklılaşsa da temel olarak hepsi bir karar ağacı yapısındadır.  $D_b$  eğitim veri kümesinin kök düğümüne atanması ile ağaç oluşturulmaya başlanır. Öncelikle  $D_b$  eğitim veri kümesindeki her bir  $X_i$  özniteliğinin her bir bölünme noktası belirlenir.  $D_b$  eğitim veri kümesi,  $X_i$  özniteliğinin her bir bölünme noktasına göre bölünerek, alt düğümler oluşturulur. Oluşturulan alt düğümlere göre modelin hata miktarı hesaplanır.  $X_i$  özniteliğini en iyi bölen bölünme noktası seçilir. Ardından tüm öznitelikler için modelin hata miktarını en aza indiren  $X^*$  özniteliği seçilir.  $X^*$  özniteliğinin en iyi bölünme noktasına göre ağaç dallara ayrılır. Maksimum ağaç derinliği, bir düğümde bölme için ele alınan minimum eleman sayısı ve yeni bir düğümde olması gereken minimum eleman sayısı gibi durdurma kriteri sağlanana kadar oluşan alt düğümler için bölünme işlemi tekrarlanır. Durdurma kriterine ulaşan düğümler yaprak düğüm olarak belirlenir. Bir ağaç oluşturulduktan sonra, istenmeyen alt ağaçlar veya düğümler bulunabilir. Şekil 3.1’de AID Regresyon ağacı algoritmasının temel adımları Morgan ve Sonquist (1963) dikkate alınarak verilmiştir.

---

**Algoritma 1:** Regresyon ağacı

---

**Girdi:**  $D_t$  eğitim veri kümesi, durdurma kriteri

**Çıktı:** Regresyon ağacı

**Yöntemin adımları:**

- 1) Kök düğümünden başla.
  - 2) **While** durdurma kriteri sağlanmıyorsa **Do**
  - 3)     **For**  $i = 1$  **to**  $m$  **Do** // her bir düğüm için
  - 4)         Her bir  $X_i$  özneliği için, çocuk düğümlerdeki hata miktarları toplamını en aza indiren bölünme noktasını belirle.
  - 5)     **End For**
  - 6)     Minimum hata miktarına sahip olan  $X^*$  özneliğe göre düğümün en iyi bölünme noktasını belirle.
  - 7)     Düğümü iki çocuk düğüme böl.
  - 8) **End While**  
Her çocuk düğüm için 2. adıma geri dön.
- 

**Şekil 3.1.** Regresyon ağacı algoritmasının temel adımları

### 3.2.2. K-en yakın komşu yöntemi

$K$ -en yakın komşu (KNN) yönteminin temel yaklaşımı, tahmin edilecek nesneye en yakın (benzer) nesnelerin bulunmasıdır. KNN yönteminin temel unsurları eğitim veri kümesi, benzerlik ölçüsü ve  $K$  en yakın komşu sayısıdır. Bir  $x^T$  test nesnesi KNN yöntemi ile tahmin edilmek istendiğinde öncelikle verilen  $D_b$  eğitim veri kümesindeki her bir nesne ile  $x^T$  test nesnesi arasındaki uzaklıklar hesaplanır.  $x^T$  test nesnesine en yakın  $K$  tane eğitim nesnesi belirlenir. Belirlenen  $K$  tane eğitim nesnesinin çıktı değişkenlerinin ortalaması alınır. KNN algoritmasının temel adımları Şekil 3.2’de verilmiştir.

KNN algoritması test veri kümesindeki her nesne ile eğitim veri kümesindeki her bir nesne arasındaki uzaklığı hesapladığı için gerçek zamanlı uygulamalarda ve eğitim veri kümesinin büyük olduğu durumlarda verimsiz bellek kullanımı problemine neden olmaktadır. KNN algoritmasında dikkat edilmesi gerek bir diğer konu benzerlik ölçüsünün seçimidir. Nesneler arasındaki uzaklıklar hesaplanmadan önce veri kümesine normalizasyon yapılmalıdır.

---

**Algoritma 2:**  $K$ -en Yakın Komşu Algoritması

---

**Girdi:**  $D_b$  eğitim veri kümesi,  $x^T$  test nesnesi,  $K$ , uzaklık ölçüsü

**Çıktı:**  $x^T$ 'nin tahmin değeri

**Yöntemin adımları:**

- 1)  $x^T$  test nesnesi ile  $D_b$  eğitim veri kümesindeki her bir nesne arasındaki uzaklığın hesapla.
- 2)  $D_b$  eğitim veri kümesinden  $x^T$ 'ye en yakın  $K$  nesneyi seç.
- 3)  $x^T$  için tahmin değeri hesapla:

$$y_{x^T} = \frac{\sum_{i=1}^K y_i}{K}$$

---

**Şekil 3.2.**  $K$ -en yakın komşu algoritmasının temel adımları

Bununla birlikte  $K$  değerinin doğru belirlenmesi model performansının doğru hesaplanması için oldukça önemlidir.  $K$  değerinin küçük seçilmesi modelin gürültüye duyarlılığını arttırırken, büyük  $K$  değeri başka sınıflardan komşuların tahmin sonuçlarına dahil edilmesine neden olmaktadır.

**3.2.3. Benzemezlik (dissimilarity) ölçüsü**

Veri madenciliğinde mesafe, nesnelerin birbirine yakın veya uzak olmasının anlamını açıklamanın somut bir yoludur (Alfeilat ve ark. 2019). İki nesne olan  $x_p$  ve  $x_q$  arasındaki uzaklık  $d(x_p, x_q)$  olarak gösterilir. Bu tezde  $L_p$  Minkowski mesafe ölçüleri ailesinden olan Öklid mesafe ölçüsü kullanılmıştır. Öklid mesafe ölçüsü Denklem 3.1'deki gibi hesaplanmaktadır.

$$d(x_p, x_q) = \sqrt{\sum_{j=1}^m |x_{pj} - x_{qj}|^2} \quad (3.1)$$

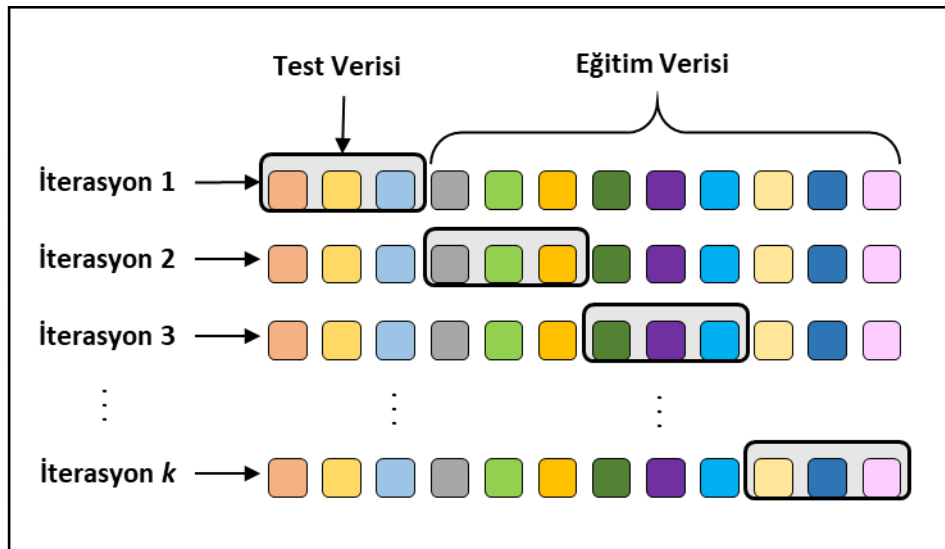
Mesafe  $[0,1]$  aralığında olduğunda, buna karşılık gelen benzerlik (similarity) ölçüsü  $s(x_p, x_q)$  Denklem 3.2'deki gibi hesaplanmaktadır.

$$s(x_p, x_q) = 1 - d(x_p, x_q) \quad (3.2)$$

### 3.2.4. Performans değerlendirme yöntemleri

Bir modelin yeni bir nesneyi doğru tahmin etme gücüne tahmin doğruluğu denir. Modelin tahmin doğruluğunu belirlemek için kullanılan birçok performans değerlendirme yöntemi bulunmaktadır. Veri kümelerini rastgele örneklenmiş bölümlere ayrılarak tahmin doğruluğunu değerlendirmeye yönelik kullanılan yaygın teknikler dışarıda tutma (holdout), rastgele alt örnekleme (random subsampling), çapraz geçерleme (cross validation) ve önyüklemedir (bootstrap). Model performansını değerlendirmek için bu tür tekniklerin kullanılması genel hesaplama süresini artırır, ancak model seçimi için faydalıdır (Han ve Kamber 2011). Bu tezde çapraz geçerleme ve önyikleme yöntemleri kullanılmıştır.

Çapraz geçerlemede, veriler rastgele olarak birbirini tekrar etmeyen her biri eşit büyüklükte  $D_1, D_2, \dots, D_k$  olarak gösterilen  $k$  alt kümeye veya “katlara” bölünür. Eğitim ve test  $k$  kez yapılır. İterasyon  $i$ 'de bölüm  $D_i$  test kümesi olarak ayrılır ve kalan bölümler modelin eğitimi için kullanılır. Modelin hata tahmini,  $k$  tekrardan gelen toplam tahmin hata miktarının başlangıçtaki toplam nesne sayısına bölünmesiyle elde edilir. Şekil 3.3'de  $k$ -kat çapraz geçerleme yöntemi şematik olarak gösterilmiştir.



Şekil 3.3.  $k$ -kat çapraz geçerleme yönteminin şematik gösterimi

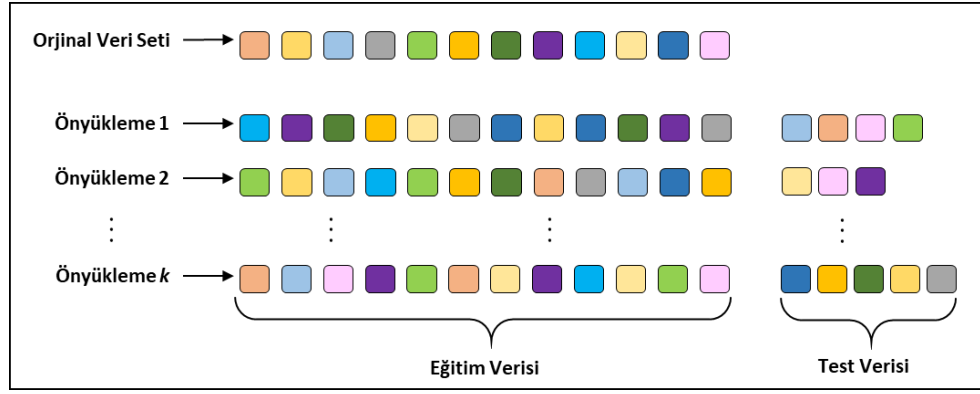
Birini dışarıda bırak çapraz geçерleme (Leave-One-Out Cross Validation – LOOCV),  $k$ 'nin başlangıçtaki nesne sayısı olarak belirlendiği özel bir  $k$ -kat çapraz geçерleme yöntemidir. LOOCV yönteminde test veri kümesi için her bir iterasyonda yalnızca bir nesne dışarıda bırakılır.

*Önyükleme* verilen eğitim nesneleri eşit olasılığa sahip diğer nesnelere ile değiştirerek örneklenir. Yani, bir nesne her seçildiğinde, tekrar seçilmesi ve eğitim kümesine yeniden eklenmesi eşit derecede olasıdır. Yaygın olarak kullanılan önyükleme yöntemlerinden biri Efron (1983), Efron ve Tibshirani (1997) tarafından geliştirilen 0.632 önyüklemesidir.

0.632 önyüklemede  $n$  nesneye sahip bir veri kümesi için tüm nesnelere eşit  $1/n$  seçilme olasılığına sahiptir ve  $n$  kez değiştirme yapılır. Değıştirme yapılırken her bir nesnenin seçilmeme olasılığı  $(1 - 1/n)$ 'dir. Önyükleme sonunda bir nesnenin seçilmeme olasılığı  $(1 - 1/n)^n$  olacaktır.  $n$  çok büyük bir sayı olduğunda her bir nesnenin seçilmeme olasılığı  $e^{-1} = 0.386$ 'ya yakınsar. Sonuç olarak orijinal veri kümelerinin %36,8'si test veri kümesini oluştururken, %63,2'si eğitim kümesini oluşturmaktadır. Önyükleme  $k$  kez tekrarlanır ve modelin doğruluğu (Accuracy of the model –  $ACC(M)$ ) aşağıdaki gibi tahmin edilir (Denklem 3.3):

$$Acc(M) = \sum_{i=1}^k 0.632 \times Acc(M_i)_{testk\u00fcmesi} + 0.368 \times Acc(M_i)_{eğitimk\u00fcmesi} \quad (3.3)$$

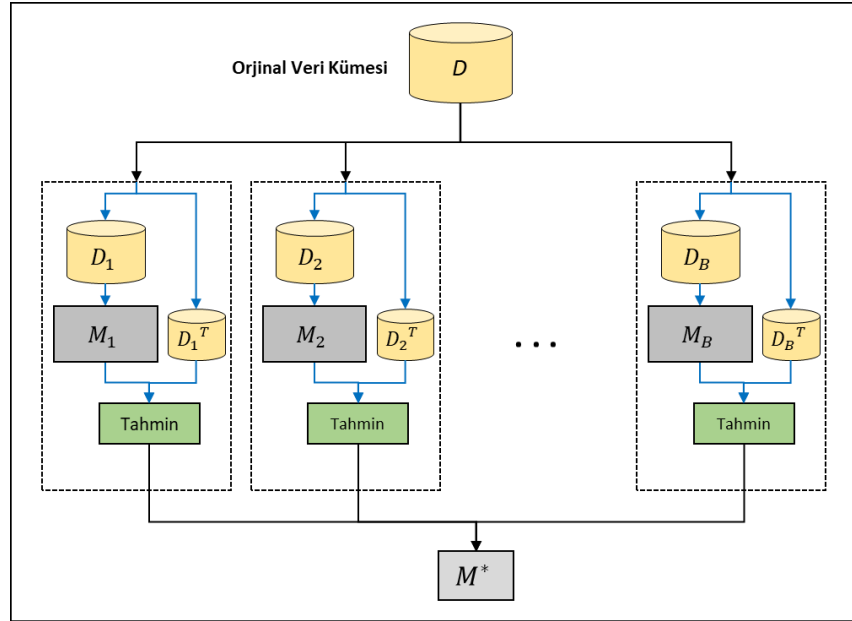
Bu denklemde kullanılan  $Acc(M_i)_{testk\u00fcmesi}$   $i$ . önyükleme örnekleme ile elde edilen modelin test verileri ile hesaplanan doğruluğudur.  $Acc(M_i)_{eğitimk\u00fcmesi}$ ,  $i$ . önyükleme örnekleme ile elde edilen modelin eğitim verileri ile hesaplanan doğruluğudur. Önyükleme yöntemi küçük veri kümelerinde daha doğru sonuçlar vermektedir (Han ve Kamber 2011). Şekil 3.4'de önyükleme yönteminin şematik gösterimi verilmiştir.



Şekil 3.4. Önyükleme yönteminin şematik gösterimi

### 3.2.5. Torbalama ve güçlendirme yöntemleri

Sınıflandırıcı ve tahmin edici doğruluğunu iyileştirmek için yaygın olarak kullanılan yöntemlerden biri de topluluk yöntemlerdir. Topluluk yöntemler geliştirilmiş bir bileşik model,  $M^*$  oluşturmak amacıyla bir dizi öğrenilmiş  $B$  modelini (sınıflandırıcılar veya tahmin ediciler),  $M_1, M_2, \dots, M_B$  birleştirir. Hem torbalama hem de güçlendirme, sınıflandırma ve regresyon için kullanılabilir (Han ve Kamber 2011). Şekil 3.5'te topluluk yöntem şeması verilmiştir.



Şekil 3.5. Topluluk yöntemlerin şematik gösterimi



- **Torbalama**

Torbalama yönteminde  $n$  nesneden oluşan veri kümesinde önyükleme yöntemi kullanılarak  $B$  adet eğitim kümesi oluşturulur. Belirlenen bir tahmin modeli  $B$  eğitim kümelerine uygulanarak  $B$  adet tahmin edici oluşturulur. Yeni bir test nesnesini tahmin etmek için, her tahmin edici ile bir tahmin değeri hesaplanır. Yeni test nesnesinin nihai tahmin değeri tüm tahmin değerlerinin ortalaması alınarak belirlenir. Torbalama algoritmasının (Breiman 1996a) temel adımları Şekil 3.6'da verilmiştir.

---

**Algoritma 3:** Torbalama Algoritması

---

**Girdi:**  $D$  veri kümesi,  $B$  topluluktaki model sayısı, tahmin algoritması

**Çıktı:**  $M^*$  birleşik model

**Yöntemin adımları:**

- 1) **For**  $b = 1$  **to**  $B$  **Do** //  $B$  tane model oluşturulması.
- 2)  $D$ 'nin değiştirilerek örnekleme ile  $D_b$  önyükleme örneklemini oluştur.
- 3)  $D_b$  kullanılarak  $M_b$  modelini oluştur.
- 4) **End For**
- 5) Her bir  $M_b$  modelinde test nesnesinin tahmin edilmesi ve tahmin değerlerinin ortalamasını al.

---

**Şekil 3.6.** Torbalama algoritmasının temel adımları (Breiman 1996a)

- **Güçlendirme**

Güçlendirme algoritmalarında eğitim kümesindeki her bir nesnenin bir ağırlığı vardır. Eğitim işleminden sonra her bir tahmin edicinin tahmin hatası hesaplanarak eğitim nesnelerin ağırlıkları güncellenir. Yeni bir test nesnesinin nihai tahmin değeri belirlenirken her tahmin edicinin doğruluğuna bağlı olarak tahmin değerlerinin ağırlıklı ortalaması alınır. Bu tezde Shrestha ve Solomatine (2005) tarafından geliştirilen Adaboost.RT algoritması dikkate alınmıştır. Adaboost.RT algoritmasının temel adımları Şekil 3.7'de verilmiştir.

---

**Algoritma 4:** Adaboost.RT Algoritması

---

**Girdi:**  $D$  veri kümesi, tahmin algoritması,  $B$  iterasyon sayısı (tamsayı), eşik değeri  $\varphi$  ( $0 < \varphi < 1$ )

**Çıktı:** Topluluk model

**Yöntemin adımları:**

- 1) İterasyon sayısını  $b = 1$  olarak belirlenmesi.
- 2) Her  $i$  nesne ağırlığının  $w_b(i) = 1/n$  olarak dağıtılması.
- 3) Hata oranı  $\varepsilon_b = 0$  olarak belirlenmesi.
- 4) **While**  $b \leq B$
- 5) Tahmin algoritmasını çağır,  $w_b$  dağılımını hesapla.
- 6) Regresyon modelini oluştur.

$$f_b(x) \rightarrow y^T$$

- 7) Her eğitim örneği için mutlak göreceli hatayı (Absolute Relative Error – ARE) hesapla.

$$ARE_b(i) = \left| \frac{f_b(x_i) - y_i}{y_i} \right|$$

- 8)  $f_b(x)$ 'in hata oranı  $\varepsilon_b$  hesapla.

$$\varepsilon_b = \sum_{i:ARE_b(i)>\varphi} w_b(i)$$

- 9)  $\beta_b = \varepsilon_b^a$  hesapla ( $a$ : güç katsayısı).

- 10)  $w_b$  dağılımını aşağıda verilen formüle göre güncelle.

$$w_{b+1}(i) = \frac{w_b(i)}{Z_b} \times \begin{cases} \beta_b, & \text{eğer } ARE_b(i) \leq \varphi \\ 1, & \text{diğer durumlarda} \end{cases}$$

Burada  $Z_b$ ,  $w_{b+1}$  bir dağılım olacak şekilde seçilen bir normalleştirme faktörüdür.

$b = b + 1$  olarak ayarla.

- 11) **End While**

- 12) Topluluk modelin tahmin değerini hesapla.

$$f_{fin}(x) = \frac{\sum_b \left( \log \frac{1}{\beta_b} \right) f_b(x)}{\sum_b \left( \log \frac{1}{\beta_b} \right)}$$

---

**Şekil 3.7.** Adaboost.RT algoritmasının temel adımları (Shrestha ve Solomatine 2005)

Adaboost.RT algoritmasında model performansı diğer güçlendirme yöntemlerinde olduğu gibi hata oranı  $\varepsilon_b$  hesaplanarak değerlendirilir. Bu algoritmada iyi veya kötü tahmin edilen nesnelere belirlenirken, mutlak göreceli hata kullanılmaktadır. Herhangi bir nesne için mutlak göreceli hata, eşik değeri  $\varphi$ 'den büyükse, bu nesnenin tahmin değeri

kötü tahmin edilmiş olarak kabul edilir, aksi halde iyi tahmin edilmiştir. Adaboost.RT ile yapılan deneyler topluluk modelin performansının  $\varphi$ 'ye duyarlı olduğunu göstermiştir (Shrestha ve Solomatine 2005). Ağırlık güncelleme parametresi  $\beta_b$ ,  $\varepsilon_b$  hata oranının  $a$  güç katsayısı kuvveti ile belirlenmiştir. Bununla birlikte  $a$ 'nın değeri artıkça  $\varepsilon_b$ 'nin çok düşük olduğu zor tahmin edilen nesnelere göreceli olarak daha fazla ağırlık verilir ve model bu zor tahmin edilen nesnelere üzerine yoğunlaşır. Adaboost.RT algoritmasında  $\varepsilon_b$ 'ye bağlı olarak "doğal" bir durdurma kriteri yoktur. Son olarak farklı modellerden elde edilen tahmin değerlerinin ağırlıklı ortalaması alınarak sonuçlar birleştirilir.

### 3.2.6. Yerel aykırı değer faktörü

Yerel aykırı faktör (Local Outlier Factor – LOF) yaklaşımı, bir nesne komşuluğunda ortalama noktaya özgü mesafelerle, mesafeleri normalleştirerek küme yoğunluğundaki yerel varyasyonları ayarlar. Pratikte, normalleştirme faktörünün ortalama yerel veri yoğunluğuna karşılık geldiği (normalleştirilmiş) mesafe tabanlı bir yaklaşım olmasına rağmen, genellikle yoğunluğa dayalı bir yaklaşım olarak anlaşılmaktadır (Bramer 2007). LOF algoritmasının (Breunig ve ark. 2000) temel adımları Şekil 3.8'de verilmiştir.

LOF' ta öncelikle bir  $p$  nesnesinin komşularına olan uzaklıkları hesaplanır. Uzaklıklar küçükten büyüğe sıralandığında  $K$ . sıradaki nesnenin  $p$  nesnesine olan uzaklığı  $Kdistance(p)$  olarak belirlenir. Ardından belirlenen  $K$  değerine göre  $p$  nesnesinin  $o$  nesnesine olan ulaşılabilirlik mesafesi  $reachdist_K(p, o)$  hesaplanır. Ulaşılabilirlik mesafesi  $p$  nesnesinin  $o$  nesnesine olan uzaklığı ile  $p$  nesnesinin  $Kdistance(p)$  değerinden büyük olmalıdır. Hesaplanan ulaşılabilirlik mesafesi bir nesnenin yerel ulaşılabilirlik yoğunluğunun hesaplanması için kullanılır. Bir  $p$  nesnesi için belirlenen  $K$  adet komşuluk için hesaplanan ulaşılabilirlik mesafeleri toplanarak  $K$  sayısına bölünür ve tersi alınır. Bir  $p$  nesnesinin yerel aykırı değer faktörü belirlenirken  $p$  nesnesinin yerel ulaşılabilirlik yoğunluğu ile komşuluğundaki nesnelere yerel ulaşılabilirlik yoğunlukları oranlanarak toplanır. Son olarak komşuluk sayısına bölünür.

---

**Algoritma 5:** Yerel Aykırı Değer Faktörü Algoritması

---

**Girdi:**  $D$  veri kümesi,  $MinPts$ , uzaklık fonksiyonu,  $K$  değeri

**Çıktı:** Veri kümesindeki nesnelerin yerel aykırı değer faktörü

**Yöntem adımları:**

1) Belirlenen uzaklık fonksiyonuna göre her bir  $p$  ile bir  $o \in D$  nesnesi arasındaki  $d(p, o)$  mesafesi hesaplanır.

2) Her bir  $p$  nesnesinin  $K$ . en yakın komşusu ile arasındaki mesafe  $Kdistance(p)$  hesaplanır.

3) Her bir  $p$  nesnesinin  $K$ -komşuluk mesafesi  $N_{K-distance(p)}(p)$  hesaplanır.

$$N_{Kdistance(p)}(p) = N_K(p) = \{q \in D \setminus \{p\} \mid d(p, q) \leq Kdistance(p)\}$$

4) Her bir  $p$  nesnesinin  $o$  nesnesine olan ulaşılabilirlik mesafesi  $reachdist_K(p, o)$  hesaplanır.

$$reachdist_K(p, o) = \max \{Kdistance(o), d(p, o)\}$$

5) Her bir  $p$  nesnesinin yerel ulaşılabilirlik yoğunluğu  $lrd_{MinPts}(p)$  hesaplanır.

$$lrd_{MinPts}(p) = 1 / \left( \frac{\sum_{o \in N_{MinPts}(p)} reachdist_{MinPts}(p, o)}{|N_{MinPts}(p)|} \right)$$

6) Her bir  $p$  nesnesinin yerel aykırı değer faktörü  $LOF_{MinPts}(p)$  hesaplanır.

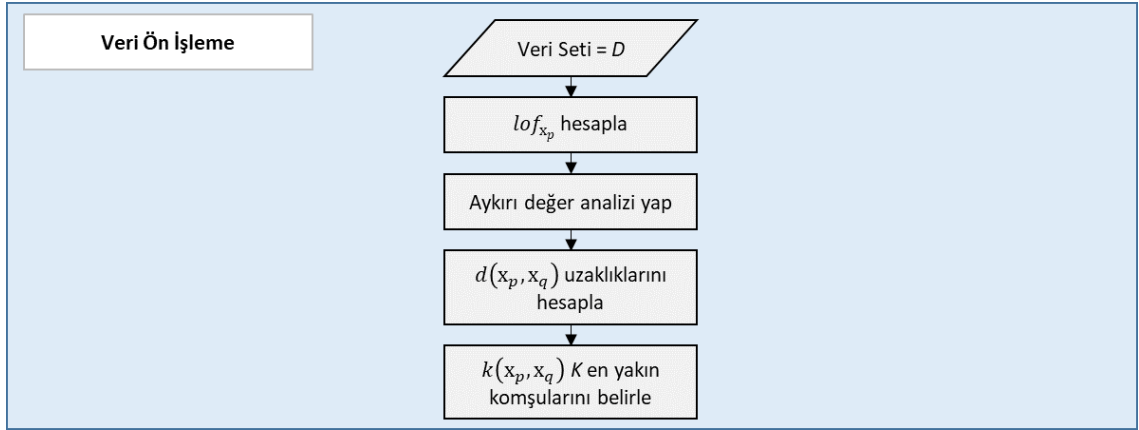
$$LOF_{MinPts}(p) = \frac{\sum_{o \in N_{MinPts}(p)} \frac{lrd_{MinPts}(o)}{lrd_{MinPts}(p)}}{|N_{MinPts}(p)|}$$

---

**Şekil 3.8.** Yerel aykırı değer faktörü algoritmasının temel adımları

### 3.2.7. Önerilen yaklaşımlar

Bu tezde tahmin problemlerinin çözümü için ağaç tabanlı yöntemler ve komşuluk tabanlı yöntemler üzerinde çalışmalar gerçekleştirilmiştir. Önerilen ağaç tabanlı yöntemler Regresyon Ağacı (RA), Torbalama Regresyon Ağacı (TRA) ve Güçlendirme Regresyon ağacı (GRA) algoritmalarıdır. Komşuluk tabanlı yöntemler ise  $K$ -En Yakın Komşuluk (KNN) ve Torbalama  $K$ -En Yakın Komşuluk (TKNN) algoritmalarıdır. Önerilen yöntemler iki aşamadan oluşmaktadır. Bu aşamalar veri ön işleme ve tahmin modelinin oluşturulmasıdır. Şekil 3.9'da veri ön işleme aşamasının işlem adımları verilmiştir. Şekil 3.11'de ağaç tabanlı yöntemler için, Şekil 3.12'de komşuluk tabanlı yöntemler için tahmin modellerinin oluşturulması aşamasının işlem adımları göstermektedir.



**Şekil 3.9.** Veri ön işleme akış şeması

Veri ön işleme adımı hem ağaç tabanlı yöntemler hem de komşuluk tabanlı yöntemler için aynıdır. Komşuluk tabanlı yöntemler ile tahmin yapılırken veri kümesindeki aykırı veriler tahmin değerinde yanıltıcı olmaktadır. Bu sebeple öncelikle veri kümelerine aykırı değer analizi yapılmaktadır. Aykırı değer analizinde yerel aykırı değer faktörü bir eşik değerinin üzerinde olan nesnelere aykırı veri olarak kabul edilir. Tespit edilen aykırı nesnelere veri kümesinden çıkarıldıktan sonra her bir  $p$  nesnesinin,  $q$  nesnesine olan uzaklığı  $d(x_p, x_q)$  ve uzaklık sıralaması  $k(x_p, x_q)$  hesaplanarak uzaklık ve sıralama matrisleri oluşturulur (Şekil 3.10).

---

**Algoritma 6:** Uzaklık matrisinin ve sıralama matrisinin oluşturulması

---

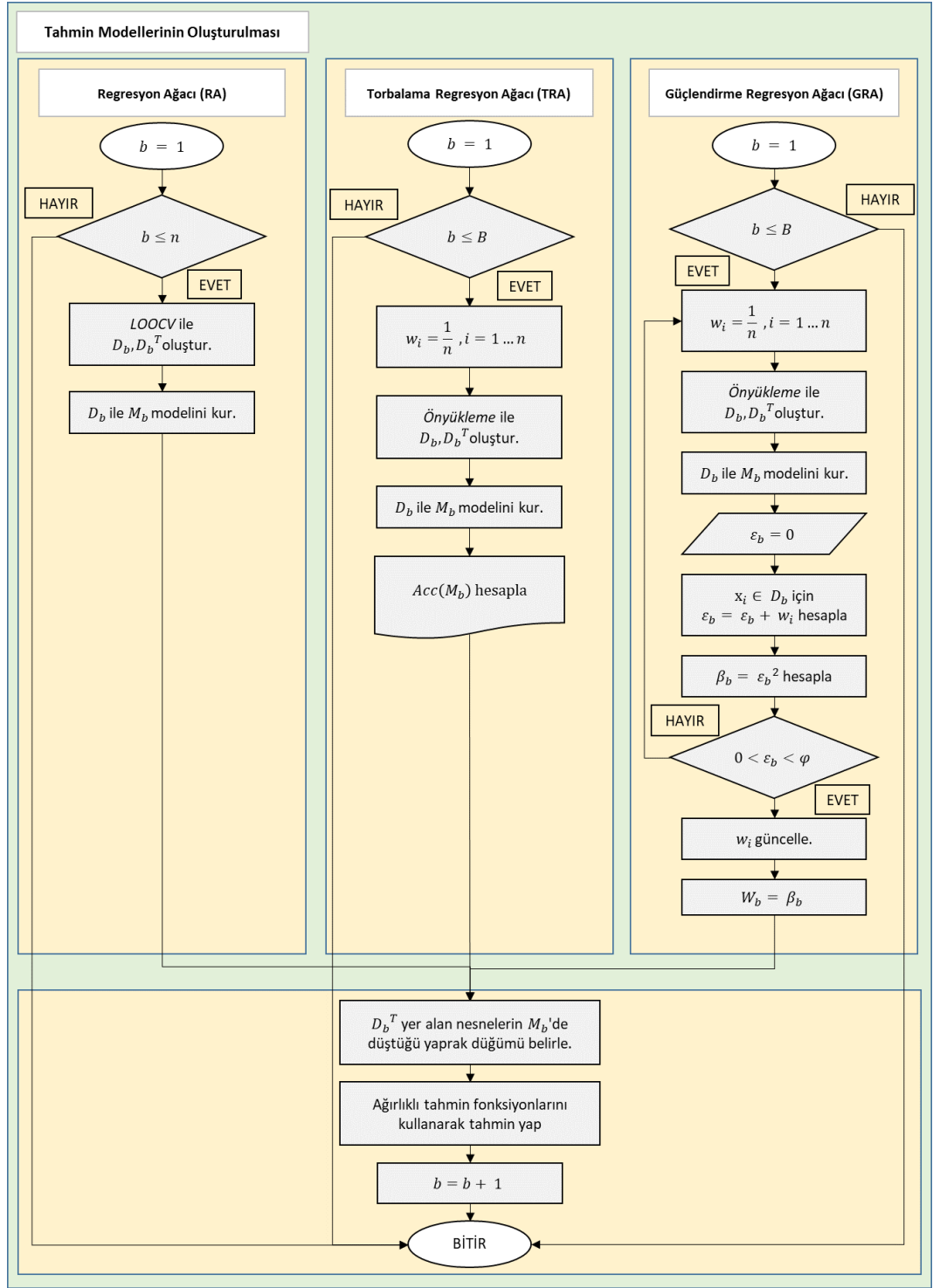
**Girdi:**  $D$  veri kümesi, öklid mesafesi

**Çıktı:** Uzaklık ve sıralama matrisleri

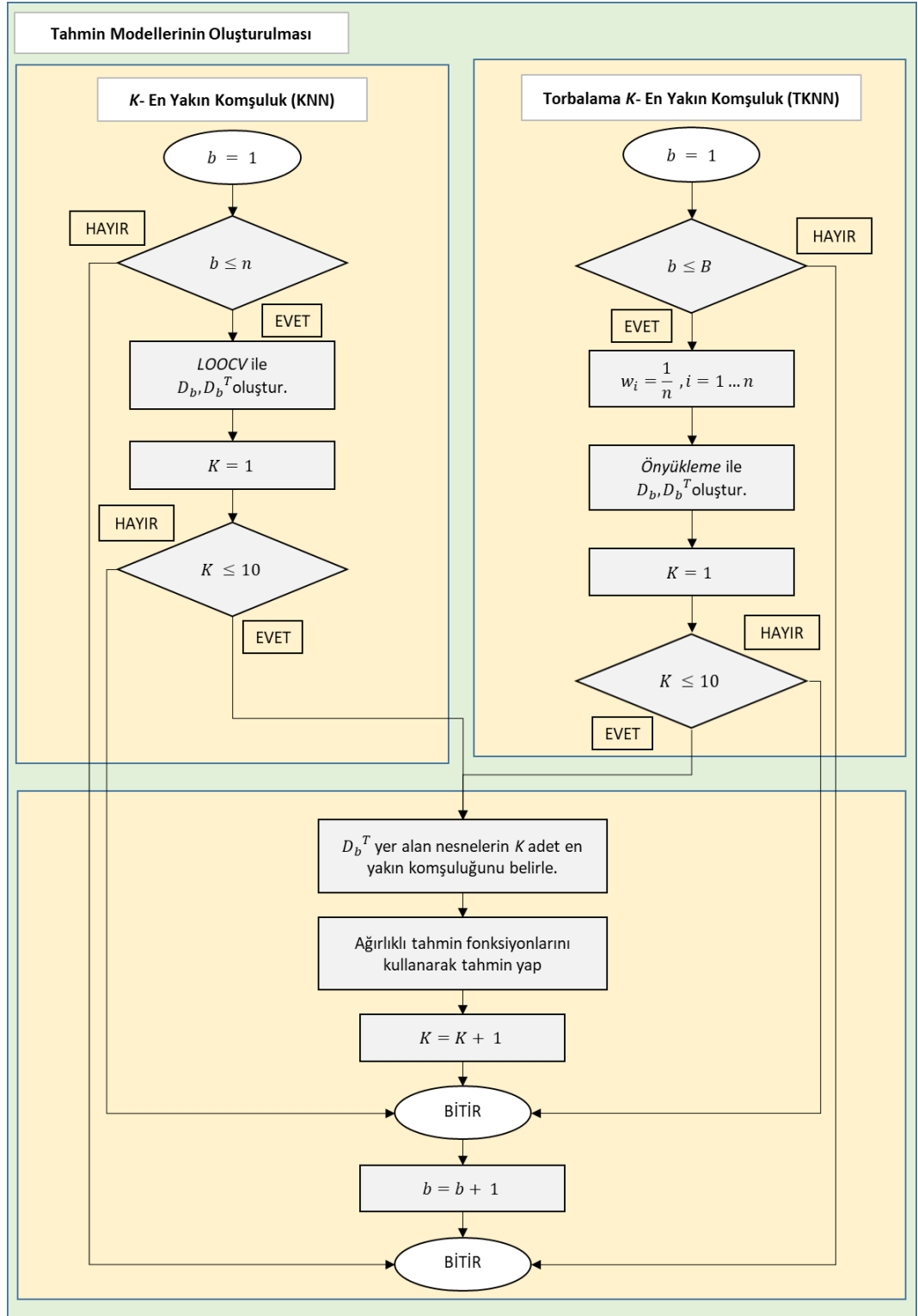
**Yöntemin adımları:**

- 1) Kategorik öznitelikler için ikili değişken olarak dönüştür.
  - 2) **For**  $p = 1$  to  $n$  **Do**
  - 3)     **For**  $q = 1$  to  $n$  **Do**
  - 4)         Denklem 3.1' e göre uzaklık matrisi  $d(x_p, x_q)$  hesapla.
  - 5)     **End For**
  - 6)      $d(x_p, x_q)$  değerlerini sıralayarak  $k(x_p, x_q)$  değerlerini belirle.
  - 7)     Sıralama matrisini oluştur.
  - 8) **End For**
- 

**Şekil 3.10.** Uzaklık matrisinin ve sıralama matrisinin oluşturulması



**Şekil 3.11.** Ağaç tabanlı yöntemlerin akış şeması



**Şekil 3.12.** Komşuluk tabanlı yöntemlerin akış şeması

Ağaç tabanlı yöntemlerde tahmin modellerinin eğitim veri kümeleri  $D_b$  oluşturulurken RA algoritmasında LOOCV yönteminden, TRA ve GRA algoritmalarında 0,632 ön

yükleme performans değerlendirme yönteminden yararlanılmıştır. 0,632 ön yükleme yöntemi ile eğitim veri kümesi oluşturulurken, TRA algoritmalarında nesnelere eşit ağırlıklıdır  $w_i = \frac{1}{n}$ , bu da her bir nesnenin eşit seçilme olasılığına sahip olduğunu göstermektedir. GRA algoritmasında ise başlangıçta eşit ağırlığa sahip olan nesnelere oluşturulan ağaçların hata miktarına  $\varepsilon$  göre güncellenmektedir. Oluşturulacak bir sonraki ağaçta her bir nesne farklı seçilme olasılığına sahip olmaktadır.

Ağaç tabanlı yöntemlerde tahmin modelinin eğitilmesi aşamasında kullanılan  $M_b$  modelinin temel adımları Şekil 3.13'te verilmiştir. TRA algoritmasında oluşturulan her bir modelin doğruluğu  $Acc(M_b)$  hesaplanarak, birleşik model  $M^*$  oluşturulurken her bir modelin tahmin değerine katkısı doğruluğu ile orantılı olarak dahil edilmiştir. GRA algoritmasında ise ağacın güçlendirme değerine  $\beta_b$  bağlı olarak ağaçlar birbirinden farklı ağırlıklar almaktadır. Böylece birleşik model  $M^*$  oluşturulurken yüksek hataya sahip modellerin tahmin değerine olan etkisi düşürülürken, doğruluğu yüksek olan modellerin tahmin değerine etkisi arttırılmaktadır. Ağaç tabanlı yöntemlerin tahmin modelleri RA algoritması için Ek 1'de, TRA algoritması için Ek 2'te, GRA algoritması için Ek 3'te detaylı olarak verilmiştir.

Komşuluk tabanlı yöntemlerde tahmin modellerinin eğitim veri kümeleri  $D_b$  oluşturulurken KNN algoritmasında LOOCV yöntemini, TKNN algoritmasında 0,632 ön yükleme performans değerlendirme yöntemi kullanılmıştır. Eğitim veri kümesi ve test veri kümesi oluşturulduktan sonra her bir test nesnesinin  $K$ -en yakın komşulukları belirlenmektedir.

Komşuluk tabanlı yöntemlerin tahmin modelleri KNN algoritması için Ek 4'te, TKNN algoritması için Ek 5'te verilmiştir.



---

**Algoritma 7:** Önerilen ağaç tabanlı yöntemlerde  $M_b$  eğitim modelinin oluşturulması

---

**Girdi:**  $D_b$  eğitim veri kümesi,  $derinlik_{max}$ ,  $n_{min}$ ,  $\epsilon$

**Çıktı:** Regresyon ağacı

**Yöntemin adımları:**

- 1)  $j = 1$
  - 2)  $derinlik_j = 0$
  - 3)  $Düğüm_j = D_b \setminus \setminus$  kök düğümü
  - 4)  $\bar{Y}_j$  hesapla
  - 5)  $RMSE_j$  hesapla.
  - 6) **Repeat**
  - 7)     **For**  $i = 1$  **to**  $m$  **Do** //veri kümesindeki her bir öznitelik için
  - 8)              $X_i$  özniteliğini en iyi bölen bölünme noktasını belirle.
  - 9)              $X_i$  özniteliğinden düğümü bölmenin hatası  $RMSE_{X_i}$ :
  - 10)     **End For**
  - 11)      $Düğüm_j$ 'yi en iyi bölen  $X$  özniteliğini seç.  
$$RMSE_{bölünme} = \min_{i=1 \dots (m-1)} (RMSE_{X_i})$$
  - 12)     **If**  $([RMSE_j - RMSE_{bölünme}] / RMSE_j) > \epsilon$  **And**  
 $derinlik_j < derinlik_{max}$  **And**  $nesneSayısı_j \geq n_{min}$  **Then**
  - 13)              $Düğüm_{j+1}$  ve  $Düğüm_{j+2}$  oluştur.
  - 14)              $\bar{Y}_{j+1}$  hesapla.
  - 15)              $\bar{Y}_{j+2}$  hesapla.
  - 16)              $derinlik_{j+1} = derinlik_j + 1$
  - 17)              $derinlik_{j+2} = derinlik_j + 1$
  - 18)     **Else**
  - 19)              $Düğüm_j$  düğümünü yaprak düğüm olarak belirle.
  - 20)     **End If**
  - 21)      $j = j + 1$
  - 22) **Until**
- 

**Şekil 3.13.** Önerilen ağaç tabanlı yöntemlerde  $M_b$  eğitim modelinin oluşturulması

Tahmin değerinin belirlenmesi için hem ağaç tabanlı yöntemlerde hem de komşuluk tabanlı yöntemlerde dokuz adet ağırlıklı tahmin fonksiyonu (ATF) önerilmiştir. Ağırlıklı tahmin fonksiyonları bir nesnenin çıktı değerini belirlerken kullanılan eğitim nesnelerinin ağırlıklarını belirlemektedir. Regresyon ağaçlarının yaprak düğümlerinde ATF fonksiyonlarının kullanımı ile karmaşık tahmin fonksiyonlarının oluşturduğu hesaplama maliyeti azaltılmak istenmektedir. Ayrıca sadece yaprak düğüme düşen nesnelerin

ortalamaları alan yaklaşımların tahminlerini iyileştirilmek amaçlanmıştır. Önerilen ağırlıklı tahmin fonksiyonu ATF1, ATF2, ATF3, ATF4, ATF5, ATF6, ATF7, ATF8 ve ATF9 olarak tanımlanmaktadır. Regresyon ağacı ve KNN yöntemlerindeki eşit ağırlıklı tahmin fonksiyonu ATF1 olarak adlandırılmıştır. Diğer ağırlıklı tahmin fonksiyonları nesnelerin birbirine olan uzaklıklarına, uzaklık sıralamalarına ve yerel aykırı değer faktörlerine göre farklı ağırlık değerleri hesaplamaktadır. Önerilen ağırlıklı tahmin fonksiyonlarını kullanılarak bir  $x_i$  nesnenin tahmin değeri  $y_i^T$  sırasıyla aşağıda verilen denklemlerdeki gibi hesaplanmaktadır.

$$\mathbf{ATF1:} \quad y_i^T = \frac{\sum_{j=1}^{n_i^*} y_j}{n_i^*} \quad (3.4)$$

Burada  $n_i^*$  ağaç tabanlı yöntemlerde  $x_i$  nesnesinin düştüğü yaprak düğümündeki nesne sayısını ifade etmektedir. Komşuluk tabanlı yöntemlerde bu sayı belirlenen  $K$  değerine karşılık gelmektedir.  $y_j$  ise  $x_i$  nesnesinin düştüğü yaprak düğümdeki nesnelere kümesi  $D_i^*$  içindeki her bir nesnenin çıktı değeridir. Komşuluk tabanlı yöntemlerde  $D_i^*$  belirlenen  $K$  değerine karşılık gelen en yakın komşularının oluşturduğu kümedir.

$$\mathbf{ATF2:} \quad y_i^T = \frac{\sum_{j=1}^{n_i^*} \left( \frac{1}{k(x_i, x_j)} \times y_j \right)}{\sum_{j=1}^{n_i^*} \frac{1}{k(x_i, x_j)}} \quad (3.5)$$

Burada  $k(x_i, x_j)$   $x_j$  nesnesinin  $x_i$  nesnesine göre uzaklık sıralamasıdır ve Şekil 3.10'daki gibi hesaplanmaktadır. Yoğunluk farkı ve aykırı değer içeren veri kümelerinde sıralama tabanlı yöntemler yetersiz olabilmektedir. Bu nedenle, ağırlıklı tahmin fonksiyonlarında nesnelerin komşuluk ilişkilerinin yanı sıra yoğunluk ve yerel aykırılıkları dikkate alarak hesaplanan yerel aykırı değer faktörü eklenmiştir.

$$\mathbf{ATF3:} \quad y_i^T = \frac{\sum_{j=1}^{n_i^*} \left( \frac{1}{k(x_i, x_j)} \times \frac{lof_{x_j}}{lof_{x_i}} \times y_j \right)}{\sum_{j=1}^{n_i^*} \left( \frac{1}{k(x_i, x_j)} \times \frac{lof_{x_j}}{lof_{x_i}} \right)} \quad (3.6)$$

Burada  $lof_{x_i}$ ,  $x_i$  nesnesinin yerel aykırı değer faktörüdür.  $lof_{x_j}$  ise  $D_i^*$  veri kümesindeki  $j$ . nesnesinin yerel aykırı değer faktörüdür ve Şekil 3.8'deki gibi hesaplanmaktadır.

$$\mathbf{ATF4:} \quad y_i^T = \frac{\sum_{j=1}^{n_i^*} \left( \frac{1}{k(x_i, x_j)} \times \frac{\min(lof_{x_j}, lof_{x_i})}{\max(lof_{x_j}, lof_{x_i})} \times y_j \right)}{\sum_{j=1}^{n_i^*} \left( \frac{1}{k(x_i, x_j)} \times \frac{\min(lof_{x_j}, lof_{x_i})}{\max(lof_{x_j}, lof_{x_i})} \right)} \quad (3.7)$$

$$\mathbf{ATF5:} \quad y_i^T = \frac{\sum_{j=1}^{n_i^*} \left( \frac{1}{k(x_i, x_j)} \times \frac{1}{lof_{x_j}} \times y_j \right)}{\sum_{j=1}^{n_i^*} \left( \frac{1}{k(x_i, x_j)} \times \frac{1}{lof_{x_j}} \right)} \quad (3.8)$$

$$\mathbf{ATF6:} \quad y_i^T = \frac{\sum_{j=1}^{n_i^*} \left( \frac{1}{d(x_i, x_j)} \times y_j \right)}{\sum_{j=1}^{n_i^*} \frac{1}{d(x_i, x_j)}} \quad (3.9)$$

Burada  $d(x_i, x_j)$ ,  $x_j$  nesnesinin  $x_i$  nesnesine göre Öklid uzaklığıdır ve Şekil 3.10'daki gibi hesaplanmaktadır.

$$\mathbf{ATF7:} \quad y_i^T = \frac{\sum_{j=1}^{n_i^*} \left( \frac{1}{d(x_i, x_j)} \times \frac{lof_{x_j}}{lof_{x_i}} \times y_j \right)}{\sum_{j=1}^{n_i^*} \left( \frac{1}{d(x_i, x_j)} \times \frac{lof_{x_j}}{lof_{x_i}} \right)} \quad (3.10)$$

$$\mathbf{ATF8:} \quad y_i^T = \frac{\sum_{j=1}^{n_i^*} \left( \frac{1}{d(x_i, x_j)} \times \frac{\min(lof_{x_j}, lof_{x_i})}{\max(lof_{x_j}, lof_{x_i})} \times y_j \right)}{\sum_{j=1}^{n_i^*} \left( \frac{1}{d(x_i, x_j)} \times \frac{\min(lof_{x_j}, lof_{x_i})}{\max(lof_{x_j}, lof_{x_i})} \right)} \quad (3.11)$$

$$\mathbf{ATF9:} \quad y_i^T = \frac{\sum_{j=1}^{n_i^*} \left( \frac{1}{d(x_i, x_j)} \times \frac{1}{lof_{x_j}} \times y_j \right)}{\sum_{j=1}^{n_i^*} \left( \frac{1}{d(x_i, x_j)} \times \frac{1}{lof_{x_j}} \right)} \quad (3.12)$$

## 4. BULGULAR ve TARTIŞMA

Bu bölümde geliştirilen algoritmaların karşılaştırmalı değerlendirme veri kümelerine ve sac şekillendirme kalıp imalatı yapan bir firmanın verilerine uygulanması ile elde edilen bulgular verilerek tartışılmaktadır. Bölüm 4.1’de karşılaştırmalı değerlendirme veri kümeleri açıklanmıştır. Bölüm 4.2’de tahmin sonuçlarına göre model performanslarının değerlendirilmesi için kullanılan performans metrikleri verilmiştir. Bölüm 4.3’de karşılaştırmalı değerlendirme veri kümelerinin sayısal sonuçları verilmiştir. Bölüm 4.4’te tahmin yöntemlerinin karşılaştırılması yapılmıştır. Bölüm 4.5’te sac metal kalıpcılık sektöründe üretim sürelerinin tahmini için yapılan vaka analizi çalışması verilmiştir.

### 4.1. Karşılaştırmalı Değerlendirme Veri Kümeleri

Bu tezde önerilen yaklaşımların performansının değerlendirilmesi amacıyla KEEL veri havuzundan (Alcalá-Fdez ve ark. 2011) seçilen dokuz adet karşılaştırmalı değerlendirme (benchmark) veri kümesi kullanılmıştır. Seçilen veri kümeleri *Diabetes*, *MachineCPU*, *Dee*, *AutoMpg6*, *Baseball*, *Laser*, *Concrete*, *Ele-2* ve *Friedman*’dır. Karşılaştırmalı değerlendirme veri kümelerinin özellikleri Çizelge 4.1’de verilmiştir.

- *Diabetes* veri kümesi, çocuklarda insüline bağımlı “diabetes mellitus” modellerini etkileyen faktörlerin araştırılmasıyla ilgilidir. 43 nesne ve gerçek sayılardan oluşan iki sürekli girdi değişkeni içermektedir. Çıktı değişkeni sürekli değerler almaktadır.
- *MachineCPU* veri kümesi CPU'nun yayınlanan göreceli performansını (Published Relative Performance – PRP) yaklaşık olarak belirlemek için oluşturulmuştur. 209 nesneye ve tamsayılardan oluşan altı sürekli girdi değişkenine sahiptir. Çıktı değişkeni tamsayılardan oluşan sürekli değerler almaktadır.
- *Dee* veri kümesi İspanya’da günlük ortalama elektrik enerjisi fiyatını tahmin etmek için 2003 yılına ait gerçek veriler kullanılarak oluşturulmuştur. Veri

kümesinde 365 nesne ve gerçek sayılardan oluşan altı sürekli girdi değişkeni içermektedir. *Dee* veri kümesinin çıktı değişkeni gerçek sayılardan oluşmaktadır.

- *AutoMpg6* veri kümesi galon başına mil (Mpg) cinsinden şehir döngüsü (city-cycle) yakıt tüketimiyle ilgilidir. Veri kümesinde 392 nesne, üç tamsayı girdi değişkeni ve iki gerçek sayılı girdi değişkeni bulunmaktadır. *AutoMpg6* veri kümesinin çıktı değişkeni gerçek sayılardan oluşmaktadır.
- *Baseball* veri kümesi, atıcılar hariç, hem 1991 hem de 1992 sezonlarında en az bir maç oynayan Major League Beyzbol oyuncularının 1992 yılı maaşlarını içermektedir. Veri kümesinde 337 nesne, 14 tamsayıdan oluşan sürekli girdi değişkeni ve iki gerçek sayılı girdi değişkeni bulunmaktadır. Çıktı değişkeni tamsayılardan oluşan sürekli değerler almaktadır.
- *Laser* veri kümesi, karmakarışık bir durumda bir Uzak Kızılötesi Lazer'den kaydedilen, gözlemlenen tek bir miktarın tek değişkenli bir zaman kayıdır. *Laser* veri kümesinde 993 nesne ve gerçek sayılardan oluşan dört sürekli girdi değişkeni yer almaktadır. Bu veri kümesinin çıktı değişkeni gerçek sayılardan oluşan sürekli değerler almaktadır, ayrıca tekrarlanan nesnelere veri kümesinden kaldırılmıştır.
- *Concrete* veri kümesi beton dayanımını belirlemek için oluşturulmuştur. Veri kümesinde 1030 nesne, gerçek sayılardan oluşan yedi sürekli girdi değişkeni ve bir tamsayı sürekli girdi değişkeni yer almaktadır. Bu veri kümesinin çıktı değişkeni gerçek sayılardan oluşan sürekli değerler almaktadır.
- *Ele-2* veri kümesi 1056 nesne ve gerçek sayılardan oluşan dört sürekli girdi değişkeni içermektedir. Çıktı değişkeni gerçek sayılardan oluşan sürekli bir değişkendir.
- *Friedman* veri kümesi 1200 nesne ve gerçek sayılardan oluşan beş sürekli girdi değişkeni içermektedir. Çıktı değişkeni beş özneliğe bağlı bir fonksiyon ile

belirlenen gerçek sayılardan oluşan sürekli bir değişkendir. Veri kümesinde gürültü bulunmaktadır.

**Çizelge 4.1.** Karşılaştırmalı değerlendirme veri kümelerinin özellikleri

Veri Kümesi	Öznitelik Sayısı	Nesne Sayısı	Çıktı Değişkeni
<i>Diabetes</i>	2	43	Gerçek Sayı
<i>MachineCPU</i>	6	209	Tamsayı
<i>Baseball</i>	16	337	Tamsayı
<i>Dee</i>	6	365	Gerçek Sayı
<i>AutoMpg6</i>	5	392	Gerçek Sayı
<i>Laser</i>	5	996	Gerçek Sayı
<i>Concrete</i>	8	1030	Gerçek Sayı
<i>Ele-2</i>	4	1056	Gerçek Sayı
<i>Friedman</i>	5	1200	Gerçek Sayı

#### 4.2. Performans Metrikleri

Tahmin problemlerinde, tahmin edilen değerlerin gerçek bilinen değerden ne kadar uzaklıkta olduğuna bakılarak model performansları değerlendirilmektedir. Performans metrikleri olarak bilinen bu ölçümler,  $y_i$  ile  $y_i^T$  arasındaki hatayı ölçer. Bu tezde kullanılan performans metrikleri ortalama mutlak yüzde hata (Mean Absolute Percentage Error – MAPE) ve hata kareler ortalamasının kareköküdür (Root Mean Square Error – RMSE). Hata performans metriklerinin küçük değerler alması, tahmin edilen değerlerin gerçek değerlere daha yakın olduğunu göstermektedir. Performans metriklerinin hesabı Denklem 4.1 ve Denklem 4.2’de verilmiştir.

$$\text{MAPE} = 100 * \frac{\sum_{i=1}^n \frac{|y_i - y_i^T|}{y_i}}{n} \quad (4.1)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - y_i^T)^2}{n}} \quad (4.2)$$

### 4.3. Karşılaştırmalı Değerlendirme Veri Kümelerinin Sayısal Sonuçları

Bu bölümde önerilen yaklaşımların, karşılaştırmalı değerlendirme veri kümelerine uygulanması ile elde edilen sayısal sonuçlar açıklanmıştır. Elde edilen sayısal sonuçlara göre önerilen yaklaşımlar karşılaştırılmıştır. Bu tezde önerilen yaklaşımların geliştirilmesi ve tüm analiz çalışmaları için R istatistiksel hesaplama yazılımı (R Core Team 2019) kullanılmıştır.

Tahmin modelleri oluşturulmadan önce veri kümelerindeki numerik değişkenlere 0-1 normalizasyonu yapılmıştır. Kategorik değişkenler ikili değişkenlere dönüştürülmüştür. Sonrasında veri ön işleme aşamasına geçilmiştir. Regresyon ağaçlarında maksimum derinlik beş kabul edilmiştir. Bir düğümdeki minimum nesne sayısı, veri kümesindeki nesne sayısının onda birinden daha fazlaysa ve maksimum derinlik sağlanmadıysa bölünme işlemine iyileşme miktarı (0,01) da dikkate alınarak devam edilmiştir. KNN ve TKNN algoritmalarında  $K$  en yakın komşu sayısı 1'den 10'a kadar değiştirilmiştir. Model ağacı algoritması kıyaslama amacıyla kullanılmıştır. RA ve KNN modellerinin tahmin doğruluğunu belirlemek için birini dışarıda bırak çapraz geçişleme (LOOCV) performans değerlendirme yöntemi kullanılmıştır. TRA, GRA ve TKNN modellerinin model doğruluğu 0,632 önyükleme performans değerlendirme yöntemi ile beş yüz tekrar yapılarak hesaplanmıştır. Karşılaştırmalı değerlendirme veri kümelerinin Model ağacı algoritmasına göre sonuçları Bölüm 4.4'te verilmiştir.

Bu tezde elde edilen sayısal sonuçlar orijinal veri kümeleri ve aykırı değer analizi yapılan veri kümeleri için iki alt başlık altında açıklanacaktır. Çizelge 4.2 ve Çizelge 4.3'de önerilen ağaç tabanlı yöntemlerin, orijinal veri kümelerine uygulanması ile elde edilen MAPE ve RMSE sonuçları sırasıyla verilmiştir. Ek 6 ve Ek 7'de önerilen komşuluk tabanlı yöntemlerin, orijinal veri kümelerine uygulanması ile elde edilen MAPE ve RMSE sonuçları verilmiştir. Aykırı değer analizi yapılan Diabetes, MachineCPU, AutoMpg6, Baseball, Laser, Concrete ve Ele-2 karşılaştırmalı değerlendirme veri kümelerinde sırasıyla bir, on yedi, dört, dört, yirmi bir, yirmi altı ve iki tane aykırı değer tespit edilmiştir. Dee ve Friedman veri kümelerinde aykırı değere rastlanmamıştır. Çizelge 4.4 ve Çizelge 4.5'de önerilen ağaç tabanlı yöntemlerin, aykırı değer analizi

yapılan karşılaştırmalı değerlendirme veri kümelerine uygulanması ile elde edilen MAPE ve RMSE sonuçları sırasıyla verilmiştir. Ek 8 ve Ek 9’da önerilen komşuluk tabanlı yöntemlerin, aykırı değer analizi yapılan karşılaştırmalı değerlendirme veri kümelerine uygulanması ile elde edilen MAPE ve RMSE sonuçları verilmiştir. Verilen tablolarda koyu işaretlenen değerler her bir veri kümesi için en iyi performans sonucunu göstermektedir.

**Çizelge 4.2.** Orijinal veri kümelerine ağaç tabanlı yöntemlerin uygulanması ile elde edilen MAPE sonuçları (%)

Veri Kümesi	Model	Kullanılan Ağırlıklı Tahmin Fonksiyonu								
		ATF1	ATF2	ATF3	ATF4	ATF5	ATF6	ATF7	ATF8	ATF9
<i>Diabetes</i>	RA	9,42	9,32	9,39	9,34	9,32	10,03	10,08	10,04	9,98
	TRA	8,76	8,62	8,62	8,66	8,62	<b>8,57</b>	8,59	8,65	8,69
	GRA	8,74	8,63	8,63	8,63	8,63	8,68	8,64	8,68	8,68
<i>MachineCPU</i>	RA	45,06	37,57	37,73	37,54	37,61	39,17	39,36	39,14	39,1
	TRA	40,87	35,79	35,82	<b>35,56</b>	35,7	35,78	35,77	35,77	35,91
	GRA	40,79	36,21	36,22	36,03	36,13	36,16	36,16	36,06	36,08
<i>Baseball</i>	RA	37,33	26,78	26,6	26,56	27,01	22,69	22,64	<b>22,54</b>	22,85
	TRA	31,58	25,35	25,56	25,07	25,45	23,52	23,56	23,34	23,54
	GRA	31,95	25,71	25,96	25,41	25,69	23,78	23,93	23,56	23,73
<i>Dee</i>	RA	9,42	9,24	9,3	9,2	9,19	10,36	10,37	10,35	10,36
	TRA	8,54	8,33	8,36	<b>8,3</b>	<b>8,3</b>	8,38	8,42	8,37	8,35
	GRA	8,56	8,36	8,39	8,31	8,31	8,41	8,44	8,4	8,39
<i>AutoMPG6</i>	RA	68,79	71,93	72,24	72,2	71,57	78,12	78,23	78,12	78,01
	TRA	66,55	65,56	65,92	65,53	<b>65,25</b>	65,98	66,27	65,97	65,75
	GRA	66,67	66,12	66,48	66,1	65,79	66,31	66,59	66,28	66,05
<i>Laser</i>	RA	9,5	7,97	<b>7,95</b>	7,96	8	8,23	8,19	8,24	8,26
	TRA	8,56	7,96	7,96	7,91	7,97	8,41	8,41	8,38	8,42
	GRA	8,59	8,04	8,04	7,99	8,04	8,45	8,46	8,42	8,44
<i>Concrete</i>	RA	25,21	18,95	18,95	18,86	18,98	18,93	18,95	18,9	18,94
	TRA	23,24	18,98	19,93	18,83	19,05	20,17	20,14	20,13	20,21
	GRA	22,89	18,87	<b>18,83</b>	18,76	18,92	20,43	20,4	20,39	20,47
<i>Ele2</i>	RA	26,57	21,99	22	21,96	21,98	19,97	19,97	<b>19,96</b>	<b>19,96</b>
	TRA	24,28	21,85	21,85	21,82	21,84	20,25	20,26	20,23	20,24
	GRA	24,28	21,94	21,94	21,91	21,94	20,39	20,4	20,37	20,38
<i>Friedman</i>	RA	20,19	14,16	14,14	14,17	14,18	15,43	15,43	15,42	15,43
	TRA	17,13	12,82	12,76	12,83	12,87	15,97	15,94	15,98	15,99
	GRA	16,76	12,46	<b>12,4</b>	12,47	12,52	15,58	15,55	15,59	15,6



Çizelge 4.2’de en iyi performans gösteren modellere bakıldığında üç veri kümesinde RA, dört veri kümesinde TRA ve iki veri kümesinde GRA seçilmiştir. ATF1, ATF2 ve ATF7’nin hiçbir veri kümesinde üstün gelmediği görülmüştür. ATF3’ün en çok tercih edilen ağırlıklı tahmin fonksiyonu olduğu görülmektedir.

**Çizelge 4.3.** Orijinal veri kümelerine ağaç tabanlı yöntemlerin uygulanması ile elde edilen RMSE sonuçları

Veri kümesi	Model	Kullanılan Ağırlıklı Tahmin Fonksiyonu								
		ATF1	ATF2	ATF3	ATF4	ATF5	ATF6	ATF7	ATF8	ATF9
<i>Diabetes</i>	RA	0,69	0,68	0,69	0,69	0,69	0,72	0,73	0,73	0,73
	TRA	0,63	<b>0,61</b>	0,62	0,62	<b>0,61</b>	<b>0,61</b>	<b>0,61</b>	<b>0,61</b>	0,62
	GRA	0,63	0,63	0,63	0,63	0,63	0,63	0,63	0,63	0,63
<i>MachineCPU</i>	RA	72,93	70,71	69,78	70,11	71,64	69,19	69,32	69,47	69,07
	TRA	61,08	61,58	60,95	60,9	62,29	59,83	<b>59,25</b>	59,28	60,5
	GRA	63,1	63,53	62,85	62,84	64,24	61,59	60,91	60,99	62,34
<i>Baseball</i>	RA	18,85	13,63	13,3	13,74	14,04	13,49	13,32	13,62	13,7
	TRA	15,93	12,52	<b>12,37</b>	12,51	12,81	13,06	12,93	13,03	13,09
	GRA	15,9	12,57	12,42	12,57	12,85	13,14	13,03	13,09	13,17
<i>Dee</i>	RA	0,48	0,47	0,47	0,47	0,47	0,54	0,54	0,54	0,54
	TRA	0,43	<b>0,42</b>	<b>0,42</b>	<b>0,42</b>	<b>0,42</b>	0,43	0,43	<b>0,42</b>	<b>0,42</b>
	GRA	0,43	<b>0,42</b>	<b>0,42</b>	<b>0,42</b>	<b>0,42</b>	0,43	0,43	0,43	0,43
<i>AutoMPG6</i>	RA	740,26	727,29	729,31	728,3	725,46	790,16	791,86	790,53	788,69
	TRA	686	676,95	678,48	676,99	<b>675,7</b>	681,23	682,54	681,19	680,16
	GRA	689,91	682,64	684,1	682,63	681,38	685,86	687,17	685,82	684,73
<i>Laser</i>	RA	3,16	2,88	2,89	2,88	2,88	3,17	3,17	3,18	3,17
	TRA	2,93	2,81	2,81	<b>2,8</b>	2,81	2,95	2,95	2,95	2,95
	GRA	2,94	2,82	2,83	2,81	2,83	2,97	2,97	2,96	2,96
<i>Concrete</i>	RA	9,45	6,85	6,84	6,83	6,87	7,18	7,2	7,17	7,18
	TRA	8,23	6,82	6,83	<b>6,76</b>	6,83	7,34	7,35	7,3	7,34
	GRA	8,18	6,87	6,88	6,8	6,87	7,37	7,39	7,34	7,37
<i>Ele2</i>	RA	516,97	263,91	264,33	263,11	263,67	156,24	156,24	<b>156,03</b>	156,28
	TRA	435,66	269,17	269,19	268,22	269,21	164,35	164,42	164,01	164,3
	GRA	429,11	270,68	270,66	269,81	270,76	168,27	168,35	167,94	168,21
<i>Friedman</i>	RA	3,12	2,28	2,28	2,28	2,28	2,55	2,55	2,55	2,55
	TRA	2,65	2,07	2,06	2,07	2,08	2,51	2,51	2,51	2,51
	GRA	2,61	2,02	<b>2,01</b>	2,02	2,03	2,46	2,46	2,46	2,47

Çizelge 4.3’de verilen orijinal kümelerine ağaç tabanlı yöntemlerin uygulanması ile elde edilen RMSE sonuçlarına göre RA bir veri kümesinde, TRA yedi veri kümesinde, GRA ise bir veri kümesinde en iyi performans değerine sahiptir. Ağırlıklı tahmin fonksiyonlarında sadece ATF1 hiçbir veri kümesinde tercih edilmezken, yakın performans gösteren birçok ATF vardır.

**Çizelge 4.4.** Aykırı değer analizi yapılan veri kümelerine ağacı tabanlı yöntemlerin uygulanması ile elde edilen MAPE sonuçları (%)

Veri kümesi	Model	Kullanılan Ağırlıklı Tahmin Fonksiyonu								
		ATF1	ATF2	ATF3	ATF4	ATF5	ATF6	ATF7	ATF8	ATF9
<i>Diabetes</i>	RA	10,28	10,41	10,51	10,54	10,53	11,48	11,53	11,48	11,43
	TRA	8,44	8,15	<b>8,08</b>	8,15	8,15	8,15	8,16	8,16	8,19
	GRA	8,28	8,46	8,35	8,43	8,41	8,48	8,48	8,48	8,47
<i>MachineCPU</i>	RA	48	38,75	38,85	38,32	38,69	38,13	38,17	38,17	38,2
	TRA	39,38	35,03	35,11	<b>34,82</b>	35	35,03	35,02	34,86	34,97
	GRA	39,21	35,11	35,21	34,96	35,13	35,19	35,22	35,09	35,18
<i>Baseball</i>	RA	33,13	24,22	24,1	24,14	24,39	21,17	<b>21,12</b>	<b>21,12</b>	21,27
	TRA	31,28	25,01	22,95	24,85	24,87	25,21	22,88	22,77	23,02
	GRA	32,08	25,59	25,47	25,43	25,73	23,65	23,51	23,48	23,74
<i>Dee</i>	RA	9,42	9,24	9,3	9,2	9,19	10,36	10,37	10,35	10,36
	TRA	8,54	8,33	8,36	<b>8,3</b>	<b>8,3</b>	8,38	8,42	8,37	8,35
	GRA	8,56	8,36	8,39	8,31	8,31	8,41	8,44	8,4	8,39
<i>AutoMPG6</i>	RA	62,56	66,11	66,49	66,34	65,7	72,38	72,41	72,45	72,35
	TRA	62,39	61,67	61,98	61,69	61,36	61,58	61,84	61,57	61,33
	GRA	62	61,79	62,13	61,79	61,45	61,45	61,72	61,42	<b>61,18</b>
<i>Laser</i>	RA	9,34	8,06	8,1	8,04	8,03	8,92	8,93	8,9	8,89
	TRA	8,03	7,59	7,57	7,54	7,58	7,88	7,88	7,84	7,87
	GRA	8,03	7,55	7,56	<b>7,53</b>	7,56	7,84	7,84	7,82	7,85
<i>Concrete</i>	RA	25,22	18,75	18,73	18,69	18,77	18,76	18,77	18,72	18,76
	TRA	22,94	18,78	18,74	<b>18,67</b>	18,83	20,35	20,32	20,3	20,39
	GRA	22,89	18,87	18,83	18,76	18,92	20,43	20,4	20,39	20,47
<i>Ele2</i>	RA	26,58	21,99	21,98	21,97	21,99	19,97	19,97	<b>19,96</b>	19,97
	TRA	24,24	21,84	21,84	21,82	21,85	20,27	20,28	20,26	20,27
	GRA	24,3	21,93	21,93	21,91	21,93	20,43	20,44	20,42	20,43
<i>Friedman</i>	RA	20,19	14,16	14,14	14,17	14,18	15,43	15,43	15,42	15,43
	TRA	17,13	12,82	12,76	12,83	12,87	15,97	15,94	15,98	15,99
	GRA	16,76	12,46	<b>12,4</b>	12,47	12,52	15,58	15,55	15,59	15,6

Çizelge 4.4’de verilen aykırı değer analizi yapılan karşılaştırmalı değerlendirme veri kümelerine ağacı tabanlı yöntemlerin uygulanması ile elde edilen MAPE sonuçlarına göre iki veri kümesinde RA, dört veri kümesinde TRA ve üç veri kümesinde GRA iyi sonuç vermektedir. ATF1, ATF2 ve ATF6 hiçbir veri kümesinde tercih edilmemiştir.

**Çizelge 4.5.** Aykırı değer analizi yapılan veri kümelerine ağacı tabanlı yöntemlerin uygulanması ile elde edilen RMSE sonuçları

Veri kümesi	Model	Kullanılan Ağırlıklı Tahmin Fonksiyonu								
		ATF1	ATF2	ATF3	ATF4	ATF5	ATF6	ATF7	ATF8	ATF9
<i>Diabetes</i>	RA	0,75	0,74	0,74	0,74	0,74	0,79	0,79	0,79	0,78
	TRA	0,6	<b>0,59</b>	<b>0,59</b>	<b>0,59</b>	<b>0,59</b>	<b>0,59</b>	<b>0,59</b>	<b>0,59</b>	<b>0,59</b>
	GRA	0,6	0,61	0,61	0,61	0,61	0,61	0,61	0,61	0,61
<i>MachineCPU</i>	RA	55,08	43,32	43,19	43,21	43,45	40,41	40,41	40,51	40,39
	TRA	38,56	35,63	<b>35,47</b>	35,63	35,8	36,56	36,42	36,57	36,65
	GRA	38,89	36,29	36,29	36,45	36,98	36,2	36,91	36,96	37,05
<i>Baseball</i>	RA	16,48	11,86	11,84	11,81	11,89	12,23	12,22	12,23	12,26
	TRA	15,06	11,72	11,66	11,67	11,78	11,73	11,7	11,67	11,74
	GRA	15,04	11,71	11,68	<b>11,62</b>	11,75	11,86	11,86	11,81	11,87
<i>Dee</i>	RA	0,48	0,47	0,47	0,47	0,47	0,54	0,54	0,54	0,54
	TRA	0,43	<b>0,42</b>	<b>0,42</b>	<b>0,42</b>	<b>0,42</b>	0,43	0,43	<b>0,42</b>	<b>0,42</b>
	GRA	0,43	<b>0,42</b>	<b>0,42</b>	<b>0,42</b>	<b>0,42</b>	0,43	0,43	0,43	0,43
<i>AutoMPG6</i>	RA	730,92	721,11	722,82	722,04	719,44	789,8	791,1	790,23	788,68
	TRA	691,68	684,07	685,31	683,74	<b>683,04</b>	686,82	687,85	686,46	685,95
	GRA	690,77	685,17	684,99	684,1	686,74	686,39	687,59	686,64	685,95
<i>Laser</i>	RA	3,17	2,91	2,92	2,9	2,89	3,31	3,33	3,31	3,31
	TRA	2,68	<b>2,57</b>	<b>2,57</b>	<b>2,57</b>	2,58	2,64	2,64	2,63	2,64
	GRA	2,69	2,58	2,58	<b>2,57</b>	2,58	2,65	2,65	2,64	2,65
<i>Concrete</i>	RA	9,37	6,84	6,85	6,81	6,85	7,08	7,09	7,06	7,08
	TRA	8,14	6,78	6,79	<b>6,74</b>	6,79	7,37	7,37	7,34	7,37
	GRA	8,09	6,81	6,82	6,76	6,81	7,39	7,39	7,35	7,39
<i>Ele2</i>	RA	517,44	264,12	264,45	263,37	263,93	156,32	156,3	<b>156,11</b>	156,36
	TRA	432,17	267,24	267,22	266,31	267,32	164,45	164,51	164,13	164,41
	GRA	427,48	266,58	266,59	265,67	266,62	168,29	168,38	167,97	168,23
<i>Friedman</i>	RA	3,12	2,28	2,28	2,28	2,28	2,55	2,55	2,55	2,55
	TRA	2,65	2,07	2,06	2,07	2,08	2,51	2,51	2,51	2,51
	GRA	2,61	2,02	<b>2,01</b>	2,02	2,03	2,46	2,46	2,46	2,47

Çizelge 4.5’de verilen aykırı değer analizi yapılan karşılaştırmalı değerlendirme veri kümelerine ağacı tabanlı yöntemlerin uygulanması ile elde edilen RMSE sonuçlarına göre bir veri kümesinde RA, altı veri kümesinde TRA ve iki veri kümesinde GRA en iyi performansı göstermiştir. ATF1 hiçbir veri kümesi tarafından tercih edilmezken, belirli bir veri kümesi için benzer performans gösteren ağırlıklı tahmin fonksiyonları vardır.

Ağaç tabanlı modellerin MAPE ve RMSE sonuçlarına bakıldığında TRA’nın diğer modellere göre daha fazla tercih edildiği görülmüştür. Algoritmaların orijinal hallerini gösteren ATF1’in ise hiçbir veri kümesinde üstünlük göstermediği görülmüştür.

#### **4.4. Tahmin Yöntemlerinin Karşılaştırılması**

Birden fazla veri kümesi üzerinde iki veya daha fazla modelin sonuçları karşılaştırılırken parametrik olmayan istatistiksel testler kullanılmaktadır. Birden fazla veri kümesinde karşılaştırma yapılırken alınan ölçümler, veri kümelerinin bağımsız olmasından dolayı varyans farklılıklarına neden olmaktadır. Parametrik olmayan istatistiksel testler normallik veya varyans homojenliği gibi herhangi bir varsayıma dayanmamaktadır (Demšar 2006).

Bu tezde önerilen modellerin hem ikili hem de çoklu karşılaştırmaları yapılmıştır. Çoklu karşılaştırmalar için Friedman testi uygulanırken, ikili karşılaştırmalar için Wilcoxon eşleştirilmiş işaretli sıra testi (Wilcoxon’s paired signed-rank test) uygulanmıştır.

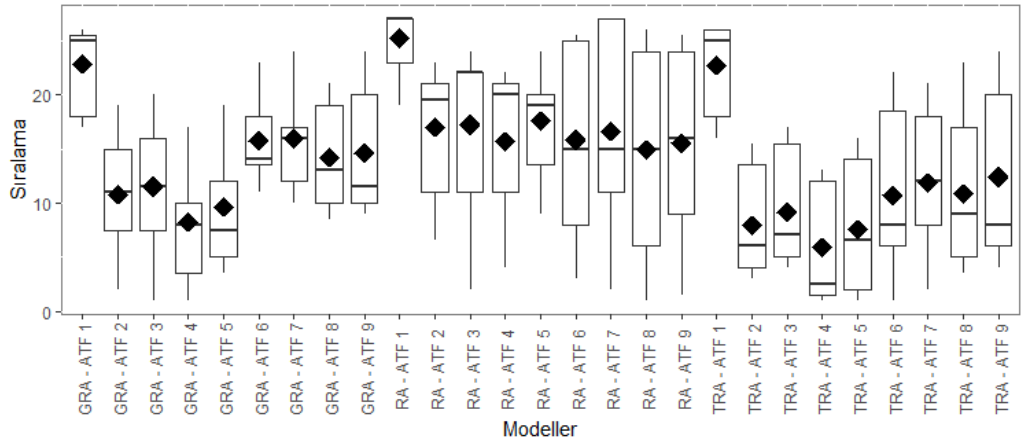
Yapılan testlerde öncelikle ağaç tabanlı yöntemler ve komşuluk tabanlı yöntemler kendi içlerinde değerlendirilmiştir. Ağaç tabanlı yöntemler için yapılan testlerde 27 karşılaştırma modeli (3 tane ağaç tabanlı model x 9 tane ATF) dikkate alınmıştır. Komşuluk tabanlı yöntemlerde farklı  $K$  değeri ile modeller çalıştırıldığı için 180 karşılaştırma modeli (2 tane komşuluk tabanlı model x  $K = 1, 2, \dots, 10$  x 9 tane ATF) dikkate alınmıştır. Hem ağaç tabanlı yöntemler hem de komşuluk tabanlı yöntemler için karşılaştırmalı değerlendirme veri kümelerinin aykırı değer analizi öncesinde ve sonrasında elde edilen MAPE ve RMSE sonuçları ayrı ayrı değerlendirilmiştir.

#### 4.4.1. Friedman testi

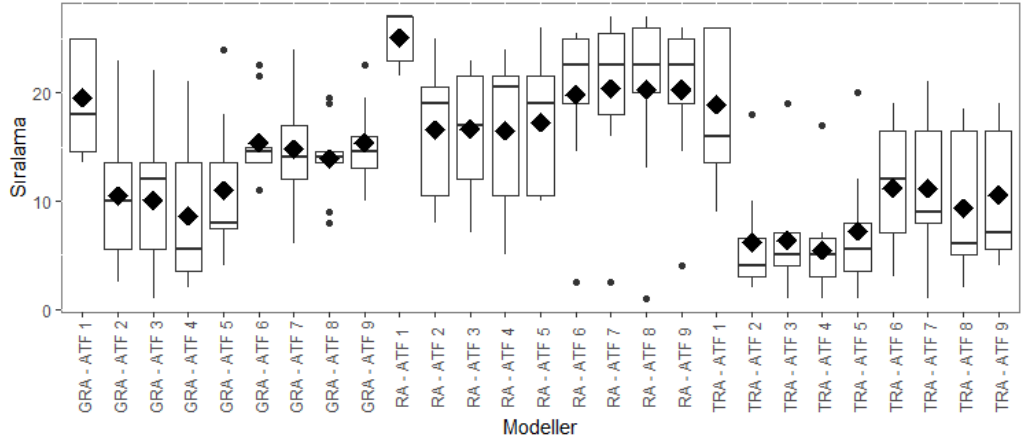
Friedman (1937, 1940) tarafından geliştirilen Friedman testi parametrik iki yönlü varyans analizinin parametrik olmayan bir benzeridir. Testin amacı birçok veri kümesi ve rakip modeller arasında anlamlı bir farklılık olup olmadığını belirlemektir. Friedman testinin ilk adımı her bir veri kümesi için karşılaştırılan modellerin orijinal sonuçlarının sıralanmasıdır.  $V$  adet veri kümesi ve  $h$  adet karşılaştırma modeli için  $r_i^j$ ,  $j$ . karşılaştırma modelinin  $i$ . veri kümesindeki sıralamasıdır. Her bir karşılaştırma modeli için  $R_j$  ortalama sıralama Denklem 4.3'deki gibi hesaplanmaktadır.

$$R_j = \frac{1}{V} \sum_{i=1}^V r_i^j \quad (4.3)$$

Şekil 4.1 ve Şekil 4.2'de önerilen ağaç tabanlı yöntemlerin, orijinal veri kümelerine uygulanması ile elde edilen MAPE ve RMSE sonuçlarına göre  $r_i^j$  değerlerinin kutu bilyık diyagramları verilmiştir. Ek 10 ve Ek 11'de önerilen komşuluk tabanlı yöntemlerin, orijinal veri kümelerine uygulanması ile elde edilen MAPE ve RMSE sonuçlarına göre  $R_j$  değerleri verilmiştir.

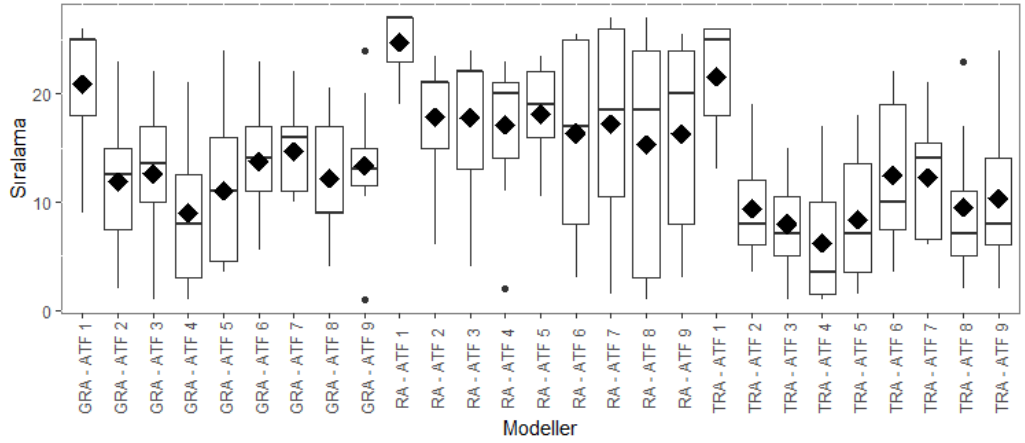


**Şekil 4.1.** Orijinal veri kümelerine ağaç tabanlı yöntemlerin uygulanması ile elde edilen MAPE sonuçlarına göre  $r_i^j$  değerlerinin kutu bilyık diyagramı

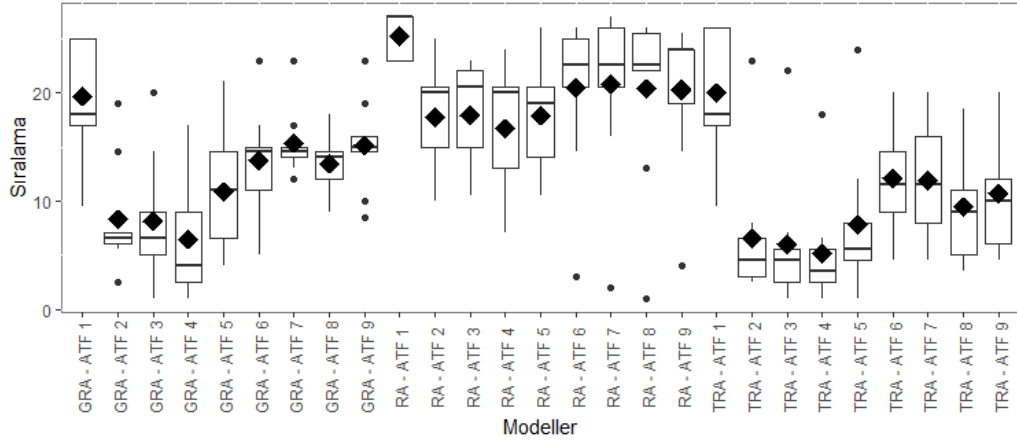


**Şekil 4.2.** Orijinal veri kümelerine ağaç tabanlı yöntemlerin uygulanması ile elde edilen RMSE sonuçlarına göre  $r_t^j$  değerlerinin kutu bıyık diyagramı

Şekil 4.3 ve Şekil 4.4’de önerilen ağaç tabanlı yöntemlerin, aykırı değer analizi yapılan veri kümelerine uygulanması ile elde edilen MAPE ve RMSE sonuçlarına göre  $r_t^j$  değerlerinin kutu bıyık diyagramları verilmiştir. Ek 12 ve Ek 13’de önerilen komşuluk tabanlı yöntemlerin, aykırı değer analizi yapılan veri kümelerine uygulanması ile elde edilen MAPE ve RMSE sonuçlarına göre  $R_j$  değerleri verilmiştir.



**Şekil 4.3.** Aykırı değer analizi yapılan veri kümelerine ağaç tabanlı yöntemlerin uygulanması ile elde edilen MAPE sonuçlarına göre  $r_t^j$  değerlerinin kutu bıyık diyagramı



**Şekil 4.4.** Aykırı değer analizi yapılan veri kümelerine ağaç tabanlı yöntemlerin uygulanması ile elde edilen RMSE sonuçlarına göre  $r_t^j$  değerlerinin kutu bilyık diyagramı

MAPE ve RMSE sonuçlarına göre verilen kutu bilyık diyagramları incelendiğinde TRA ve GRA sıralama değerlerinin RA algoritmasına göre daha küçük olduğu görülmektedir. RA-ATF1, TRA-ATF1 ve GRA-ATF1 yöntemlerinin sıralama değerlerinin diğer yöntemlere göre büyük olması önerilen komşuluk tabanlı ağırlıklı tahmin fonksiyonlarının tahmin sonuçlarının daha iyi olduğunu göstermektedir.

Friedman istatistiği  $V$  ve  $h$  yeterince büyük olduğunda ( $V > 10$  ve  $h > 5$ )  $h - 1$  serbestlik dereceli  $\chi_F^2$  göre dağıtılır. Çizelge 4.6'da orijinal veri kümeleri için Friedman testi ile hesaplanan değerleri  $\chi_F^2$  ve ki-kare tablo değerleri  $\chi^2$  verilmiştir. Çizelge 4.7'de aykırı değer analizi sonrası Friedman testi ile hesaplanan değerleri  $\chi_F^2$  ve ki-kare tablo değerleri  $\chi^2$  verilmiştir. Friedman testi için  $H_0$  ve  $H_1$  hipotezleri aşağıda verilmiştir.

$H_0$ : Yöntemlerin performanslarında farklılık yoktur.

$H_1$ : Yöntemlerden en az birinin performansında farklılık vardır.

**Çizelge 4.6.** Orijinal veri kümeleri için Friedman test sonuçları

	MAPE performans kriterine göre sonuçlar		RMSE performans kriterine göre sonuçlar	
	Hesaplanan $\chi_F^2$	Ki-kare tablo değeri $\chi^2$	Hesaplanan $\chi_F^2$	Ki-kare tablo değeri $\chi^2$
<b>Ağaç Tabanlı Yöntemler</b>	<b>86,00</b>	38,885 ( $d_f = 26, p = 0,05$ )	<b>104,95</b>	38,885 ( $d_f = 26, p = 0,05$ )
<b>Komşuluk Tabanlı Yöntemler</b>	195,54	211,217 ( $d_f = 179, p = 0,05$ )	<b>497,65</b>	211,217 ( $d_f = 179, p = 0,05$ )

Ağaç tabanlı yöntemlerin orijinal veri kümelerinde MAPE ve RMSE sonuçlarına göre  $\chi^2 < \chi_F^2$  olduğu için  $H_0$  hipotezi reddedilmektedir. Komşuluk tabanlı yöntemlerin orijinal veri kümelerinde MAPE sonuçlarına göre  $\chi^2 > \chi_F^2$  olduğu için  $H_0$  hipotezi kabul edilmektedir. RMSE sonuçlarına göre  $\chi^2 < \chi_F^2$  olduğu için  $H_0$  hipotezi reddedilmektedir. Orijinal veri kümelerinde yapılan Friedman testlerinde sadece komşuluk tabanlı yöntemlerin MAPE sonuçlarına karşılaştırılan modeller arasında istatistiksel olarak anlamlı bir farklılık görülmemiştir.

**Çizelge 4.7.** Aykırı değer analizi yapılan veri kümeleri için Friedman test sonuçları

	MAPE performans kriterine göre sonuçlar		RMSE performans kriterine göre sonuçlar	
	Hesaplanan $\chi_F^2$	Ki-kare tablo değeri $\chi^2$	Hesaplanan $\chi_F^2$	Ki-kare tablo değeri $\chi^2$
<b>Ağaç Tabanlı Yöntemler</b>	<b>76,20</b>	38,885 ( $d_f = 26, p = 0,05$ )	<b>120,83</b>	38,885 ( $d_f = 26, p = 0,05$ )
<b>Komşuluk Tabanlı Yöntemler</b>	146,34	211,217 ( $d_f = 179, p = 0,05$ )	<b>637,75</b>	211,217 ( $d_f = 179, p = 0,05$ )

Aykırı değer analizi yapılan veri kümeleri için Friedman test sonuçları, orijinal veri kümeleri için Friedman test sonuçlarına benzerdir. Yapılan Friedman testlerinde sadece komşuluk tabanlı yöntemlerin MAPE sonuçlarına göre karşılaştırılan modeller arasında istatistiksel olarak anlamlı bir farklılık görülmemiştir.

#### 4.4.2. Wilcoxon eşleştirilmiş işaretli sıra testi

Önerilen modellerin ikili karşılaştırmaları için iki aşamalı Wilcoxon eşleştirilmiş işaretli sıra testi (Wilcoxon's paired signed-rank test) yapılmıştır. Birinci aşamada aralarında



istatistiksel olarak anlamlı fark bulunan modeller belirlenmiştir. İkinci aşamada aralarında istatistiksel olarak anlamlı fark olan modeller sıralanarak en iyi performans gösteren modeller belirlenmiştir. İkili karşılaştırma için kurulan hipotez testi aşağıda verilmiştir.

$H_0$ : Karşılaştırılan iki modelin hata miktarı eşittir.

$H_1$ : Karşılaştırılan iki modelden birinci modelin hata miktarı, ikinci modele göre daha azdır.

Çizelge 4.8 ve Çizelge 4.9’da verilen en iyi performans gösteren ilk iki sıradaki modellere bakıldığında ağaç tabanlı yöntemlerin topluluk versiyonları olarak geliştirilen TRA ve GRA yöntemlerinin öne çıktığı görülmektedir. Bununla birlikte ağaç tabanlı yöntemler ile tahmin yapılırken yaprak düğümünde eşit ağırlıklı tahmin fonksiyonu (ATF1) kullanmak yerine nesnelere birbirine olan uzaklık sıralamalarını, uzaklıklarını ya da yerel aykırı değer faktörlerini dikkate alan ağırlıklı tahmin fonksiyonlarının (ATF2, ATF3, ATF4, ATF5, ATF8, ATF9) daha iyi sonuç verdiği görülmüştür. Komşuluk tabanlı yöntemler için en iyi performans gösteren ilk iki sıradaki modeller incelendiğinde TKNN yönteminin KNN yönteminden her zaman daha iyi sonuç verdiği dair yeterli kanıt yoktur. Ancak hem KNN yönteminde hem de TKNN yönteminde eşit ağırlıklı tahmin fonksiyonu (ATF1) yerine nesnelere farklı ağırlık değerlerini (ATF2, ATF6, ATF7, ATF8, ATF9) kullanan modellerin daha iyi performans gösterdiği görülmüştür.

**Çizelge 4.8.** Orijinal veri kümelerinde ağaç ve komşuluk tabanlı yöntemler için MAPE ve RMSE sonuçlarına göre seçilen modeller

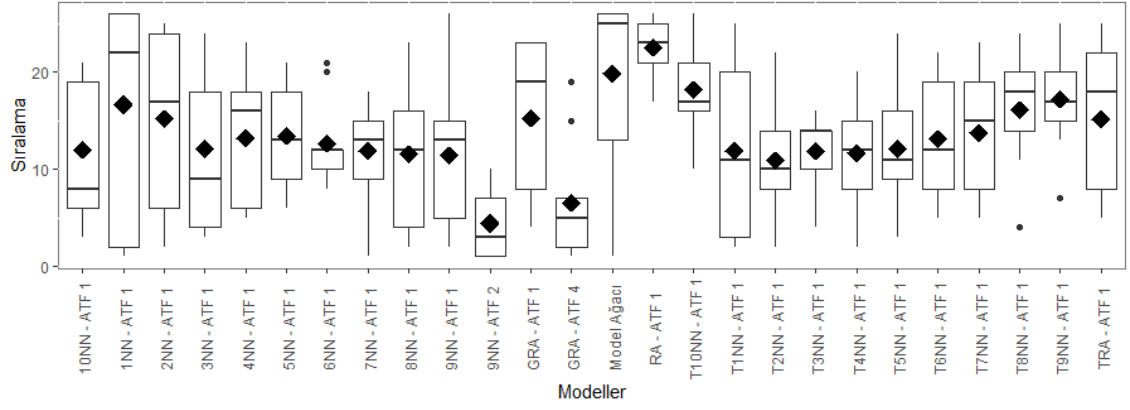
	MAPE performans kriterine göre sonuçlar		RMSE performans kriterine göre sonuçlar	
	Ağaç tabanlı yöntemler	Komşuluk tabanlı yöntemler	Ağaç tabanlı yöntemler	Komşuluk tabanlı yöntemler
<b>1. Seçim</b>	GRA-ATF4	9NN-ATF2	TRA-ATF8	T4NN-ATF8
<b>2. Seçim</b>	TRA-ATF2 TRA-ATF4 TRA-AFT5 GRA-ATF2 GRA-ATF3 GRA-ATF5	10NN-ATF2	TRA-ATF2 TRA-ATF3 TRA-ATF4 TRA-ATF9	T5NN-ATF8

**Çizelge 4.9.** Aykırı değer analizi yapılan veri kümelerinde ağaç ve komşuluk tabanlı yöntemler için MAPE ve RMSE sonuçlarına göre seçilen modeller

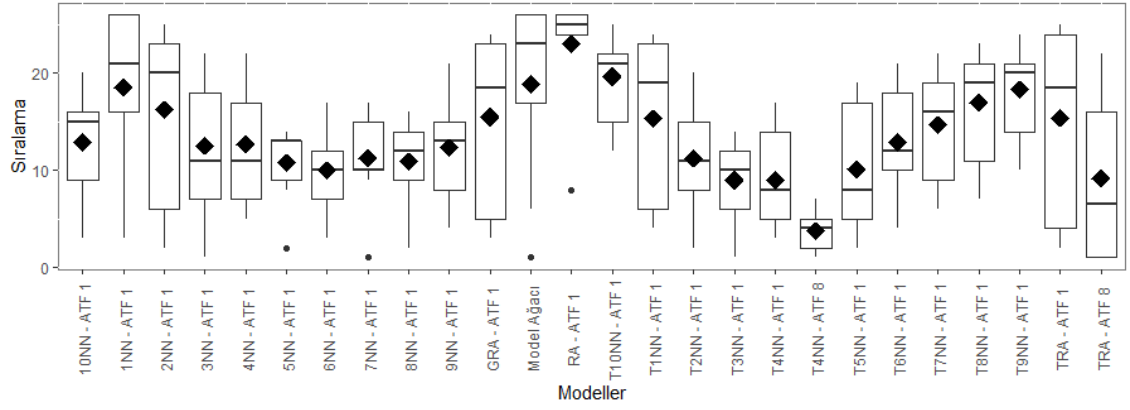
	MAPE performans kriterine göre sonuçlar		RMSE performans kriterine göre sonuçlar	
	Ağaç tabanlı yöntemler	Komşuluk tabanlı yöntemler	Ağaç tabanlı yöntemler	Komşuluk tabanlı yöntemler
<b>1. Seçim</b>	TRA-ATF4	9NN-ATF2	TRA-ATF8	T4NN-ATF8
<b>2. Seçim</b>	TRA-ATF3	10NN-ATF2	TRA-ATF4 TRA-ATF9 GRA-ATF4	7NN-ATF6 7NN-ATF7 7NN-ATF8 7NN-ATF9

Yapılan Friedman testleri ve Wilcoxon eşleştirilmiş işaretli sıra testinde ağaç tabanlı yöntemler ve komşuluk tabanlı yöntemler kendi aralarında gruplandırılarak incelenmiştir. Çalışmanın bundan sonraki aşamasında Model ağacı, ağaç tabanlı yöntemlerin ve komşuluk tabanlı yöntemlerin eşit ağırlıklı tahmin fonksiyonları (ATF1) ve Wilcoxon eşleştirilmiş işaretli sıra testi sonucuna göre en iyi performans gösteren yöntemler birlikte değerlendirilmiştir. Daha önce yapılan istatistiksel analizlerde olduğu gibi yeni karşılaştırma havuzları aykırı değer analizi çalışmalarına ve seçilen modellerin MAPE ve RMSE sonuçlarına göre ayrı ayrı değerlendirilmiştir. Oluşturulan yeni karşılaştırma havuzlarında yer alan modellerin  $r_i^j$  değerlerinin kutu bıyık diyagramları Şekil 4.5, Şekil 4.6, Şekil 4.7 ve Şekil 4.8’de verilmiştir.

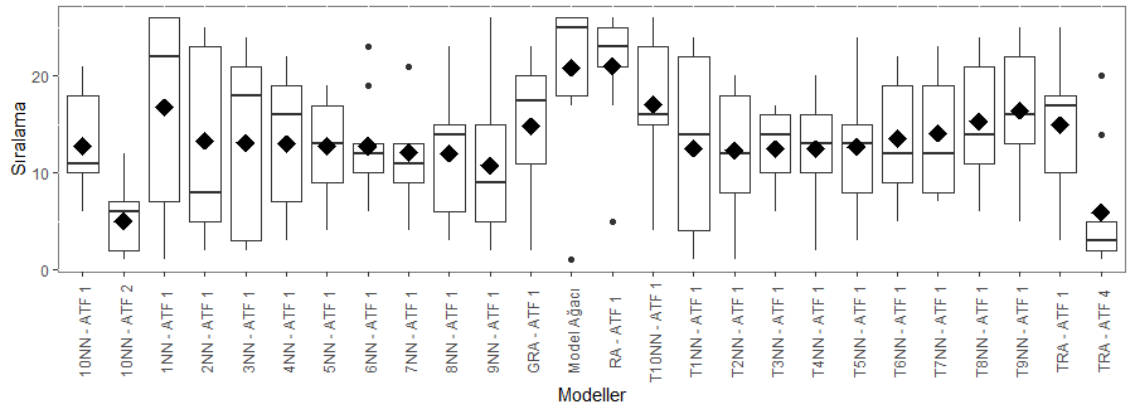
Orijinal veri kümelerinin MAPE sonuçlarına göre yapılan değerlendirmede geliştirilen GRA – ATF4 ve 9NN – ATF2 yöntemlerinin hem Model ağacı hem de geliştirilen diğer yöntemlerin eşit tahmin fonksiyonlu modellerine göre daha iyi performansa sahip oldukları görülmektedir (Şekil 4.5). RMSE sonuçlarına göre yapılan değerlendirmede TRA – ATF8 ve T4NN – ATF8 yöntemlerinin daha iyi performansa sahip oldukları görülmektedir (Şekil 4.6).



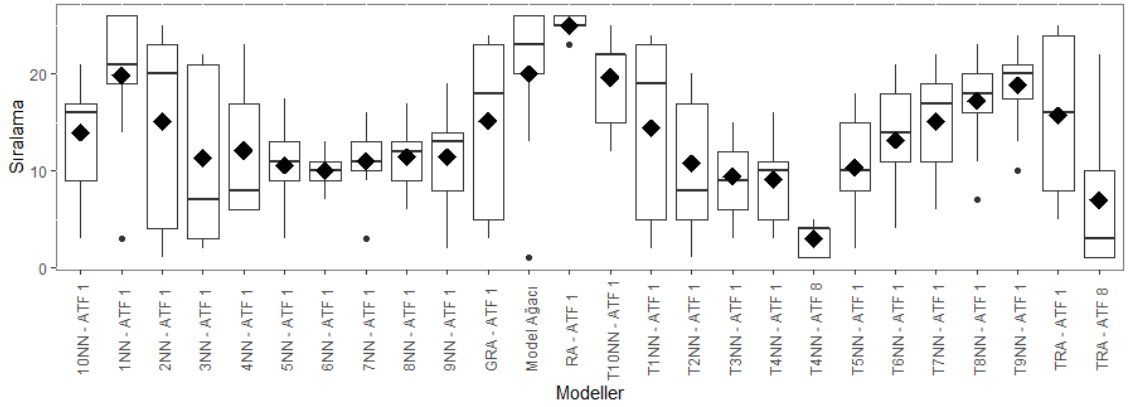
Şekil 4.5. Orijinal veri kümelerinin MAPE sonuçlarına göre belirlenen karşılaştırma modellerinin  $r_i^j$  değerlerinin kutu bilyik diyagramı



Şekil 4.6. Orijinal veri kümelerinin RMSE sonuçlarına göre belirlenen karşılaştırma modellerinin  $r_i^j$  değerlerinin kutu bilyik diyagramı



Şekil 4.7. Aykırı değer analizi yapılan veri kümelerinin MAPE sonuçlarına göre belirlenen karşılaştırma modellerinin  $r_i^j$  değerlerinin kutu bilyik diyagramı



**Şekil 4.8.** Aykırı değer analizi yapılan veri kümelerinin RMSE sonuçlarına göre belirlenen karşılaştırma modellerinin  $r_i^j$  değerlerinin kutu bilyık diyagramı

Aykırı değer analizi yapılan karşılaştırmalı değerlendirme veri kümelerinin MAPE sonuçlarına göre yapılan değerlendirmede geliştirilen TRA – ATF4 ve 10NN-ATF2 yöntemlerinin hem Model ağacı hem de geliştirilen diğer yöntemlerin eşit tahmin fonksiyonlu modellerine göre daha iyi performansa sahip oldukları görülmektedir (Şekil 4.7). RMSE sonuçlarına göre yapılan değerlendirmede TRA – ATF8 ve T4NN – ATF8 yöntemlerinin daha iyi performansa sahip oldukları görülmektedir (Şekil 4.8). Oluşturulan dört karşılaştırma havuzu için en iyi performansa sahip ilk iki sırada yer alan modeller Çizelge 4.10’da özetlenmiştir.

**Çizelge 4.10.** Model ağacı, ağaçlı tabanlı yöntemler ve komşuluk tabanlı yöntemlerin birlikte değerlendirilmesi ile en iyi performansa sahip yöntemlerin listesi

	Seçim sırası	MAPE sonuçlarına göre	RMSE sonuçlarına göre
Aykırı Değer Analizi Öncesi	1	9NN – ATF2	T4NN – ATF8
	2	GRA – ATF4	TRA – ATF8
Aykırı Değer Analizi Sonrası	1	TRA – ATF4	T4NN – ATF8
	2	10NN – ATF2	TRA – ATF8

Orijinal veri kümeleri ile yapılan karşılaştırmada MAPE sonuçlarına göre KNN ve GRA algoritmasının, RMSE sonuçlarına göre TKNN ve TRA algoritmasının en iyi sonucu verdiği görülmektedir. Aykırı değer analizi sonrası yapılan karşılaştırmada MAPE sonuçlarına göre TRA ve KNN algoritmasının, RMSE sonuçlarına göre TKNN ve TRA algoritmasının en iyi sonucu verdiği görülmektedir. ATF2, ATF4 ve ATF8 en iyi sonucu

veren yöntemlerde kullanılan ağırlıklı tahmin fonksiyonlarıdır. Model ağacı ve önerilen yaklaşımların orijinal hallerini temsil eden ATF1 yapılan karşılaştırma sonuçlarına göre ilk sıralarda tercih edilmemektedir. Çizelge 4.10'a göre tez kapsamında önerilen komşuluk tabanlı yöntemlerin daha iyi sonuç verdiği görülmektedir.

#### **4.5. Sac Metal Kalıpcılık Sektöründe Üretim Sürelerinin Tahmini için Vaka Analizi**

Siparişe dayalı imalat sistemlerinde üretim sürelerinin doğru ve hızlı tahmini üretim planlama, çizelgeleme ve kapasite planlama çalışmaları için oldukça önemlidir. Üretim sürelerinin belirlenmesi için sistematik bir tahmin yöntemi olmayan firmalarda güvenilir üretim planları yapılamamaktadır. Bu durum müşteriye verilen teslim tarihinde gecikmelere, öngörülemeyen kapasite boşluklarına, kaynakların verimsiz kullanılmasına neden olmaktadır. Gün geçtikçe artan maliyetler nedeniyle firmalar rekabet gücünü kaybetmektedir.

Siparişe dayalı imalat sistemlerinde, üretim süresi tahmini yapan uygulamalar vardır. Doğan ve Birant (2021) son yirmi yılda üretim süreçlerini iyileştirmek için kullanılan ana yaklaşımları ve algoritmaları derlemiştir. Weichert ve ark. (2019), 2008-2018 yılları arasında imalat sanayinde ürün kalitesi ve süreç iyileştirme ile ilgili yapılmış çalışmaları derlemiştir. Cheng ve ark. (2018) büyük veri çağında veri madenciliği tekniklerinin gelişimini gözden geçirmektedir ve 2010'dan itibaren yapılan çalışmaları analiz ederek veri madenciliği tekniklerinin üretim sistemlerindeki uygulamalarını tartışmıştır. Köksal ve ark. (2011) imalat sanayiindeki veri madenciliği uygulamaları üzerine 1997'den 2007'ye kadar olan literatürü içeren kapsamlı bir inceleme yapmıştır. Choudhary ve ark. (2009) veri madenciliğinin üretim sistemlerindeki uygulamalarını özetlemiştir.

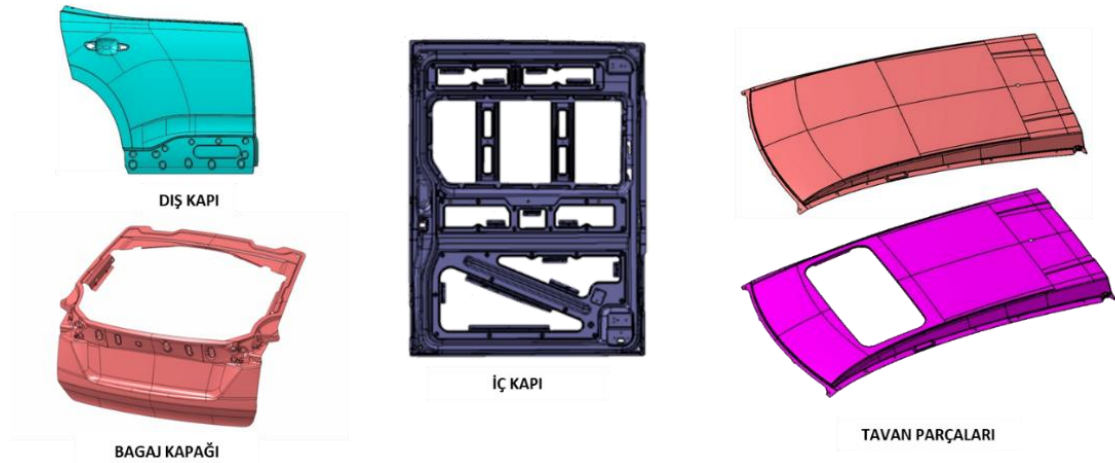
Siparişe dayalı imalat yapan sektörlerde üretim süresi tahmini için veri madenciliği yöntemlerinin kullanıldığı görülmektedir. Örneğin, gemi imalatında gerekli olan adam-saat tahmini için Hur ve ark. (2015) çoklu lineer regresyon modeli ile sınıflandırma ve regresyon ağacı modellerini karşılaştırmıştır. Azadeh ve ark. (2013), hem kesin hem de belirsiz ortamlardaki üretim sürelerinin tahmini için bulanık regresyon, geleneksel regresyon, yapay sinir ağı, adaptif ağ tabanlı bulanık çıkarım sistemi ve genetik algoritma

yöntemleri karşılaştırılmıştır. Lingitz ve ark. (2018), yarı iletken üretimi yapan bir firmada teslim süresi tahmini için makine öğrenmesi algoritmalarını karşılaştırmıştır. Bu algoritmalar doğrusal regresyon modelleri ve regresyon ağacı algoritmalarıdır. Tirkel (2013), veri tabanlarındaki bilgi keşfini kullanarak, yarı iletken üretiminde akış zamanının tahmini için karar ağacı modellerini ve sinir ağı modellerini karşılaştırmıştır.

Bu tezde, sac şekillendirme kalıbı üreten bir firmanın geçmiş üretim verilerini kullanarak üretim sürelerinin gerçekçi, hızlı ve sistematik bir şekilde belirlenmesi için önerilen yaklaşım kullanılmıştır. Literatüre bakıldığında kalıp imalat süresi tahmini için veri madenciliği yöntemlerini kullanan bir çalışmaya rastlanmamıştır.

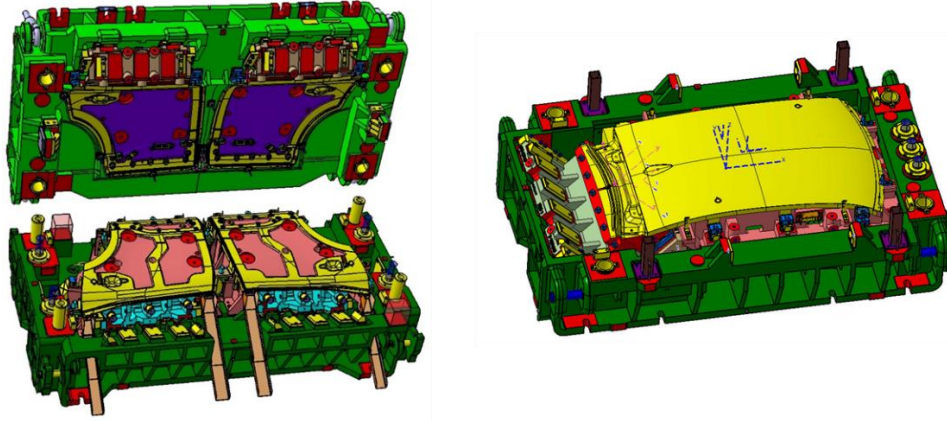
#### 4.5.1. Üretim ortamı

Vaka çalışması, otomotiv sektöründe yoğun olarak kullanılan yüksek mukavemetli sacların şekillendirilmesi için kalıp üreten bir firmada yapılmıştır. Firmada gövde ve şasi sac parçaları, aracın görsel dış ve iç panel parçaları (kapı, tavan, çamurluk), emniyet parçaları (direksiyon mili, arka aks, crash box, salıncak kolu) gibi birçok ürün için kalıp imalatı yapılmaktadır. Şekil 4.9’da firmanın kalıp imalatını yaptığı parçalara örnekler verilmiştir.



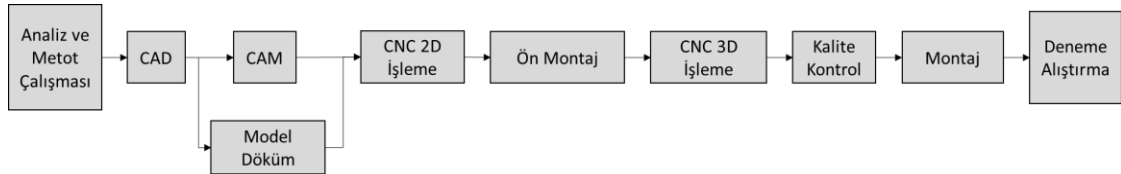
Şekil 4.9. Firmada üretilen sac şekillendirme kalıplarına ait parça örnekleri

Bir metal sac nihai parçaya dönüşene kadar, birçok kalıptan geçmektedir. Her bir kalıp üretilecek parçanın geometrine bağlı olarak farklı kalıp işlemleri (çekme, kesme, bükme, kamlı delme vb.) içermektedir. Şekil 4.10’da firmanın ürettiği kalıp örnekleri verilmiştir.



**Şekil 4.10.** Sac şekillendirme kalıp örnekleri

Kalıpların üretim sırasında yerine getirdiği işlemler farklı olsa da her bir kalıbın imalat süreci birbirine benzer adımlardan oluşmaktadır. Şekil 4.11’de temel kalıp imalat süreci gösterilmiştir.



**Şekil 4.11.** Temel kalıp imalat süreci

Kalıp imalatında, siparişe dayalı imalat yapılması ve yüksek ürün çeşitliliği nedeniyle üretim süreleri kalıptan kalıba değişmektedir. Firmada iki aşamalı üretim süresi tahmini yapılmaktadır. İlk aşamada müşteriden gelen parça bilgisine göre bir bütçe belirlenerek, siparişin alınabilmesi için teklif verilmektedir. Teklif hazırlanırken hem malzeme maliyetleri hem de işçilik maliyeti dikkate alınmaktadır. Teklif aşamasında doğru işçilik maliyeti tahmini yapmak rekabetçi teklifler verebilmesi için büyük avantaj sağlamaktadır. Kalıp imalatı müşteri standartlarına göre yapıldığı için malzeme temininde ortak pazar

kullanılmakta ve bu sebeple malzeme maliyeti rakip tekliflerle oldukça benzer olmaktadır. Firmaya büyük rekabet avantajı sağlayan kısım işçilik maliyetlerinin doğru öngörülmesidir. İşçilik maliyetinin tahmininde geçmişte tamamlanan kalıpların üretim süreleri uzman kişiler tarafından dikkate almaktadır. Üretim süresi ikinci aşama olarak üretim planlama çalışmalarında kullanılmaktadır. Üretim planlama çalışmaları, müşteriden siparişin gelmesi ile başlamaktadır. Müşteriden gelen parça verilerine göre öncelikle analiz ve metot çalışmaları yapılarak parçanın imalatı için gerekli olan kalıp sayısı ve kalıpların temel işlemleri ortaya çıkmaktadır. Sonrasında her bir kalıbın Bilgisayar Destekli Tasarımı (Computer Aided Design – CAD) belirli bir tasarımcı tarafından yapılmaktadır. Kalıp tasarımları nihai halini aldığı anda tedarikçilerde model ve döküm aşaması başlamaktadır. Firmada ise Bilgisayar Destekli İmalat (Computer Aided Manufacturing – CAM) çalışmaları başlamaktadır. CAM işlemlerinin tamamlanması ve parçaların dökümden gelmesi ile İki Boyutlu Bilgisayarlı Sayısal Kontrol (Two Dimensional Computer Numerical Control – CNC 2D) talaşlı imalat işlemleri yapılmaktadır. Üç Boyutlu Bilgisayarlı Sayısal Kontrol (Three Dimensional Computer Numerical Control – CNC 3D) talaşlı imalat işlemlerinden önce gerekirse parçaların ön montajı yapılmaktadır. CNC 3D işlemleri tamamlanan parçalar ölçülmektedir. Kalıpların pres altında yapılacak çalışmalara hazırlanması için kalıp parçalarının (alt gövde, üst gövde, çelikler, kamlar, tamponlar, sürtünme plakaları, gazlı yaylar vb.) montajı yapılmaktadır. Montaj çalışmalarının tamamlanması ile birlikte pres altında deneme alıştırılmaları yapılarak parçanın müşterinin istediği ölçüsel toleranslar içinde basılması sağlanmaktadır. Parça istenilen kalite standardına ulaşana kadar pres altında deneme alıştırma çalışmaları tekrarlanmaktadır.

Müşteriden gelen her bir siparişin farklı kalıp üretim sürelerinde yapılması ve proje periyotlarının birbirinden farklı olması üretim planlama çalışmalarının önemini daha da arttırmaktadır. Müşteri siparişlerinin karşılanması ve kalıpların zamanında yapılması için üretim sürelerinin doğru tahmin edilmesi gerekmektedir.



#### 4.5.2. Veri kümesi tanımlama

Mevcut sistemde, proje teklifleri ve üretim planları için gerekli olan üretim süreleri geçmişte yapılan çalışmaların uzman kişiler tarafından değerlendirilmesi ile belirlenmektedir.

Bu tezde kullanılan veri kümesi, Eser ve ark. (2019) tarafından yapılan ön çalışmada uzman bilgisi dikkate alınarak oluşturulmuştur. Öncelikle uzman bilgisine göre kalıp üretim sürelerini etkileyebilecek faktörler belirlenerek dört kategoriye ayrılmıştır. Bu kategoriler teknik özellikler (genişlik, uzunluk, yükseklik, ağırlık vb.), fonksiyonel özellikler (kesme kalıbı, çekme kalıbı, delme kalıbı vb.), üretim ortamının iş yükü (atölyedeki sipariş sayısı) ve firma tecrübesidir (beceri, deneyim, eğitim vb.). Kalıp üretim süresini etkileyen faktörlerin belirlenen kategorilere göre dağılımı Şekil 4.12’de verilmiştir.

Bu tezde vaka çalışması veri kümesi *Kalıp* veri kümesi olarak adlandırılmıştır. *Kalıp* veri kümesi firmanın 2015-2018 yılları arasındaki tamamladığı kalıpları içermektedir. *Kalıp* veri kümesinde 85 nesne, dört gerçek sayılı, yedi tam sayılı ve otuz sekiz nominal girdi değişkeni bulunmaktadır. Ayrıca çıktı değişkeni tamsayılardan oluşan sürekli değerler almaktadır.

Üretim Ortamının İş yükü	
Öznitelik	Öznitelik Türü
Minimum CNC işi sayısı	Tam sayı
Ortalama CNC işi sayısı	Gerçek sayı
Maksimum CNC işi sayısı	Tam sayı
Maksimum CAD işi sayısı	Tam sayı
Maksimum montaj işi sayısı	Tam sayı

Firma Tecrübesi	
Öznitelik	Öznitelik Türü
Tasarım Operatörü	Kategorik
Tasarım operatörünün üretim tecrübesi	Kategorik
Tasarım tecrübesi	Tam sayı
Montaj ekibi	Kategorik

Fonksiyonel Özellikler	
Öznitelik	Öznitelik Türü
Operasyon zorluğu	Kategorik
Parça transfer türü	Kategorik
Isıl işlem kontrolü	Kategorik
Isıl işlem türü	Kategorik
Revizyon sayısı	Tam sayı
Revizyon kontrol	Kategorik
İlave işçilik kontrol	Kategorik
Ölçüsel müdahale kontrolü	Kategorik
Açınım kesme	Kategorik
Çekme	Kategorik
Çevre kesme	Kategorik
Delme	Kategorik
Kamlı Delme	Kategorik
Kesme	Kategorik
Kamlı Kesme	Kategorik
Bükme	Kategorik
Kamlı Bükme	Kategorik
Kalibre	Kategorik

Teknik Özellikler	
Öznitelik	Öznitelik Türü
Parça grubu	Kategorik
Tasarım türü	Kategorik
Teslimat türü	Kategorik
Görsellik	Kategorik
Parça türü	Kategorik
Simetrik parça	Kategorik
Ek parça kontrolü	Kategorik
Ölçüsel zorunluluk	Kategorik
Sac mukavemeti	Kategorik
Sac alanı	Gerçek sayı
Parça Zorluğu	Kategorik
Tasarım revizyon sayısı	Tam sayı
Parçanın yapılma sayısı	Tam sayı
Kalıp yüksekliği	Gerçek sayı
Kalıp alanı	Gerçek sayı
Tampon sayısı	Tam sayı
Göz sayısı	Tam sayı
Tij kontrol	Kategorik
Tij / azot seçimi	Kategorik
Kalıp tesiri	Kategorik
Ayrılabilir dişi	Kategorik
İlave plaka kontrol	Kategorik
Isıl işlem kontrol	Kategorik
Bir sactan çıkan parça sayısı	Tam sayı

**Şekil 4.12.** Kalıp üretim süresine etki eden faktörlerin kategorilere ayrılması

#### 4.5.3. Vaka analizi bulguları

Sürekli değişkenler birbirinden farklı değer aralıklarında dağılmaktadır. Bu sebeple veri önışleme aşamasında öncelikle sürekli değişkenlere 0-1 normalizasyonu uygulanmıştır.

Kategorik deęişkenlerin her biri ikili deęişkenlere dönüştürülmüştür. *Kalıp* veri kümesinde eksik veri bulunmamaktadır. Aykırı veri analizi yapılarak veri kümesinde aykırı veri olmadığı tespit edilmiştir. Bu sebeple *Kalıp* veri kümesinde yapılan çalışmalarda aykırı deęer analizi öncesi ve sonrası ayırımı bulunmamaktadır. Ancak veri kümesinde deęişken sayısının çok olmasından dolayı öznitelik seçimi yapılarak model performansındaki deęişim incelenmiştir. Öznitelik seçimi yapılırken Eser ve ark. (2019) tarafından yapılan çalışmada incelenen modeller arasında en az bir model tarafından tercih edilen tüm öznitelikler dikkate alınmıştır.

Çizelge 4.11 ve Çizelge 4.12’de *Kalıp* veri kümesinin ağaç tabanlı yöntemler için MAPE ve RMSE sonuçları verilmiştir. *Kalıp* veri kümesinin komşuluk tabanlı yöntemler için MAPE ve RMSE sonuçları EK 14 ve EK 15’da verilmiştir.

**Çizelge 4.11.** *Kalıp* veri kümesinin ağaç tabanlı yöntemler için MAPE sonuçları (%)

		Kullanılan Ağırlıklı Tahmin Fonksiyonu								
Veri Kümesi	Model	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Kalıp</i>	RA	23,82	22,49	22,55	22,44	<b>22,43</b>	23,97	23,98	23,95	23,95
	TRA	25,06	24,62	24,63	24,62	24,6	25,12	25,14	25,11	25,09
	GRA	27,56	27,39	27,4	27,39	27,39	27,55	27,56	27,55	27,54

**Çizelge 4.12.** *Kalıp* veri kümesinin ağaç tabanlı yöntemler için RMSE sonuçları

		Kullanılan Ağırlıklı Tahmin Fonksiyonu								
Veri Kümesi	Model	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Kalıp</i>	RA	906,87	898,55	<b>898,51</b>	898,18	898,88	1009,51	1009,33	1009,25	1009,67
	TRA	966,73	964,36	964,77	964,7	964,06	978,17	978,58	978,31	977,73
	GRA	1052,27	1048,52	1048,66	1048,51	1048,28	1053,19	1053,32	1053,15	1053,02

Öznitelik seçimi sonrası veri kümesi *Kalıp-2* olarak adlandırılmıştır. *Kalıp-2* veri kümesinin ağaç tabanlı yöntemler için MAPE ve RMSE sonuçları Çizelge 4.13 ve Çizelge 4.14 verilirken, komşuluk tabanlı yöntemler için MAPE ve RMSE sonuçları Ek 16 ve Ek 17’de verilmiştir. *Kalıp* veri kümesi ve *Kalıp-2* veri kümesinin ağaç tabanlı yöntemlerde ve komşuluk tabanlı yöntemlerde en iyi performans gösteren modeller ve aldıkları deęerler Model ağacı yöntemi ile Çizelge 4.15’de karşılaştırılmıştır.

**Çizelge 4.13.** Kalıp-2 veri kümesinin ağaç tabanlı yöntemler için MAPE sonuçları (%)

		Kullanılan Ağırlıklı Tahmin Fonksiyonu								
Veri Kümesi	Model	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
Kalıp-2	RA	24,38	23,61	<b>23,6</b>	23,61	23,62	25,98	25,97	25,98	26,01
	TRA	24,6	24,18	24,17	24,19	24,19	24,67	24,67	24,68	24,68
	GRA	26,88	26,74	26,74	26,74	26,74	26,89	26,88	26,89	26,89

**Çizelge 4.14.** Kalıp-2 veri kümesinin ağaç tabanlı yöntemler için RMSE sonuçları

		Kullanılan Ağırlıklı Tahmin Fonksiyonu								
Veri Kümesi	Model	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
Kalıp-2	RA	915,85	909,45	<b>908,55</b>	910,35	910,32	1050,43	1049,68	1050,27	1051,14
	TRA	959,63	944,24	944,04	944,32	944,32	960,29	960,17	960,34	960,43
	GRA	1036,42	1031,04	1030,99	1030,94	1031,14	1036,18	1036,14	1036,09	1036,32

**Çizelge 4.15.** Model ağacı, ağaç tabanlı yöntemler ve komşuluk tabanlı yöntemlerin vaka çalışması için karşılaştırılması

	MAPE performans kriterine göre sonuçlar			RMSE performans kriterine göre sonuçlar		
	Model ağacı	Ağaç tabanlı modeller	Komşuluk tabanlı modeller	Model ağacı	Ağaç tabanlı modeller	Komşuluk tabanlı modeller
Kalıp	29,82	22,43 (RA-ATF 5)	27,5 (KNN-ATF 5)	1139,90	898,51 (RA-ATF 3)	1120,05 (KNN-ATF 3 KNN-ATF 4)
Kalıp-2	29,33	23,6 (RA-ATF 3)	25,83 (KNN-ATF 5)	1112,54	908,55 (RA-ATF 3)	1084,23 (TKNN-ATF 3)

Çizelge 4.15'e bakıldığında ağaç tabanlı yöntemlerde topluluk yaklaşımın iyi sonuç vermediği görülmüştür. Ancak nesnelere farklı ağırlıklar ile dikkate alan ağırlıklı tahmin fonksiyonlarının, eşit ağırlık tahmin fonksiyonuna (ATF1) göre daha iyi sonuç verdiği görülmüştür. Komşuluk tabanlı yöntemlerde topluluk yaklaşımının kullanılması belirgin bir üstünlük sağlamamıştır. Ağaç tabanlı yöntemlerde olduğu gibi komşuluk tabanlı yöntemler içinde en iyi sonuçlar önerilen ağırlıklı tahmin fonksiyonları ile sağlanmıştır. Hem ağaç tabanlı yöntemlerin hem de komşuluk tabanlı yöntemlerin Model ağacından daha iyi performans gösterdiği sonucuna ulaşılmıştır. Kalıp veri kümesinde nesne sayısının az olması topluluk yöntemlerde verinin iyi temsil edilememesine neden olmaktadır.

## 5. SONUÇ

Bu tezde tahmin problemlerinin çözümleri için komşuluk tabanlı regresyon ağacı yöntemleri kullanılmıştır. Sürekli çıktı değişkenine sahip veri kümelerine ağaç tabanlı yöntemler ve komşuluk tabanlı yöntemler uygulanmıştır. En iyi tahmin sonucunun elde edilebilmesi için aykırı değer analizinin, topluluk yöntemlerin ve ağırlıklı tahmin fonksiyonlarının model performansına etkileri incelenmiştir.

Karşılaştırmalı değerlendirme veri kümeleri kullanılarak elde edilen bulgular, ağaç tabanlı yöntemler içinde topluluk tabanlı GRA ve TRA algoritmalarının, RA algoritmasından hem MAPE hem de RMSE değerlerine göre daha iyi sonuç verdiğini göstermektedir. Komşuluk tabanlı yöntemler MAPE sonuçlarına göre değerlendirildiğinde KNN algoritmasının TKNN algoritmasına göre daha iyi performans gösterdiği görülürken, RMSE sonuçlarına göre TKNN algoritması daha iyi sonuç vermektedir. Bu sebeple komşuluk tabanlı yöntemler için topluluk yöntemlerin daha iyi sonuç verdiği kesin olarak söylenememektedir.

Komşuluk tabanlı ağırlıklı tahmin fonksiyonları ile geliştirilen algoritmaların performansları incelendiğinde topluluk yöntem kullanılması bağımsız olarak sadece ağırlıklı tahmin fonksiyonlarının kullanılması dahi tahmin sonuçlarını iyileştirdiği görülmüştür. Ayrıca komşuluk tabanlı ağırlıklı tahmin fonksiyonları ile geliştirilen algoritmaların bir karşılaştırma modeli olan Model ağacından ve literatürdeki temel versiyonlarından (ATF1) daha iyi sonuç verdiği görülmüştür.

Karşılaştırmalı değerlendirme veri kümelerinde aykırı değer analizi sonrası model performanslarının doğruluğunun çoğunlukla arttığı görülmüştür. Bu durum sürekli çıktı değişkenine sahip veri kümelerinde tahmin modeli geliştirirken aykırı değer analizinin önemini bir kez daha ortaya koymaktadır.

Yapılan vaka çalışması MAPE ve RMSE değerleri incelendiğinde torbalama regresyon ağacı ve güçlendirme regresyon ağacı yöntemlerinin tahmin performansını iyileştirmediği görülmüştür. Ancak ağırlıklı tahmin fonksiyonları kalıp veri kümesine uygulanan tüm

ağaç tabanlı ve komşuluk tabanlı yöntemlerde iyi sonuç vermiştir. Vaka analizindeki veri kümesi siparişe dayalı üretilen sınırlı sayıda kalıbı içermektedir. Topluluk yöntemlerde ise veri kümesinin alt kümelere bölünerek eğitiminin yapılması veri kümesinin temsil gücünü azaltmaktadır.

Sonuç olarak karşılaştırmalı değerlendirme veri kümelerinde hem topluluk yöntemlerin kullanılması hem de ağırlıklı tahmin fonksiyonlarının kullanılması tahmin doğruluğunu arttırmaktadır. Ağaç tabanlı yöntemlerde bir nesnenin tahmin değeri belirlenirken, aynı düğüme düşen nesnelerin komşuluk derecelerinin ve her bir nesnenin veri kümesi içindeki yerel aykırı değer faktörünün dikkate alınması tahmin problemlerinin çözümü için literatüre yeni bir yaklaşım kazandırmaktadır.

Tüm yaklaşımlar birlikte değerlendirildiğinde, tahmin problemlerinin çözümünde öncelikle aykırı değer analizi ile veri ön işleme yapılması önerilmektedir. Her veri kümesinde nesnelere ve öznitelikler arasında farklı ilişkiler bulunduğu için komşuluk tabanlı yöntemlerde tahmin sonuçlarını etkileyen önemli kriterlerden biri olan  $K$  değeri ayrıca ele alınmalıdır.

Gelecek çalışmalarda, komşuluk tabanlı ağırlıklandırma yaklaşımlarının karar ağaçlarında kullanımı ele alınabilir. Önerilen komşuluk tabanlı regresyon ağaçları budama mekanizması ile bütünleştirilebilir. Regresyon ağaçları oluşturulurken bölen öznitelik seçiminde komşuluk tabanlı yaklaşımlardan yararlanılabilir. Ayrıca kalıp veri kümesinde gürültü analizleri yapılarak ve veri kümesinin problemi temsil gücü artırılarak çalışmalar yapılabilir.

## KAYNAKLAR

- Abu Alfeilat, H.A., Hassanat, A.B., Lasassmeh, O., Tarawneh, A.S., Alhasanat, M.B., Eyal Salman, H.S., Prasath, V.S. 2019.** Effects of distance measure choice on K-Nearest neighbor classifier performance: A review. *Big data*, 7(4): 221-248.
- Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F. 2011.** Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic & Soft Computing*, 17.
- Alexander, W.P., Grimshaw, S.D. 1996.** Treed regression. *Journal of Computational and Graphical Statistics*, 5(2): 156-175.
- Alberg, D., Last, M., Kandel, A. 2012.** Knowledge discovery in data streams with regression tree methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1): 69-78.
- Aggarwal, C.C. 2015.** Outlier analysis. In *Data mining*. Springer, Cham, 237-263.
- Avnimelech, R., Intrator, N. 1999.** Boosting regression estimators. *Neural computation*, 11(2): 499-520.
- Azadeh, A., Ziaefar, A., Pichka, K., Asadzadeh, S. M. 2013.** An intelligent algorithm for optimum forecasting of manufacturing lead times in fuzzy and crisp environments. *International Journal of Logistics Systems and Management*, 16(2): 186-210.
- Bakar, Z.A., Mohemad, R., Ahmad, A., Deris, M.M. 2006.** A comparative study for outlier detection techniques in data mining. In 2006 IEEE conference on cybernetics and intelligent systems, June, 2006, 1-6.
- Barrash, S., Shen, Y., Giannakis, G.B. 2019.** Scalable and Adaptive KNN for Regression Over Graphs. In 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), December, 2019, IEEE, 241-245.
- Batista, G.E.A.P.A., Silva, D.F. 2009.** How k-nearest neighbor parameters affect its performance. In Argentine symposium on artificial intelligence, 1-12.
- Bay, S.D. 1998.** Combining Nearest Neighbor Classifiers Through Multiple Feature Subsets. In ICML, July, 1998, Vol. 98, 37-45.
- Bhatia, N. 2010.** Vandana. Survey of Nearest Neighbor Techniques. *Int J Comput Sci Inf Secur*, 8:302–305.
- Biau, G., Devroye, L., Dujmović, V., Krzyżak, A. 2012.** An affine invariant k-nearest neighbor regression estimate. *Journal of Multivariate Analysis*, 112, 24-34.

- Bramer, M. 2007.** Principles of data mining. Springer, London, 526.
- Breiman, L. 1994.** Bagging predictors. Technical Report 421, Department of Statistics University of California Berkeley.
- Breiman, L. 1996a.** Bagging predictors. *Machine learning*, 24(2): 123-140.
- Breiman, L. 1996b.** Heuristics of instability and stabilization in model selection. *The annals of statistics*, 24(6): 2350-2383.
- Breiman, L. 1999.** Prediction games and arcing algorithms. *Neural computation*, 11(7): 1493-1517.
- Breiman, L. 2001.** Random forests. *Machine learning*, 45(1): 5-32.
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A. 1984.** Classification and regression trees. *Pacific Grove Wadsworth: Belmont*.
- Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J. 2000.** LOF: identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD international conference on management of data, May, 2000, USA, 93-104.
- Brown, D.E., Pittard, C.L., Park, H. 1996.** Classification trees with optimal multivariate decision nodes. *Pattern Recognition Letters*, 17(7): 699-703.
- Chaudhuri, P., Huang, M.C., Loh, W.Y., Yao, R. 1994.** Piecewise-polynomial regression trees. *Statistica Sinica*, 143-167.
- Chauhan, P., Shukla, M. 2015.** A review on outlier detection techniques on data stream by using different approaches of K-Means algorithm. In 2015 International Conference on Advances in Computer Engineering and Applications, March, 2015, *IEEE*, 580-585.
- Chen, Z., Li, B., Han, B. 2017.** Improve regression accuracy by using an attribute weighted KNN approach. In 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), July, 2017, *IEEE*, 1838-1843.
- Cheng, Y., Chen, K., Sun, H., Zhang, Y., Tao, F. 2018.** Data and knowledge mining with big data towards smart production. *Journal of Industrial Information Integration*, 9: 1-13.
- Choudhary, A. K., Harding, J. A., Tiwari, M. K. 2009.** Data mining in manufacturing: a review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, 20(5): 501.
- Cover, T., Hart, P. 1967.** Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1): 21-27.



- Czajkowski, M., Kretowski, M. 2016.** The role of decision tree representation in regression problems—An evolutionary perspective. *Applied Soft Computing*, 48, 458-475.
- Demšar, J. 2006.** Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan), 1-30.
- Diao, L., Weng, C. 2019.** Regression tree credibility model. *North American Actuarial Journal*, 23(2): 169-196.
- Dogan, A., & Birant, D. 2020.** Machine Learning and Data Mining in Manufacturing. *Expert Systems with Applications*, 114060.
- Domeniconi, C., Yan, B. 2004.** Nearest neighbor ensemble. In Proceedings of the 17th International Conference on Pattern Recognition, August, 2004, IEEE, Vol. 1, 228-231.
- Domingues, R., Filippone, M., Michiardi, P., Zouaoui, J. 2018.** A comparative evaluation of outlier detection algorithms: Experiments and analyses. *Pattern Recognition*, 74: 406-421.
- Drucker, H. 1997.** Improving regressors using boosting techniques. In ICML, July, 1997, 107-115.
- Dudani, S.A. 1976.** The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics*, (4), 325-327.
- Dusseldorp, E., Conversano, C., Van Os, B.J. 2010.** Combining an additive and tree-based regression model simultaneously: STIMA. *Journal of Computational and Graphical Statistics*, 19(3): 514-530.
- Dusseldorp, E., Meulman, J.J. 2004.** The regression trunk approach to discover treatment covariate interaction. *Psychometrika*, 69(3): 355-374.
- Efron, B. 1983.** Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American statistical association*, 78(382): 316-331.
- Efron, B., Tibshirani, R. 1997.** Improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association*, 92(438): 548-560.
- Evelyn, F., Hodges, J.L. 1951.** Discriminatory analysis nonparametric discrimination: Consistency properties. Technical report.
- Eser, G., İnkaya, T., Ekdik, O. 2019.** Predicting the Production Times in Die Manufacturing Process Using Data Mining Methods. In 10th International Symposium on Intelligent Manufacturing and Service Systems, Sakarya, Turkey, September, 2019.
- Farrelly, C.M. 2017.** KNN Ensembles for Tweedie Regression: The Power of Multiscale Neighborhoods. arXiv preprint arXiv:1708.02122.

- Feely, R. 2000.** Predicting stock market volatility using neural networks. *BA (Mod) Dissertation*.
- Freund, Y., Schapire, R. 1996.** Experiment with a new boosting algorithm. In 13th International Conference on Machine Learning, Bari, Italy, July, 1996, 148-156.
- Freund, Y., Schapire, R. 1997.** A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55 (1): 119-139.
- Friedman, M. 1937.** The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, 32(200): 675-701.
- Friedman, M. 1940.** A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics*, 11(1): 86-92.
- Friedman, J., Hastie, T., Tibshirani, R. 2000.** Additive Logistic Regression: A Statistical View of Boosting. *The Annals of Statistics*, 28 (2): 337-374
- García, V., Sánchez, J.S., Marqués, A.I., Martínez-Peláez, R. 2018.** A regression model based on the nearest centroid neighborhood. *Pattern Analysis and Applications*, 21(4): 941-951.
- Gomes, C.M. A., Lemos, G.C., Jelihovschi, E.G. 2020.** Comparing the Predictive Power of the CART and CTREE algorithms. *Avaliação Psicológica*, 19(1): 87-96.
- Guyader, A., Hengartner, N. 2013.** On the mutual nearest neighbors estimate in regression. *The Journal of Machine Learning Research*, 14(1): 2361-2376.
- Haixiang, G., Yijing, L., Yanan, L., Xiao, L., Jinling, L. 2016.** BPSO-Adaboost-KNN ensemble learning algorithm for multi-class imbalanced data classification. *Engineering Applications of Artificial Intelligence*, 49: 176-193.
- Han, E.H.S., Karypis, G., Kumar, V. 2001.** Text categorization using weight adjusted k-nearest neighbor classification. In Pacific-asia conference on knowledge discovery and data mining, April, 2001, Berlin, Heidelberg, Springer, 53-65.
- Han, J., Pei, J., Kamber, M. 2011.** Data mining: concepts and techniques. Elsevier. 743 pp.
- Hawkins, D.M. 1980.** Identification of outliers. London, Chapman and Hall, Vol:11
- Hothorn, T., Hornik, K., Zeileis, A. 2006.** Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics*, 15(3): 651-674.

- Hur, M., Lee, S. K., Kim, B., Cho, S., Lee, D., Lee, D. 2015.** A study on the man-hour prediction system for shipbuilding. *Journal of Intelligent Manufacturing*, 26(6): 1267-1279.
- Johansson, U., Linusson, H., Löfström, T., Boström, H. 2018.** Interpretable regression trees using conformal prediction. *Expert systems with applications*, 97, 394-404.
- Kass, G.V. 1980.** An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(2): 119-127.
- Kim, H., Loh, W.Y. 2001.** Classification trees with unbiased multiway splits. *Journal of the American Statistical Association*, 96(454): 589-604.
- Kosasih, R., Fahrurrozi, A., Handhika, T., Sari, I., Lestari, D.P. 2020.** Travel Time Estimation for Destination In Bali Using kNN-Regression Method with Tensorflow. In IOP Conference Series: Materials Science and Engineering, May, 2020, Vol. 854, Vol.1, pp. 012061.
- Köksal, G., Batmaz, İ., Testik, M.C. 2011.** A review of data mining applications for quality improvement in manufacturing industry. *Expert systems with Applications*, 38(10): 13448-13467.
- Kramer, O. 2011.** Dimensionality reduction by unsupervised k-nearest neighbor regression. In 2011 10th International Conference on Machine Learning and Applications and Workshops, December, 2011, IEEE, 1:275-278.
- Lee, S.K., Kang, P., Cho, S. 2014.** Probabilistic local reconstruction for k-NN regression and its application to virtual metrology in semiconductor manufacturing. *Neurocomputing*, 131, 427-439.
- Li, D.F., Guan, W. 2020.** Algorithm Based on KNN and Multiple Regression for the Missing-Value Estimation of Sensors. *Journal of Highway and Transportation Research and Development*, 14(2): 7-15.
- Lingitz, L., Gallina, V., Ansari, F., Gyulai, D., Pfeiffer, A., Monostori, L. 2018.** Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer. *Procedia CIRP*, 72, 1051-1056.
- Loh, W.Y. 2009.** Improving the precision of classification trees. *The Annals of Applied Statistics*, 1710-1737.
- Loh, W.Y. 2011.** Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1): 14-23.
- Loh, W.Y. 2014.** Fifty years of classification and regression trees. *International Statistical Review*, 82(3): 329-348.

- Loh, W.Y., He, X., Man, M. 2015.** A regression tree approach to identifying subgroups with differential treatment effects. *Statistics in medicine*, 34(11): 1818-1833.
- Loh, W.Y., Shih, Y.S. 1997.** Split selection methods for classification trees. *Statistica sinica*, 815-840.
- Loh, W.Y., Vanichsetakul, N. 1988.** Tree-structured classification via generalized discriminant analysis. *Journal of the American Statistical Association*, 83(403): 715-725.
- Mac Namee, B., Cunningham, P., Byrne, S., Corrigan, O. I. 2002.** The problem of bias in training data in regression problems in medical decision support. *Artificial intelligence in medicine*, 24(1): 51-70.
- Morgan, J.N., Sonquist, J.A. 1963.** Problems in the analysis of survey data, and a proposal. *Journal of the American statistical association*, 58(302): 415-434.
- Morgan, J.N., Messenger, R.C. 1973.** THAID, a sequential analysis program for the analysis of nominal scale dependent variables.
- Nourali, H., Osanloo, M. 2020.** A regression-tree-based model for mining capital cost estimation. *International Journal of Mining, Reclamation and Environment*, 34(2): 88-100.
- Opitz, D., Maclin, R. 1999.** Popular ensemble methods: An empirical study. *Journal of artificial intelligence research*, 11, 169-198.
- Quinlan, J.R. 1992.** Learning with continuous classes. In 5th Australian joint conference on artificial intelligence, November, 1992, Australia, Vol. 92, 343-348.
- Quinlan, J.R. 1996.** Bagging, boosting and C4.5. In Proceedings of the 13th national conference on artificial intelligence, August, 1996, Portland, OR, USA, 725-730.
- Painsky, A., Rosset, S. 2016.** Cross-validated variable selection in tree-based methods improves predictive performance. *IEEE transactions on pattern analysis and machine intelligence*, 39(11): 2142-2153.
- Rana, P., Pahuja, D., Gautam, R. 2014.** A critical review on outlier detection techniques. *International Journal of Science and Research*, 3(12): 2394-2403.
- Schapire, R. 1990.** The strength of weak learnability. *Machine Learning*, 5(2): 197-227.
- Shrestha, D.L., Solomatine, D.P. 2006.** Experiments with AdaBoost. RT, an improved boosting scheme for regression. *Neural computation*, 18(7): 1678-1710.
- Song, Y., Huang, J., Zhou, D., Zha, H., Giles, C.L. 2007.** Iknn: Informative k-nearest neighbor pattern classification. In European Conference on Principles of Data Mining and Knowledge Discovery, September, 2007, Berlin, Heidelberg, Springer, 248-264.

- Su, Y., Ding, J. 2019.** Variable Grouping Based Bayesian Additive Regression Tree. arXiv preprint arXiv:1911.00922.
- Su, X., Wang, M., Fan, J. 2004.** Maximum likelihood regression trees. *Journal of Computational and Graphical Statistics*, 13(3): 586-598.
- Sutton, C.D. 2005.** Classification and regression trees, bagging, and boosting. *Handbook of statistics*, 24: 303-329.
- Tirkel, I. 2013.** Forecasting flow time in semiconductor manufacturing using knowledge discovery in databases. *International Journal of Production Research*, 51(18): 5536-5548.
- Torgo, L. 1997.** Functional models for regression tree leaves. In Proceedings of the Fourteenth International Conference on Machine Learning, July, 1997, Burlington, MA, 385-393.
- Taunk, K., De, S., Verma, S., Swetapadma, A. 2019.** A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. In 2019 International Conference on Intelligent Computing and Control Systems (ICCS), May, 2019, *IEEE*, 1255-1260.
- Wang, Y., Chaib-draa, B. 2016.** KNN-based Kalman filter: An efficient and non-stationary method for Gaussian process regression. *Knowledge-Based Systems*, 114: 148-155.
- Wang, Y., Witten, I.H. 1996.** Induction of model trees for predicting continuous classes. *Working paper series, Department of Computer Science*, University of Waikato.
- Weichert, D., Link, P., Stoll, A., Rüping, S., Ihlenfeldt, S., Wrobel, S. 2019.** A review of machine learning for the optimization of production processes. *The International Journal of Advanced Manufacturing Technology*, 1-14.
- Wettschereck, D., Dietterich, T.G. 1994.** Locally adaptive nearest neighbor algorithms. *In Advances in Neural Information Processing Systems*, 184-191.
- Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.H., Steinbach, M., Hand, D.J., Steinberg, D. 2008.** Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1): 1-37.
- Yao, H., Fu, X., Yang, Y., Postolache, O. 2018.** An incremental local outlier detection method in the data stream. *Applied Sciences*, 8(8): 1248.
- You, L., Peng, Q., Xiong, Z., He, D., Qiu, M., Zhang, X. 2020.** Integrating aspect analysis and local outlier factor for intelligent review spam detection. *Future Generation Computer Systems*, 102: 163-172.

**Zemel, R. S., Pitassi, T. 2001.** A gradient-based boosting algorithm for regression problems. *In Advances in neural information processing systems*, 696-702.

## EKLER

- EK 1** Önerilen regresyon ağacı ile tahmin modeli algoritması
- EK 2** Önerilen torbalama regresyon ağacı ile tahmin modeli algoritması
- EK 3** Önerilen güçlendirme regresyon ağacı ile tahmin modeli algoritması
- EK 4** Önerilen KNN algoritması
- EK 5** Önerilen Torbalama KNN algoritması
- EK 6** Orijinal veri kümelerinde KNN ve TKNN için MAPE değerleri (%)
- EK 7** Orijinal veri kümelerinde KNN ve TKNN için RMSE değerleri
- EK 8** Aykırı değer analizi yapılan veri kümelerinde KNN ve TKNN için MAPE değerleri (%)
- EK 9** Aykırı değer analizi yapılan veri kümelerinde KNN ve TKNN için RMSE değerleri
- EK 10** Orijinal veri kümelerine komşuluk tabanlı yöntemlerin uygulanması ile elde edilen MAPE sonuçlarına göre  $r_i^j$  değerlerinin dağılımı
- EK 11** Orijinal veri kümelerine komşuluk tabanlı yöntemlerin uygulanması ile elde edilen RMSE sonuçlarına göre  $r_i^j$  değerlerinin dağılımı
- EK 12** Aykırı değer analizi yapılan veri kümelerine komşuluk tabanlı yöntemlerin uygulanması ile elde edilen MAPE sonuçlarına göre  $r_i^j$  değerlerinin dağılımı
- EK 13** Aykırı değer analizi yapılan veri kümelerine komşuluk tabanlı yöntemlerin uygulanması ile elde edilen RMSE sonuçlarına göre  $r_i^j$  değerlerinin dağılımı
- EK 14** *Kalıp* veri kümesinin komşuluk tabanlı yöntemler için MAPE sonuçları (%)
- EK 15** *Kalıp* veri kümesinin komşuluk tabanlı yöntemler için RMSE sonuçları
- EK 16** Öznitelik seçimi sonrası *Kalıp 2* veri kümesinin komşuluk tabanlı yöntemler için MAPE sonuçları (%)
- EK 17** Öznitelik seçimi sonrası *Kalıp 2* veri kümesinin komşuluk tabanlı yöntemler için RMSE sonuçlar

## EK 1 Önerilen Regresyon Ağacı ile Tahmin Modeli Algoritması

---

**Algoritma 8:** Önerilen regresyon ağacı ile tahmin modelinin kurulması

---

**Girdi:**  $D$  veri kümesi, *uzaklık matrisi* (Algoritma 6), *sıralama matrisi* (Algoritma 6), önerilen regresyon ağacı modeli (Algoritma 7), performans değerlendirme yöntemi (LOOCV), tahmin fonksiyonları

**Çıktı:** Regresyon ağacı tahmin modeli

**Yöntemin adımları:**

- 1) **For**  $b = 1$  to  $n$  **Do**
  - 2)  $D_b \leftarrow D \setminus \{x_b\}$  eğitim veri kümesini oluştur.
  - 3)  $D_b^T \leftarrow x_b$  test veri kümesini oluştur.
  - 4)  $D_b$  eğitim veri kümesi ile Algoritma 7’de verilen önerilen regresyon ağacını oluştur.
  - 5)  $x_b$ ’nin düştüğü yaprak düğümü belirle.
  - 6) Ağırlıklı tahmin fonksiyonlarını kullanarak  $x_b$  için tahmin değerini  $y_b$  hesapla.
  - 7) **End For**
  - 8) Her bir ağırlıklı tahmin fonksiyonu için performans metriklerini (MAPE ve RMSE) hesapla.
-



## EK 2 Önerilen Torbalama Regresyon Ağacı ile Tahmin Modeli Algoritması

---

**Algoritma 9:** Önerilen torbalama regresyon ağacı ile tahmin modelinin kurulması

---

**Girdi:**  $D$  veri kümesi, *uzaklık matrisi* (Algoritma 6), *sıralama matrisi* (Algoritma 6), önerilen regresyon ağacı modeli (Algoritma 7), performans değerlendirme yöntemi (0,632 önyükleme), tahmin fonksiyonları,  $B$

**Çıktı:** Torbalama Regresyon ağacı tahmin modeli

**Yöntem adımları:**

- 1) **For**  $b = 1$  to  $B$  **Do**
  - 2)      $D_b$ , eğitim veri kümesini 0,632 önyükleme yöntemi ile oluştur.
  - 3)      $D_b^T \leftarrow D \setminus \{D_b\}$  test veri kümesini oluştur.
  - 4)      $D_b$  eğitim veri kümesi ile Algoritma 7’de verilen önerilen regresyon ağacını oluştur.
  - 5)     Modelin eğitim hatasını  $Acc(M_b)_{D_b}$  hesapla.
  - 6)     **For**  $i = 1$  to  $n^T$  **Do**
  - 7)          $x_i$  nesnesinin düştüğü yaprak düğümü belirle.
  - 8)         Ağırlıklı tahmin fonksiyonlarını kullanarak  $x_i$  için tahmin değerini  $y_i$  hesapla.
  - 9)     **End For**
  - 10)    **End For**
  - 11) **End For**
  - 12) Her bir ağırlıklı tahmin fonksiyonu için performans metriklerini (MAPE ve RMSE) hesapla.
-

### EK 3 Önerilen Güçlendirme Regresyon Ağacı ile Tahmin Modeli Algoritması

---

**Algoritma 10:** Önerilen güçlendirme regresyon ağacı ile tahmin modelinin kurulması

---

**Girdi:**  $D$  veri kümesi, *uzaklık matrisi* (Algoritma 6), *sıralama matrisi* (Algoritma 6), önerilen regresyon ağacı modeli (Algoritma 7), performans değerlendirme yöntemi (0,632 önyükleme), tahmin fonksiyonları,  $k, \varphi, \theta$

**Çıktı:** Güçlendirme Regresyon ağacı tahmin modeli

**Yöntem adımları:**

- 1) **For**  $i = 1$  to  $n$  **Do**
  - 2)      $w_1(x_i) = 1/n$
  - 3) **End For**
  - 4) **For**  $b = 1$  to  $B$  **Do**
  - 5)      $D_b$ , eğitim veri kümesini ve  $w_b(x)$  nesne ağırlıklarını kullanarak 0,632 önyükleme yöntemi ile oluştur.
  - 6)      $D_b^T \leftarrow D \setminus \{D_b\}$  test veri kümesini oluştur.
  - 7)      $D_b$  eğitim veri kümesi ile Algoritma 7'de verilen önerilen regresyon ağacını oluştur.
  - 8)      $\varepsilon_b = 0$
  - 9)     **For**  $i = 1$  to  $n$  **Do**
  - 10)         **If**  $x_i \in D_b$  **Then**
  - 11)              $\varepsilon_b = \varepsilon_b + w_b(x_i)$
  - 12)         **End If**
  - 13)     **End For**
  - 14)      $\beta_b = \varepsilon_b^2$
  - 15)     **If**  $0 < \varepsilon_b < \theta$  **Then**
  - 16)         **For**  $i = 1$  to  $n$  **Do**  $\parallel x_i \in D_b$  için
  - 17)             **If**  $\left| \frac{(x_{im} - y_i)}{x_{im}} \right| \leq \varphi$  **Then**
  - 18)                  $w_b(x_i) = w_b(x_i) \times \beta_b$
  - 19)             **Else**
  - 20)                  $w_b(x_i) = w_b(x_i) \times 1$
  - 21)             **End If**
  - 22)         **End For**
  - 23)          $w_b(x_i)$  ağırlıklarını normalize et
  - 24)          $W = \beta_b$
  - 25)         **Break**
  - 26)     **Else**
  - 27)          $w_b(x_i) = 1/n$
  - 28)     **End If**
  - 29)     **For**  $i = 1$  to  $n^T$  **Do**
  - 30)          $x_i$  nesnesinin düştüğü yaprak düğümü belirle.
  - 31)         Ağırlıklı tahmin fonksiyonlarını kullanarak  $x_i$  için tahmin değerini  $y_i$  hesapla.
  - 32)     **End For**
  - 33) **End For**
  - 34) Her bir ağırlıklı tahmin fonksiyonu için performans metriklerini (MAPE ve RMSE) hesapla.
-

## EK 4 Önerilen KNN Algoritması

---

### Algoritma 11: Önerilen KNN algoritması

---

**Girdi:**  $D$  veri kümesi, *uzaklık matrisi* (Algoritma 6), *sıralama matrisi* (Algoritma 6),  $K$

**Çıktı:** Önerilen KNN algoritması

**Yöntem adımları:**

- 1) **For**  $b = 1$  to  $n$  **Do**
  - 2)      $D_b \leftarrow D \setminus \{x_b\}$  eğitim veri kümesi oluştur.
  - 3)      $D_b^T \leftarrow x_b$  test veri kümesi oluştur.
  - 4)     **For**  $j = 1$  to  $K$  **Do**
  - 5)          $x_b$ 'nin  $j$  tane en yakın komşusunu belirle.
  - 6)         Ağırlıklı tahmin fonksiyonlarını kullanarak  $x_b$  için tahmin değerini  $y_b$  hesapla.
  - 7)     **End For**
  - 8) **End For**
  - 9) Her bir ağırlıklı tahmin fonksiyonu için performans metriklerini (MAPE ve RMSE) hesapla.
-

## EK 5 Önerilen Torbalama KNN Algoritması

---

### Algoritma 12: Önerilen Torbalama KNN algoritması

---

**Girdi:**  $D$  veri kümesi, *uzaklık matrisi* (Algoritma 6), *sıralama matrisi* (Algoritma 6),  $K$ ,  $B$

**Çıktı:** Önerilen Torbalama KNN algoritması

**Yöntem adımları:**

- 1) **For**  $b = 1$  to  $B$  **Do**
  - 2)      $D_b$ , eğitim veri kümesi 0,632 önyükleme yöntemi ile oluştur.
  - 3)      $D_b^T \leftarrow D \setminus \{D_b\}$  test veri kümesini oluştur.
  - 4)     **For**  $j = 1$  to  $K$  **Do**
  - 5)         **For**  $i = 1$  to  $n^T$  **Do**
  - 6)              $x_i$ 'nin  $j$  tane en yakın komşusunu belirle.
  - 7)             Ağırlıklı tahmin fonksiyonlarını kullanarak  $x_i$  için tahmin değerini  $y_i$  hesapla.
  - 8)         **End For**
  - 9)     **End For**
  - 10) **End For**
  - 11) Her bir ağırlıklı tahmin fonksiyonu için performans metriklerini (MAPE ve RMSE) hesapla.
-

**EK 6 Orijinal Veri Kümelerinde KNN ve TKNN için MAPE Değerleri (%)**

Veri Kümesi	Model	K	Kullanılan Ağırlıklı Tahmin Fonksiyonu								
			ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Diabetes</i>	KNN	1	12,15	12,15	12,15	12,15	15,17	12,15	12,15	12,15	12,15
		2	10,06	10,20	11,24	11,24	14,41	10,40	10,40	10,40	10,40
		3	9,29	9,58	10,76	10,76	14,09	9,72	9,72	9,72	9,72
		4	9,73	9,71	10,66	10,66	13,90	9,93	9,93	9,93	9,93
		5	9,30	9,41	10,46	10,46	13,78	9,40	9,40	9,40	9,40
		6	8,88	9,30	10,31	10,31	13,40	9,29	9,29	9,29	9,29
		7	8,61	8,95	9,97	9,97	13,36	8,97	8,97	8,97	8,97
		8	8,62	8,91	9,89	9,89	13,26	9,00	9,00	9,00	9,00
		9	8,76	9,02	9,97	9,97	12,92	9,12	9,12	9,12	9,12
		10	8,88	9,12	10,02	10,02	12,98	9,18	9,18	9,18	9,18
	TKNN	1	10,21	10,21	10,21	10,21	10,21	10,21	10,21	10,21	10,21
		2	9,49	9,59	9,90	9,60	9,61	9,83	9,86	9,66	9,71
		3	9,24	9,40	9,80	9,19	9,29	9,30	9,58	9,12	9,21
		4	9,09	9,26	9,34	9,11	9,15	9,31	9,43	9,13	9,23
		5	8,93	9,24	9,39	9,06	9,07	9,12	9,38	9,15	9,19
		6	9,10	9,23	9,26	9,00	8,98	9,22	9,31	9,21	9,20
		7	9,24	9,31	9,23	9,02	9,03	9,08	9,33	8,98	9,20
		8	9,40	9,25	9,30	8,93	9,12	9,16	9,41	9,12	9,30
		9	9,30	9,21	9,36	9,13	9,17	9,22	9,30	9,21	9,40
		10	9,25	9,24	9,36	9,01	9,14	9,26	9,49	9,09	9,30
<i>MachineCPU</i>	KNN	1	39,11	39,11	39,11	39,11	38,79	39,11	39,11	39,11	39,11
		2	36,41	36,25	36,80	36,80	36,99	37,24	37,24	37,24	37,24
		3	36,18	35,64	35,86	35,86	36,25	36,22	36,22	36,22	36,22
		4	36,73	35,66	35,03	35,03	35,55	36,61	36,61	36,61	36,61
		5	37,42	35,41	34,77	34,79	35,25	36,54	36,54	36,54	36,54
		6	37,25	35,24	34,48	34,48	35,23	36,31	36,31	36,31	36,31
		7	36,49	34,85	34,45	34,45	34,91	35,75	35,75	35,75	35,75
		8	36,24	34,77	34,35	34,35	34,78	35,65	35,65	35,65	35,65
		9	35,64	34,32	34,02	34,02	34,67	35,44	35,44	35,44	35,44
		10	36,10	34,47	33,85	33,85	34,78	35,33	35,33	35,33	35,33
	TKNN	1	35,11	35,11	35,11	35,11	35,11	35,11	35,11	35,11	35,11
		2	35,59	35,17	35,29	35,12	35,06	35,16	35,21	35,19	35,12
		3	36,55	35,40	35,47	35,46	35,49	35,25	35,07	35,03	35,19
		4	36,38	35,37	35,38	35,18	35,39	34,82	34,74	34,72	34,95
		5	36,22	35,01	35,16	34,91	35,10	34,43	34,35	34,44	34,67
		6	36,12	34,95	35,01	34,74	35,12	34,35	34,09	34,13	34,40
		7	36,11	34,86	34,98	34,93	35,05	34,15	34,03	34,03	34,27
		8	36,42	35,03	35,04	34,83	35,04	34,39	34,19	34,23	34,45
		9	36,58	35,20	35,17	34,97	35,22	34,41	34,40	34,36	34,79
		10	37,07	35,26	35,24	35,18	35,34	34,71	34,43	34,64	34,90

**EK 6 Orijinal Veri Kümelerinde KNN ve TKNN için MAPE Değerleri (%)**  
(Devamı)

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Baseball</i>	KNN	1	24,95	24,95	24,95	24,95	23,37	24,95	24,95	24,95	24,95
		2	27,12	26,53	24,85	24,85	23,43	26,60	26,60	26,60	26,64
		3	26,33	25,71	24,38	24,38	23,06	25,62	25,62	25,62	25,62
		4	27,83	26,40	24,61	24,61	23,18	25,33	25,33	25,33	25,31
		5	29,48	27,11	24,83	24,83	23,28	25,60	25,60	25,60	25,60
		6	28,49	26,69	24,09	24,09	22,75	25,07	25,07	25,07	25,11
		7	29,53	27,41	24,52	24,52	23,01	25,63	25,63	25,63	25,63
		8	29,98	27,69	24,70	24,70	23,28	26,12	26,12	26,12	26,14
		9	30,54	27,85	24,91	24,91	23,33	26,45	26,45	26,45	26,45
		10	31,27	28,30	25,10	25,10	23,54	26,79	26,79	26,79	26,80
	TKNN	1	25,87	25,87	25,87	25,87	25,87	25,87	25,87	25,87	25,87
		2	28,12	27,55	27,93	27,06	27,14	27,34	27,63	26,90	27,06
		3	29,00	28,07	28,41	27,62	27,88	27,25	27,61	26,81	26,97
		4	29,90	28,58	28,82	28,16	28,47	27,52	27,75	27,08	27,27
		5	30,63	28,89	29,23	28,58	28,84	28,00	28,28	27,60	27,88
		6	31,64	29,46	29,73	29,15	29,38	28,61	28,81	28,25	28,51
		7	32,48	30,00	30,19	29,54	29,95	29,06	29,25	28,84	28,99
		8	33,21	30,36	30,59	29,98	30,40	29,38	29,53	29,08	29,37
		9	33,77	30,71	30,89	30,37	30,68	29,56	29,72	29,28	29,62
		10	34,30	30,93	31,16	30,63	31,01	29,79	30,00	29,52	29,82
<i>Dee</i>	KNN	1	10,48	10,48	10,48	10,48	11,08	10,48	10,48	10,48	10,48
		2	9,49	9,47	9,92	9,92	10,57	9,44	9,44	9,44	9,44
		3	8,84	8,90	9,52	9,52	10,25	8,80	8,80	8,80	8,80
		4	8,44	8,48	9,22	9,22	10,00	8,37	8,37	8,37	8,37
		5	8,31	8,30	9,02	9,02	9,85	8,23	8,23	8,23	8,23
		6	8,22	8,23	8,90	8,90	9,75	8,18	8,18	8,18	8,18
		7	8,21	8,18	8,81	8,81	9,70	8,20	8,20	8,20	8,20
		8	8,11	8,12	8,71	8,71	9,59	8,13	8,13	8,13	8,13
		9	8,11	8,09	8,68	8,68	9,55	8,13	8,13	8,13	8,13
		10	8,18	8,09	8,64	8,64	9,50	8,16	8,16	8,16	8,16
	TKNN	1	9,21	9,21	9,21	9,21	9,21	9,21	9,21	9,21	9,21
		2	8,58	8,70	8,78	8,62	8,59	8,57	8,65	8,50	8,48
		3	8,29	8,39	8,50	8,33	8,32	8,25	8,33	8,20	8,19
		4	8,19	8,27	8,38	8,21	8,19	8,15	8,23	8,10	8,10
		5	8,14	8,22	8,30	8,15	8,15	8,12	8,18	8,06	8,06
		6	8,15	8,18	8,27	8,12	8,12	8,10	8,17	8,06	8,05
		7	8,19	8,17	8,26	8,12	8,11	8,11	8,19	8,07	8,07
		8	8,25	8,18	8,25	8,12	8,12	8,15	8,23	8,10	8,09
		9	8,29	8,18	8,27	8,13	8,12	8,18	8,25	8,14	8,12
		10	8,33	8,19	8,27	8,13	8,13	8,20	8,28	8,16	8,15

**EK 6 Orijinal Veri Kümelerinde KNN ve TKNN için MAPE Değerleri (%)**  
(Devamı)

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9	
<i>AutoMpg</i>	KNN	1	73,04	73,04	73,04	73,04	63,84	73,04	73,04	73,04	73,04	
		2	69,97	69,51	69,99	69,99	60,75	69,54	69,54	69,54	69,54	
		3	69,68	68,88	69,14	69,14	59,98	69,40	69,40	69,40	69,40	
		4	69,03	67,96	68,02	68,02	59,08	68,90	68,90	68,90	68,90	
		5	68,13	67,37	67,45	67,45	58,42	68,15	68,15	68,15	68,15	
		6	68,58	67,47	67,39	67,39	58,45	68,42	68,42	68,42	68,42	
		7	67,82	66,94	67,02	67,02	58,06	67,60	67,60	67,60	67,60	
		8	67,98	67,10	67,00	67,00	57,99	67,75	67,75	67,75	67,75	
		9	67,23	66,71	66,68	66,68	57,75	67,11	67,11	67,11	67,11	
		10	66,22	66,28	66,35	66,35	57,43	66,34	66,34	66,34	66,34	
	TKNN	1	67,81	67,81	67,81	67,81	67,81	67,81	67,81	67,81	67,81	67,81
		2	67,51	67,24	67,48	67,16	67,00	67,36	67,65	67,27	67,10	
		3	67,66	67,05	67,49	66,92	66,63	67,52	67,99	67,38	67,09	
		4	67,58	66,94	67,46	66,78	66,43	67,44	68,04	67,25	66,87	
		5	67,13	66,68	67,26	66,54	66,13	67,01	67,67	66,85	66,39	
		6	66,91	66,56	67,16	66,45	66,00	66,78	67,45	66,64	66,16	
		7	66,45	66,33	66,92	66,22	65,76	66,34	67,00	66,24	65,73	
		8	66,42	66,24	66,85	66,15	65,69	66,28	66,91	66,19	65,71	
		9	66,60	66,25	66,86	66,17	65,70	66,36	66,95	66,26	65,82	
		10	66,78	66,27	66,87	66,21	65,72	66,45	67,01	66,36	65,94	
<i>Laser</i>	KNN	1	8,77	8,77	8,77	8,77	11,84	8,77	8,77	8,77	8,77	
		2	8,33	8,18	8,34	8,34	11,25	8,31	8,31	8,31	8,30	
		3	8,06	7,89	8,08	8,08	11,01	8,24	8,24	8,24	8,24	
		4	7,94	7,74	7,89	7,89	10,80	8,18	8,18	8,18	8,18	
		5	7,98	7,72	7,79	7,79	10,66	8,23	8,23	8,23	8,23	
		6	8,16	7,75	7,73	7,73	10,58	8,36	8,36	8,36	8,36	
		7	8,31	7,83	7,73	7,73	10,55	8,46	8,46	8,46	8,46	
		8	8,21	7,79	7,69	7,69	10,51	8,41	8,41	8,41	8,41	
		9	8,27	7,82	7,67	7,67	10,48	8,45	8,45	8,45	8,45	
		10	8,40	7,86	7,68	7,68	10,45	8,51	8,51	8,51	8,51	
	TKNN	1	7,90	7,90	7,90	7,90	7,90	7,90	7,90	7,90	7,90	
		2	7,85	7,71	7,71	7,70	7,73	7,76	7,77	7,74	7,77	
		3	7,92	7,69	7,69	7,67	7,72	7,83	7,82	7,79	7,84	
		4	8,06	7,74	7,75	7,71	7,77	7,91	7,92	7,88	7,93	
		5	8,19	7,80	7,80	7,76	7,82	8,02	8,03	7,97	8,03	
		6	8,27	7,84	7,84	7,80	7,86	8,08	8,11	8,05	8,10	
		7	8,32	7,86	7,85	7,81	7,89	8,14	8,14	8,09	8,15	
		8	8,37	7,89	7,87	7,84	7,90	8,20	8,19	8,14	8,21	
		9	8,45	7,91	7,90	7,87	7,94	8,24	8,24	8,20	8,25	
		10	8,53	7,95	7,93	7,90	7,97	8,29	8,30	8,27	8,31	

**EK 6 Orijinal Veri Kümelerinde KNN ve TKNN için MAPE Değerleri (%)**  
(Devamı)

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Concrete</i>	KNN	1	21,30	21,30	21,30	21,30	23,18	21,30	21,30	21,30	21,30
		2	22,81	21,16	22,37	22,37	22,96	20,13	20,13	20,13	20,13
		3	22,08	20,77	21,72	21,72	22,16	19,59	19,59	19,59	19,59
		4	22,29	20,75	21,63	21,63	21,92	19,50	19,50	19,50	19,50
		5	23,25	21,06	21,68	21,68	21,96	19,56	19,56	19,56	19,56
		6	23,82	21,35	21,65	21,65	21,88	19,57	19,57	19,57	19,57
		7	24,16	21,51	21,67	21,67	21,89	19,67	19,67	19,67	19,67
		8	24,45	21,66	21,73	21,73	21,95	19,67	19,67	19,67	19,67
		9	25,05	21,94	21,84	21,84	22,03	19,76	19,76	19,76	19,76
		10	25,39	22,09	21,91	21,91	22,10	19,89	19,89	19,89	19,89
	TKNN	1	23,50	23,50	23,50	23,50	23,50	23,50	23,50	23,50	23,50
		2	23,39	23,18	23,20	23,06	23,18	22,16	22,21	22,10	22,12
		3	23,97	23,39	23,43	23,25	23,36	22,14	22,21	22,07	22,10
		4	24,76	23,78	23,81	23,63	23,76	22,28	22,33	22,20	22,26
		5	25,34	24,07	24,11	23,92	24,07	22,44	22,48	22,36	22,43
		6	25,88	24,36	24,41	24,22	24,35	22,64	22,69	22,57	22,62
		7	26,37	24,61	24,68	24,48	24,59	22,82	22,89	22,74	22,80
		8	26,83	24,84	24,92	24,71	24,81	23,02	23,11	22,94	23,00
		9	27,28	25,05	25,14	24,92	25,02	23,23	23,31	23,13	23,20
		10	27,73	25,25	25,35	25,12	25,22	23,41	23,49	23,32	23,39
<i>Ele2</i>	KNN	1	19,42	19,42	19,42	19,42	20,39	19,42	19,42	19,42	19,42
		2	19,62	19,33	19,37	19,37	20,25	19,58	19,58	19,58	19,58
		3	20,92	19,52	19,91	19,91	20,66	19,30	19,30	19,30	19,30
		4	22,58	20,49	20,63	20,63	21,22	20,07	20,07	20,07	20,07
		5	23,59	21,00	20,95	20,95	21,50	20,54	20,54	20,54	20,54
		6	25,09	21,56	21,49	21,49	21,99	22,14	22,14	22,14	22,14
		7	25,84	21,92	21,79	21,79	22,25	22,85	22,85	22,85	22,85
		8	28,28	22,58	22,45	22,45	22,87	25,08	25,08	25,08	25,08
		9	29,00	22,93	22,80	22,80	23,19	25,84	25,84	25,84	25,84
		10	28,02	22,85	22,69	22,69	23,10	24,73	24,73	24,73	24,73
	TKNN	1	21,65	21,65	21,65	21,65	21,65	21,65	21,65	21,65	21,65
		2	23,35	22,68	22,69	22,64	22,66	21,82	21,86	21,79	21,78
		3	26,07	24,10	24,14	24,04	24,06	23,96	24,02	23,92	23,91
		4	27,60	25,02	25,05	24,97	24,99	25,34	25,33	25,33	25,34
		5	28,35	25,52	25,51	25,48	25,53	25,94	25,84	25,96	26,02
		6	28,17	25,57	25,52	25,54	25,61	25,61	25,45	25,66	25,76
		7	27,49	25,38	25,30	25,37	25,44	24,84	24,64	24,92	25,04
		8	26,92	25,19	25,11	25,19	25,27	24,20	23,99	24,28	24,40
		9	26,47	25,05	24,95	25,05	25,13	23,71	23,50	23,78	23,90
		10	26,16	24,95	24,86	24,95	25,03	23,35	23,16	23,41	23,52



**EK 6 Orijinal Veri Kümelerinde KNN ve TKNN için MAPE Değerleri (%)**  
(Devamı)

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Friedman</i>	KNN	1	15,24	15,24	15,24	15,24	15,44	15,24	15,24	15,24	15,24
		2	13,16	13,06	14,04	14,04	14,38	13,05	13,05	13,05	13,05
		3	12,41	12,23	13,39	13,39	13,76	12,26	12,26	12,26	12,26
		4	12,18	11,93	13,05	13,05	13,44	11,98	11,98	11,98	11,98
		5	11,90	11,67	12,78	12,78	13,19	11,69	11,69	11,69	11,69
		6	11,87	11,60	12,62	12,62	13,01	11,65	11,65	11,65	11,65
		7	11,71	11,48	12,47	12,47	12,86	11,50	11,50	11,50	11,50
		8	11,58	11,37	12,33	12,33	12,73	11,39	11,39	11,39	11,39
		9	11,62	11,35	12,23	12,23	12,66	11,40	11,40	11,40	11,40
		10	11,69	11,36	12,17	12,17	12,59	11,46	11,46	11,46	11,46
	TKNN	1	12,67	12,67	12,67	12,67	12,67	12,67	12,67	12,67	12,67
		2	11,83	11,87	11,86	11,86	11,89	11,77	11,76	11,77	11,80
		3	11,67	11,62	11,59	11,61	11,65	11,55	11,53	11,55	11,59
		4	11,56	11,49	11,47	11,49	11,53	11,43	11,40	11,43	11,47
		5	11,56	11,42	11,39	11,42	11,46	11,41	11,36	11,41	11,46
		6	11,64	11,40	11,35	11,40	11,45	11,46	11,39	11,46	11,52
		7	11,78	11,43	11,37	11,43	11,48	11,56	11,49	11,57	11,64
		8	11,95	11,48	11,42	11,49	11,55	11,70	11,62	11,72	11,79
		9	12,13	11,54	11,47	11,55	11,61	11,85	11,75	11,87	11,94
		10	12,31	11,62	11,54	11,62	11,69	12,01	11,91	12,03	12,11

## EK 7 Orijinal Veri Kümelerinde KNN ve TKNN için RMSE Değerleri

Veri Kümesi	Model	K	Kullanılan Ağırlıklı Tahmin Fonksiyonu								
			ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Diabetes</i>	KNN	1	0,81	0,81	0,81	0,81	1,22	0,81	0,81	0,81	0,81
		2	0,67	0,67	0,75	0,75	1,18	0,69	0,69	0,69	0,69
		3	0,62	0,64	0,71	0,71	1,18	0,64	0,64	0,64	0,64
		4	0,65	0,65	0,70	0,70	1,17	0,65	0,65	0,65	0,65
		5	0,63	0,63	0,69	0,69	1,16	0,62	0,62	0,62	0,62
		6	0,60	0,62	0,68	0,68	1,15	0,61	0,61	0,61	0,61
		7	0,59	0,60	0,66	0,66	1,14	0,60	0,60	0,60	0,60
		8	0,60	0,60	0,65	0,65	1,15	0,60	0,60	0,60	0,60
		9	0,61	0,61	0,66	0,66	1,14	0,61	0,61	0,61	0,61
		10	0,61	0,61	0,66	0,66	1,13	0,61	0,61	0,61	0,61
	TKNN	1	0,68	0,68	0,68	0,68	0,68	0,68	0,68	0,68	0,68
		2	0,63	0,63	0,65	0,64	0,63	0,65	0,65	0,64	0,64
		3	0,62	0,63	0,64	0,62	0,62	0,63	0,64	0,61	0,62
		4	0,61	0,62	0,63	0,61	0,61	0,62	0,63	0,61	0,62
		5	0,61	0,62	0,63	0,61	0,61	0,61	0,63	0,61	0,61
		6	0,62	0,62	0,62	0,60	0,60	0,61	0,63	0,61	0,61
		7	0,63	0,62	0,62	0,60	0,61	0,61	0,62	0,60	0,61
		8	0,64	0,62	0,62	0,60	0,61	0,61	0,62	0,61	0,62
		9	0,63	0,62	0,63	0,61	0,61	0,62	0,62	0,61	0,62
		10	0,64	0,62	0,62	0,61	0,61	0,62	0,63	0,61	0,62
<i>MachineCPU</i>	KNN	1	62,03	62,03	62,03	62,03	102,57	62,03	62,03	62,03	62,03
		2	63,25	61,70	60,12	60,12	103,61	62,35	62,35	62,35	62,35
		3	73,36	67,39	60,81	60,81	105,45	64,27	64,27	64,27	64,27
		4	80,61	71,96	61,99	61,99	106,79	66,23	66,23	66,23	66,23
		5	84,51	74,54	62,98	62,98	107,68	67,72	67,72	67,72	67,72
		6	87,55	76,54	63,69	63,69	108,52	69,53	69,53	69,53	69,53
		7	88,90	77,78	64,46	64,46	109,05	70,99	70,99	70,99	70,99
		8	88,70	78,24	64,77	64,77	109,51	71,74	71,74	71,74	71,74
		9	91,64	79,99	65,84	65,84	110,21	74,03	74,03	74,03	74,03
		10	90,23	79,56	65,64	65,64	110,42	72,72	72,72	72,72	72,72
	TKNN	1	62,16	62,16	62,16	62,16	62,16	62,16	62,16	62,16	62,16
		2	73,87	69,78	69,10	68,78	70,97	67,28	67,11	66,88	67,75
		3	82,53	75,32	73,59	73,17	77,83	70,65	69,68	69,36	72,17
		4	87,30	78,49	76,23	75,76	81,64	73,32	71,75	71,47	75,74
		5	89,34	80,07	77,46	77,03	83,42	74,91	72,69	72,39	78,05
		6	90,86	81,23	78,29	77,90	84,80	76,41	73,51	73,22	80,15
		7	91,99	82,09	78,78	78,41	85,84	77,72	74,14	73,87	82,01
		8	93,05	82,89	79,35	79,01	86,79	79,02	74,91	74,68	83,74
		9	93,93	83,57	79,85	79,60	87,58	80,30	75,76	75,63	85,35
		10	95,12	84,37	80,44	80,28	88,44	81,82	76,86	76,90	87,01

**EK 7 Orijinal Veri Kümelerinde KNN ve TKNN için RMSE Değerleri (Devamı)**

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Baseball</i>	KNN	1	14,26	14,26	14,26	14,26	17,89	14,26	14,26	14,26	14,26
		2	13,65	13,54	13,49	13,49	17,48	13,71	13,71	13,71	13,73
		3	12,04	12,25	12,75	12,75	17,09	12,24	12,24	12,24	12,24
		4	12,45	12,26	12,54	12,54	16,96	12,36	12,36	12,36	12,35
		5	12,89	12,37	12,44	12,44	16,94	12,56	12,56	12,56	12,56
		6	12,66	12,21	12,19	12,19	16,87	12,52	12,52	12,52	12,52
		7	12,93	12,29	12,13	12,13	16,82	12,62	12,62	12,62	12,62
		8	13,25	12,40	12,13	12,13	16,87	12,82	12,82	12,82	12,83
		9	13,41	12,44	12,09	12,09	16,86	12,84	12,84	12,84	12,84
		10	13,62	12,50	12,05	12,05	16,83	12,88	12,88	12,88	12,88
	TKNN	1	12,57	12,57	12,57	12,57	12,57	12,57	12,57	12,57	12,57
		2	12,26	12,24	12,34	12,17	12,27	12,31	12,38	12,24	12,32
		3	12,42	12,19	12,27	12,11	12,25	12,31	12,42	12,18	12,34
		4	12,85	12,34	12,41	12,25	12,40	12,53	12,62	12,39	12,50
		5	13,19	12,47	12,57	12,39	12,53	12,70	12,81	12,59	12,69
		6	13,53	12,63	12,71	12,56	12,68	12,86	12,92	12,75	12,85
		7	13,78	12,78	12,83	12,69	12,81	12,98	13,04	12,91	12,98
		8	13,99	12,84	12,90	12,78	12,91	13,06	13,11	12,98	13,06
		9	14,16	12,94	12,98	12,87	12,98	13,13	13,19	13,06	13,14
		10	14,38	13,01	13,04	12,94	13,07	13,22	13,29	13,16	13,23
<i>Dee</i>	KNN	1	0,56	0,56	0,56	0,56	0,63	0,56	0,56	0,56	0,56
		2	0,49	0,49	0,53	0,53	0,60	0,48	0,48	0,48	0,48
		3	0,44	0,45	0,50	0,50	0,58	0,44	0,44	0,44	0,44
		4	0,42	0,43	0,49	0,49	0,57	0,43	0,43	0,43	0,43
		5	0,42	0,42	0,47	0,47	0,56	0,42	0,42	0,42	0,42
		6	0,41	0,42	0,47	0,47	0,56	0,42	0,42	0,42	0,42
		7	0,42	0,42	0,46	0,46	0,56	0,42	0,42	0,42	0,42
		8	0,41	0,41	0,46	0,46	0,55	0,42	0,42	0,42	0,42
		9	0,41	0,41	0,45	0,45	0,55	0,42	0,42	0,42	0,42
		10	0,41	0,41	0,45	0,45	0,55	0,41	0,41	0,41	0,41
	TKNN	1	0,48	0,48	0,48	0,48	0,48	0,48	0,48	0,48	0,48
		2	0,43	0,44	0,45	0,44	0,44	0,43	0,44	0,43	0,43
		3	0,42	0,42	0,43	0,42	0,42	0,42	0,42	0,41	0,41
		4	0,41	0,42	0,42	0,42	0,41	0,41	0,41	0,41	0,41
		5	0,41	0,41	0,42	0,41	0,41	0,41	0,41	0,41	0,41
		6	0,41	0,41	0,42	0,41	0,41	0,41	0,41	0,41	0,41
		7	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41
		8	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41
		9	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41
		10	0,42	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41

**EK 7 Orijinal Veri Kümelerinde KNN ve TKNN için RMSE Değerleri (Devamı)**

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>AutoMpg</i>	<b>KNN</b>	<b>1</b>	843,77	843,77	843,77	843,77	840,16	843,77	843,77	843,77	843,77
		<b>2</b>	771,37	769,38	802,93	802,93	803,33	768,00	768,00	768,00	768,00
		<b>3</b>	734,25	734,33	777,48	777,48	783,49	728,90	728,90	728,90	728,90
		<b>4</b>	720,71	717,76	760,36	760,36	770,74	715,17	715,17	715,17	715,17
		<b>5</b>	709,52	706,78	748,16	748,16	760,40	705,88	705,88	705,88	705,88
		<b>6</b>	713,94	704,98	740,82	740,82	755,65	707,05	707,05	707,05	707,05
		<b>7</b>	705,75	698,32	733,02	733,02	749,57	699,08	699,08	699,08	699,08
		<b>8</b>	708,64	698,68	730,27	730,27	748,04	701,38	701,38	701,38	701,38
		<b>9</b>	711,44	698,25	726,86	726,86	745,56	703,01	703,01	703,01	703,01
		<b>10</b>	708,42	696,80	724,16	724,16	744,04	700,52	700,52	700,52	700,52
	<b>TKNN</b>	<b>1</b>	747,20	747,20	747,20	747,20	747,20	747,20	747,20	747,20	747,20
		<b>2</b>	710,61	714,48	715,65	713,46	713,55	709,26	710,58	708,02	708,28
		<b>3</b>	701,77	702,34	703,63	701,00	701,51	699,17	700,62	697,54	698,29
		<b>4</b>	702,28	698,29	699,70	696,93	697,36	697,99	699,63	696,41	696,96
		<b>5</b>	705,30	696,95	698,57	695,98	695,81	699,26	701,24	698,37	698,03
		<b>6</b>	708,97	696,74	698,56	696,17	695,44	701,38	703,50	701,08	699,92
		<b>7</b>	712,65	697,05	698,90	696,73	695,72	703,84	706,00	703,86	702,31
		<b>8</b>	717,42	698,18	700,07	697,98	696,84	707,36	709,53	707,60	705,83
		<b>9</b>	723,34	700,04	701,90	700,03	698,68	711,99	714,15	712,47	710,45
		<b>10</b>	728,73	701,70	703,66	701,86	700,35	716,30	718,38	716,78	714,69
<i>Laser</i>	<b>KNN</b>	<b>1</b>	3,25	3,25	3,25	3,25	4,46	3,25	3,25	3,25	3,25
		<b>2</b>	3,06	3,00	3,08	3,08	4,36	3,10	3,10	3,10	3,10
		<b>3</b>	2,90	2,88	2,99	2,99	4,32	3,03	3,03	3,03	3,03
		<b>4</b>	2,81	2,80	2,92	2,92	4,30	2,97	2,97	2,97	2,97
		<b>5</b>	2,78	2,77	2,87	2,87	4,28	2,95	2,95	2,95	2,95
		<b>6</b>	2,81	2,77	2,85	2,85	4,27	2,98	2,98	2,98	2,98
		<b>7</b>	2,89	2,79	2,84	2,84	4,27	3,02	3,02	3,02	3,02
		<b>8</b>	2,85	2,77	2,82	2,82	4,26	3,01	3,01	3,01	3,01
		<b>9</b>	2,86	2,77	2,80	2,80	4,26	3,01	3,01	3,01	3,01
		<b>10</b>	2,88	2,78	2,80	2,80	4,26	3,02	3,02	3,02	3,02
	<b>TKNN</b>	<b>1</b>	2,90	2,90	2,90	2,90	2,90	2,90	2,90	2,90	2,90
		<b>2</b>	2,81	2,80	2,81	2,80	2,80	2,82	2,83	2,82	2,82
		<b>3</b>	2,77	2,75	2,76	2,75	2,75	2,79	2,80	2,79	2,79
		<b>4</b>	2,80	2,75	2,76	2,74	2,75	2,81	2,82	2,80	2,80
		<b>5</b>	2,83	2,76	2,77	2,75	2,75	2,83	2,85	2,82	2,82
		<b>6</b>	2,84	2,75	2,77	2,74	2,75	2,83	2,85	2,82	2,83
		<b>7</b>	2,84	2,75	2,76	2,74	2,75	2,84	2,85	2,82	2,83
		<b>8</b>	2,84	2,75	2,76	2,74	2,75	2,83	2,84	2,82	2,84
		<b>9</b>	2,86	2,75	2,76	2,74	2,75	2,84	2,85	2,83	2,84
		<b>10</b>	2,87	2,75	2,76	2,74	2,75	2,85	2,85	2,84	2,85

**EK 7 Orijinal Veri Kümelerinde KNN ve TKNN için RMSE Değerleri (Devamı)**

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Concrete</i>	<b>KNN</b>	<b>1</b>	8,71	8,71	8,71	8,71	10,27	8,71	8,71	8,71	8,71
		<b>2</b>	8,29	7,92	8,42	8,42	9,82	7,93	7,93	7,93	7,93
		<b>3</b>	8,40	7,77	8,16	8,16	9,65	7,61	7,61	7,61	7,61
		<b>4</b>	8,50	7,74	8,05	8,05	9,59	7,58	7,58	7,58	7,58
		<b>5</b>	8,46	7,72	7,98	7,98	9,58	7,62	7,62	7,62	7,62
		<b>6</b>	8,50	7,71	7,91	7,91	9,54	7,63	7,63	7,63	7,63
		<b>7</b>	8,63	7,75	7,88	7,88	9,54	7,68	7,68	7,68	7,68
		<b>8</b>	8,70	7,78	7,87	7,87	9,56	7,70	7,70	7,70	7,70
		<b>9</b>	8,86	7,84	7,87	7,87	9,57	7,74	7,74	7,74	7,74
		<b>10</b>	8,92	7,87	7,87	7,87	9,58	7,80	7,80	7,80	7,80
	<b>TKNN</b>	<b>1</b>	8,68	8,68	8,68	8,68	8,68	8,68	8,68	8,68	8,68
		<b>2</b>	8,62	8,52	8,58	8,49	8,47	8,23	8,30	8,21	8,17
		<b>3</b>	8,66	8,45	8,53	8,42	8,40	8,11	8,20	8,09	8,04
		<b>4</b>	8,76	8,44	8,52	8,42	8,40	8,09	8,17	8,07	8,04
		<b>5</b>	8,89	8,48	8,55	8,46	8,44	8,12	8,20	8,10	8,08
		<b>6</b>	9,02	8,52	8,59	8,50	8,48	8,16	8,23	8,14	8,12
		<b>7</b>	9,12	8,55	8,62	8,53	8,52	8,20	8,27	8,17	8,16
		<b>8</b>	9,21	8,58	8,65	8,56	8,55	8,24	8,31	8,22	8,21
		<b>9</b>	9,30	8,61	8,68	8,59	8,59	8,28	8,35	8,26	8,25
		<b>10</b>	9,41	8,65	8,71	8,63	8,63	8,33	8,39	8,30	8,30
<i>Ele2</i>	<b>KNN</b>	<b>1</b>	180,60	180,60	180,60	180,60	213,28	180,60	180,60	180,60	180,60
		<b>2</b>	165,54	166,38	167,36	167,36	201,49	165,38	165,38	165,38	165,38
		<b>3</b>	215,72	178,84	202,69	202,69	235,82	164,22	164,22	164,22	164,22
		<b>4</b>	255,23	195,45	209,45	209,45	243,12	158,80	158,80	158,80	158,80
		<b>5</b>	287,64	210,94	215,96	215,96	249,20	158,90	158,90	158,90	158,90
		<b>6</b>	283,26	212,58	216,32	216,32	251,21	158,40	158,40	158,40	158,40
		<b>7</b>	284,41	214,47	216,24	216,24	252,77	158,92	158,92	158,92	158,92
		<b>8</b>	298,67	219,80	218,63	218,63	255,80	162,90	162,90	162,90	162,90
		<b>9</b>	298,62	221,39	219,27	219,27	256,35	162,73	162,73	162,73	162,73
		<b>10</b>	303,23	224,17	220,60	220,60	258,01	161,23	161,23	161,23	161,23
	<b>TKNN</b>	<b>1</b>	209,47	209,47	209,47	209,47	209,47	209,47	209,47	209,47	209,47
		<b>2</b>	275,14	252,54	252,24	251,48	252,91	196,85	196,93	196,71	196,79
		<b>3</b>	297,68	268,21	267,91	266,94	268,57	192,59	192,73	192,42	192,46
		<b>4</b>	304,82	272,98	272,79	271,72	273,24	191,78	191,87	191,63	191,71
		<b>5</b>	311,13	275,98	275,87	274,79	276,16	191,30	191,26	191,16	191,35
		<b>6</b>	316,40	278,52	278,46	277,38	278,65	190,59	190,48	190,44	190,71
		<b>7</b>	319,16	279,95	279,92	278,86	280,06	189,01	188,88	188,85	189,17
		<b>8</b>	321,67	281,31	281,30	280,27	281,41	187,94	187,84	187,74	188,08
		<b>9</b>	324,08	282,66	282,65	281,65	282,76	187,15	187,08	186,91	187,26
		<b>10</b>	327,10	284,30	284,28	283,32	284,41	186,79	186,74	186,53	186,88

**EK 7 Orijinal Veri Kümelerinde KNN ve TKNN için RMSE Değerleri (Devamı)**

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Friedman</i>	KNN	1	2,48	2,48	2,48	2,48	2,73	2,48	2,48	2,48	2,48
		2	2,13	2,12	2,28	2,28	2,57	2,11	2,11	2,11	2,11
		3	2,00	1,98	2,17	2,17	2,48	1,98	1,98	1,98	1,98
		4	1,94	1,91	2,10	2,10	2,43	1,91	1,91	1,91	1,91
		5	1,87	1,85	2,04	2,04	2,38	1,84	1,84	1,84	1,84
		6	1,86	1,83	2,01	2,01	2,35	1,82	1,82	1,82	1,82
		7	1,84	1,81	1,98	1,98	2,33	1,81	1,81	1,81	1,81
		8	1,82	1,79	1,95	1,95	2,31	1,79	1,79	1,79	1,79
		9	1,82	1,78	1,94	1,94	2,30	1,78	1,78	1,78	1,78
		10	1,83	1,78	1,93	1,93	2,29	1,79	1,79	1,79	1,79
	TKNN	1	2,04	2,04	2,04	2,04	2,04	2,04	2,04	2,04	2,04
		2	1,89	1,90	1,91	1,90	1,90	1,88	1,89	1,88	1,88
		3	1,84	1,84	1,84	1,84	1,84	1,82	1,82	1,82	1,82
		4	1,81	1,81	1,81	1,81	1,81	1,79	1,79	1,79	1,80
		5	1,81	1,79	1,79	1,79	1,80	1,78	1,78	1,78	1,79
		6	1,81	1,78	1,78	1,78	1,79	1,78	1,78	1,78	1,79
		7	1,83	1,78	1,78	1,78	1,79	1,80	1,79	1,79	1,81
		8	1,85	1,79	1,78	1,79	1,80	1,81	1,80	1,81	1,82
		9	1,87	1,79	1,78	1,79	1,80	1,83	1,81	1,83	1,84
		10	1,89	1,80	1,79	1,80	1,81	1,84	1,83	1,84	1,86

**EK 8 Aykırı Değer Analizi Yapılan Veri Kümelerinde KNN ve TKNN için MAPE  
(%) Değerleri**

Veri Kümesi	Model	K	Kullanılan Ağırlıklı Tahmin Fonksiyonu								
			ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Diabetes</i>	KNN	1	12,26	12,26	12,26	12,26	13,74	12,26	12,26	12,26	12,26
		2	9,97	10,16	11,32	11,32	12,94	10,30	10,30	10,30	10,30
		3	9,60	9,77	10,86	10,86	12,78	9,76	9,76	9,76	9,76
		4	9,73	9,74	10,79	10,79	12,77	9,85	9,85	9,85	9,85
		5	9,26	9,33	10,43	10,43	12,48	9,39	9,39	9,39	9,39
		6	9,00	9,38	10,39	10,39	12,27	9,29	9,29	9,29	9,29
		7	8,55	8,90	10,08	10,08	12,08	9,11	9,11	9,11	9,11
		8	8,74	8,91	9,97	9,97	12,03	8,85	8,85	8,85	8,85
		9	8,57	8,89	10,01	10,01	11,92	8,99	8,99	8,99	8,99
		10	8,89	8,78	9,91	9,91	11,94	9,13	9,13	9,13	9,13
	TKNN	1	10,16	10,16	10,16	10,16	10,16	10,16	10,16	10,16	10,16
		2	9,41	9,68	9,77	9,50	9,37	9,65	9,71	9,67	9,43
		3	9,09	9,23	9,55	8,95	8,90	9,32	9,37	9,09	9,11
		4	8,97	9,02	9,28	8,81	8,90	8,95	9,17	8,99	9,04
		5	8,79	8,90	9,10	8,63	8,69	8,91	9,00	8,81	8,91
		6	8,80	8,80	9,09	8,76	8,84	8,96	9,16	8,91	8,94
		7	8,99	8,97	9,11	8,90	8,86	8,98	9,19	9,03	9,10
		8	9,06	9,13	9,34	8,91	9,01	9,06	9,27	9,21	9,27
		9	9,17	9,18	9,44	8,96	9,02	9,01	9,23	9,16	9,29
		10	9,06	9,14	9,28	9,01	9,13	9,19	9,14	9,05	9,36
<i>MachineCPU</i>	KNN	1	37,27	37,27	37,27	37,27	36,01	37,27	37,27	37,27	37,27
		2	35,33	34,44	35,49	35,49	34,40	34,71	34,71	34,71	34,72
		3	38,71	36,63	35,59	35,59	34,51	37,91	37,91	37,91	37,91
		4	37,46	35,92	35,01	35,01	34,06	36,92	36,92	36,92	36,92
		5	37,12	35,54	34,71	34,71	33,81	36,11	36,11	36,11	36,11
		6	37,33	35,52	34,65	34,65	33,79	35,56	35,56	35,56	35,56
		7	37,55	35,45	34,61	34,61	33,61	35,79	35,79	35,79	35,79
		8	37,06	35,34	34,28	34,28	33,58	35,59	35,59	35,59	35,59
		9	35,53	34,70	34,02	34,02	33,37	34,87	34,87	34,87	34,87
		10	35,63	34,56	33,88	33,88	33,44	34,63	34,63	34,63	34,63
	TKNN	1	34,44	34,44	34,44	34,44	34,44	34,44	34,44	34,44	34,44
		2	36,24	35,25	35,42	35,13	35,20	35,38	35,31	35,14	35,29
		3	37,10	35,81	35,76	35,56	35,99	35,52	35,48	35,32	35,47
		4	36,59	35,65	35,40	35,31	35,61	34,73	34,61	34,54	34,84
		5	36,25	35,31	35,21	34,99	35,49	34,43	34,33	34,42	34,86
		6	35,67	34,85	34,71	34,72	35,16	34,18	33,90	33,87	34,22
		7	35,28	34,72	34,71	34,57	34,85	33,85	33,66	33,76	34,02
		8	35,02	34,48	34,36	34,21	34,61	33,64	33,45	33,65	34,01
		9	34,94	34,30	34,22	34,11	34,48	33,44	33,27	33,43	33,86
		10	34,93	34,28	34,24	34,09	34,31	33,57	33,30	33,43	33,87

**EK 8 Aykırı Değer Analizi Yapılan Veri Kümelerinde KNN ve TKNN için MAPE Değerleri (%) (Devamı)**

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Baseball</i>	KNN	1	23,85	23,85	23,85	23,85	21,90	23,85	23,85	23,85	23,85
		2	25,53	25,12	23,55	23,55	21,82	24,78	24,78	24,78	24,84
		3	27,23	25,65	23,64	23,64	21,86	24,12	24,12	24,12	24,12
		4	28,40	26,37	23,98	23,98	21,94	23,82	23,82	23,82	23,83
		5	29,10	26,67	24,04	24,04	22,04	23,97	23,97	23,97	23,97
		6	28,59	26,46	23,39	23,39	21,53	23,58	23,58	23,58	23,60
		7	29,07	26,95	23,69	23,69	21,77	24,15	24,15	24,15	24,15
		8	29,70	27,31	23,89	23,89	22,02	24,78	24,78	24,78	24,79
		9	30,95	27,78	24,28	24,28	22,18	25,14	25,14	25,14	25,14
		10	31,80	28,29	24,55	24,55	22,40	25,57	25,57	25,57	25,57
	TKNN	1	25,07	25,07	25,07	25,07	25,07	25,07	25,07	25,07	25,07
		2	28,41	27,56	27,63	27,27	27,49	27,49	27,52	27,18	27,38
		3	28,98	27,88	27,94	27,64	27,93	27,48	27,50	27,21	27,54
		4	29,67	28,29	28,20	27,98	28,42	27,73	27,63	27,34	27,77
		5	30,67	28,73	28,63	28,41	28,95	28,26	28,25	27,96	28,40
		6	31,88	29,37	29,26	29,12	29,56	28,94	28,87	28,65	29,15
		7	32,84	30,00	29,83	29,67	30,16	29,58	29,38	29,31	29,78
		8	33,58	30,36	30,19	30,01	30,51	29,98	29,83	29,68	30,18
		9	33,91	30,71	30,54	30,37	30,91	30,18	30,09	29,90	30,41
		10	34,55	30,96	30,83	30,68	31,14	30,45	30,29	30,11	30,62
<i>Dee</i>	KNN	1	10,48	10,48	10,48	10,48	11,08	10,48	10,48	10,48	10,48
		2	9,49	9,47	9,92	9,92	10,57	9,44	9,44	9,44	9,44
		3	8,84	8,90	9,52	9,52	10,25	8,80	8,80	8,80	8,80
		4	8,44	8,48	9,22	9,22	10,00	8,37	8,37	8,37	8,37
		5	8,31	8,30	9,02	9,02	9,85	8,23	8,23	8,23	8,23
		6	8,22	8,23	8,90	8,90	9,75	8,18	8,18	8,18	8,18
		7	8,21	8,18	8,81	8,81	9,70	8,20	8,20	8,20	8,20
		8	8,11	8,12	8,71	8,71	9,59	8,13	8,13	8,13	8,13
		9	8,11	8,09	8,68	8,68	9,55	8,13	8,13	8,13	8,13
		10	8,18	8,09	8,64	8,64	9,50	8,16	8,16	8,16	8,16
	TKNN	1	9,21	9,21	9,21	9,21	9,21	9,21	9,21	9,21	9,21
		2	8,58	8,70	8,78	8,62	8,59	8,57	8,65	8,50	8,48
		3	8,29	8,39	8,50	8,33	8,32	8,25	8,33	8,20	8,19
		4	8,19	8,27	8,38	8,21	8,19	8,15	8,23	8,10	8,10
		5	8,14	8,22	8,30	8,15	8,15	8,12	8,18	8,06	8,06
		6	8,15	8,18	8,27	8,12	8,12	8,10	8,17	8,06	8,05
		7	8,19	8,17	8,26	8,12	8,11	8,11	8,19	8,07	8,07
		8	8,25	8,18	8,25	8,12	8,12	8,15	8,23	8,10	8,09
		9	8,29	8,18	8,27	8,13	8,12	8,18	8,25	8,14	8,12
		10	8,33	8,19	8,27	8,13	8,13	8,20	8,28	8,16	8,15



**EK 8 Aykırı Değer Analizi Yapılan Veri Kümelerinde KNN ve TKNN için MAPE Değerleri (%) (Devamı)**

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9	
<i>AutoMpg</i>	KNN	1	69,83	69,83	69,83	69,83	63,21	69,83	69,83	69,83	69,83	
		2	67,83	67,05	67,27	67,27	60,90	66,96	66,96	66,96	66,96	
		3	66,87	65,97	66,09	66,09	59,82	66,42	66,42	66,42	66,42	
		4	64,63	64,07	64,85	64,85	58,84	64,31	64,31	64,31	64,31	
		5	64,27	63,68	64,18	64,18	58,00	64,10	64,10	64,10	64,10	
		6	65,86	64,32	64,27	64,27	57,96	65,45	65,45	65,45	65,45	
		7	65,07	63,86	63,93	63,93	57,59	64,68	64,68	64,68	64,68	
		8	64,91	63,77	63,74	63,74	57,35	64,53	64,53	64,53	64,53	
		9	63,56	63,12	63,25	63,25	56,94	63,31	63,31	63,31	63,31	
		10	63,75	63,16	63,14	63,14	56,80	63,47	63,47	63,47	63,47	
	TKNN	1	65,19	65,19	65,19	65,19	65,19	65,19	65,19	65,19	65,19	65,19
		2	64,31	64,27	64,58	64,12	63,96	64,11	64,41	63,95	63,81	
		3	63,88	63,58	64,12	63,41	63,09	63,73	64,32	63,52	63,18	
		4	64,08	63,52	64,18	63,33	62,92	63,92	64,64	63,70	63,27	
		5	64,02	63,42	64,07	63,23	62,81	63,80	64,52	63,62	63,18	
		6	63,75	63,22	63,86	63,08	62,65	63,51	64,16	63,37	62,91	
		7	63,45	63,04	63,64	62,92	62,47	63,17	63,76	63,05	62,64	
		8	63,45	62,97	63,56	62,87	62,45	63,12	63,65	63,03	62,63	
		9	63,57	62,99	63,55	62,89	62,46	63,13	63,66	63,07	62,66	
		10	63,78	63,01	63,58	62,93	62,51	63,19	63,70	63,11	62,75	
<i>Laser</i>	KNN	1	8,63	8,63	8,63	8,63	11,70	8,63	8,63	8,63	8,63	
		2	7,63	7,59	8,10	8,10	11,05	7,62	7,62	7,62	7,62	
		3	7,31	7,25	7,81	7,81	10,78	7,37	7,37	7,37	7,37	
		4	7,62	7,32	7,71	7,71	10,62	7,54	7,54	7,54	7,54	
		5	7,76	7,36	7,62	7,62	10,51	7,61	7,61	7,61	7,61	
		6	7,82	7,40	7,58	7,58	10,41	7,61	7,61	7,61	7,61	
		7	7,90	7,42	7,52	7,52	10,34	7,68	7,68	7,68	7,68	
		8	7,93	7,44	7,50	7,50	10,26	7,70	7,70	7,70	7,70	
		9	7,99	7,48	7,46	7,46	10,25	7,73	7,73	7,73	7,73	
		10	8,14	7,52	7,48	7,48	10,22	7,81	7,81	7,81	7,81	
	TKNN	1	7,52	7,52	7,52	7,52	7,52	7,52	7,52	7,52	7,52	
		2	7,29	7,24	7,24	7,20	7,23	7,24	7,25	7,22	7,25	
		3	7,58	7,33	7,31	7,29	7,35	7,45	7,44	7,41	7,47	
		4	7,77	7,44	7,41	7,39	7,45	7,58	7,58	7,55	7,59	
		5	7,89	7,49	7,46	7,44	7,52	7,68	7,66	7,63	7,69	
		6	8,02	7,53	7,51	7,50	7,56	7,76	7,74	7,74	7,78	
		7	8,06	7,58	7,54	7,54	7,61	7,82	7,80	7,78	7,85	
		8	8,14	7,61	7,58	7,57	7,63	7,87	7,82	7,83	7,90	
		9	8,21	7,63	7,62	7,60	7,66	7,91	7,90	7,88	7,95	
		10	8,28	7,68	7,65	7,64	7,71	7,95	7,95	7,93	7,99	

**EK 8 Aykırı Değer Analizi Yapılan Veri Kümelerinde KNN ve TKNN için MAPE Değerleri (%) (Devamı)**

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Concrete</i>	KNN	1	21,13	21,13	21,13	21,13	22,47	21,13	21,13	21,13	21,13
		2	22,27	20,76	21,91	21,91	22,15	19,85	19,85	19,85	19,85
		3	21,47	20,22	21,20	21,20	21,45	19,30	19,30	19,30	19,30
		4	21,99	20,43	21,24	21,24	21,35	19,48	19,48	19,48	19,48
		5	22,91	20,75	21,32	21,32	21,35	19,59	19,59	19,59	19,59
		6	23,51	21,06	21,30	21,30	21,30	19,70	19,70	19,70	19,70
		7	24,11	21,31	21,40	21,40	21,37	19,77	19,77	19,77	19,77
		8	24,25	21,42	21,41	21,41	21,39	19,90	19,90	19,90	19,90
		9	25,08	21,80	21,56	21,56	21,50	20,04	20,04	20,04	20,04
		10	25,41	21,95	21,64	21,64	21,57	20,13	20,13	20,13	20,13
	TKNN	1	22,82	22,82	22,82	22,82	22,82	22,82	22,82	22,82	22,82
		2	22,80	22,54	22,53	22,44	22,57	21,71	21,71	21,64	21,71
		3	23,54	22,84	22,82	22,69	22,89	21,88	21,88	21,78	21,89
		4	24,40	23,30	23,28	23,14	23,36	22,16	22,17	22,07	22,20
		5	25,07	23,65	23,64	23,49	23,70	22,40	22,39	22,31	22,44
		6	25,55	23,90	23,90	23,75	23,96	22,57	22,54	22,47	22,62
		7	26,02	24,14	24,13	24,00	24,20	22,75	22,72	22,66	22,81
		8	26,48	24,38	24,35	24,24	24,43	22,96	22,92	22,87	23,02
		9	26,93	24,60	24,57	24,48	24,66	23,17	23,13	23,09	23,24
		10	27,42	24,84	24,81	24,72	24,90	23,38	23,33	23,30	23,46
<i>Ele2</i>	KNN	1	19,40	19,40	19,40	19,40	20,30	19,40	19,40	19,40	19,40
		2	19,64	19,34	19,38	19,38	20,18	19,60	19,60	19,60	19,60
		3	20,94	19,53	19,93	19,93	20,62	19,32	19,32	19,32	19,32
		4	22,59	20,49	20,63	20,63	21,15	20,08	20,08	20,08	20,08
		5	23,59	21,01	20,95	20,95	21,43	20,53	20,53	20,53	20,53
		6	25,10	21,57	21,49	21,49	21,92	22,14	22,14	22,14	22,14
		7	25,86	21,93	21,80	21,80	22,17	22,87	22,87	22,87	22,87
		8	28,29	22,59	22,45	22,45	22,79	25,08	25,08	25,08	25,08
		9	29,02	22,93	22,81	22,81	23,12	25,85	25,85	25,85	25,85
		10	28,04	22,86	22,70	22,70	23,02	24,75	24,75	24,75	24,75
	TKNN	1	21,74	21,74	21,74	21,74	21,74	21,74	21,74	21,74	21,74
		2	23,38	22,75	22,77	22,71	22,74	21,87	21,91	21,84	21,83
		3	26,05	24,14	24,19	24,08	24,10	23,95	24,01	23,91	23,89
		4	27,59	25,04	25,07	24,98	25,01	25,33	25,33	25,32	25,33
		5	28,39	25,56	25,55	25,52	25,57	25,99	25,88	26,01	26,08
		6	28,18	25,58	25,54	25,56	25,62	25,63	25,47	25,68	25,78
		7	27,50	25,38	25,31	25,38	25,45	24,85	24,64	24,93	25,04
		8	26,91	25,20	25,11	25,20	25,28	24,21	23,99	24,28	24,41
		9	26,47	25,06	24,96	25,06	25,14	23,71	23,51	23,78	23,90
		10	26,17	24,97	24,88	24,97	25,06	23,37	23,18	23,43	23,54

**EK 8 Aykırı Değer Analizi Yapılan Veri Kümelerinde KNN ve TKNN için MAPE Değerleri (%) (Devamı)**

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Friedman</i>	KNN	1	15,24	15,24	15,24	15,24	15,44	15,24	15,24	15,24	15,24
		2	13,16	13,06	14,04	14,04	14,38	13,05	13,05	13,05	13,05
		3	12,41	12,23	13,39	13,39	13,76	12,26	12,26	12,26	12,26
		4	12,18	11,93	13,05	13,05	13,44	11,98	11,98	11,98	11,98
		5	11,90	11,67	12,78	12,78	13,19	11,69	11,69	11,69	11,69
		6	11,87	11,60	12,62	12,62	13,01	11,65	11,65	11,65	11,65
		7	11,71	11,48	12,47	12,47	12,86	11,50	11,50	11,50	11,50
		8	11,58	11,37	12,33	12,33	12,73	11,39	11,39	11,39	11,39
		9	11,62	11,35	12,23	12,23	12,66	11,40	11,40	11,40	11,40
		10	11,69	11,36	12,17	12,17	12,59	11,46	11,46	11,46	11,46
	TKNN	1	12,67	12,67	12,67	12,67	12,67	12,67	12,67	12,67	12,67
		2	11,83	11,87	11,86	11,86	11,89	11,77	11,76	11,77	11,80
		3	11,67	11,62	11,59	11,61	11,65	11,55	11,53	11,55	11,59
		4	11,56	11,49	11,47	11,49	11,53	11,43	11,40	11,43	11,47
		5	11,56	11,42	11,39	11,42	11,46	11,41	11,36	11,41	11,46
		6	11,64	11,40	11,35	11,40	11,45	11,46	11,39	11,46	11,52
		7	11,78	11,43	11,37	11,43	11,48	11,56	11,49	11,57	11,64
		8	11,95	11,48	11,42	11,49	11,55	11,70	11,62	11,72	11,79
		9	12,13	11,54	11,47	11,55	11,61	11,85	11,75	11,87	11,94
		10	12,31	11,62	11,54	11,62	11,69	12,01	11,91	12,03	12,11

**EK 9 Aykırı Değer Analizi Yapılan Veri Kümelerinde KNN ve TKNN için RMSE Değerleri**

Veri Kümesi	Model	K	Kullanılan Ağırlıklı Tahmin Fonksiyonu								
			ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Diabetes</i>	KNN	1	0,82	0,82	0,82	0,82	1,10	0,82	0,82	0,82	0,82
		2	0,65	0,67	0,75	0,75	1,07	0,68	0,68	0,68	0,68
		3	0,63	0,65	0,72	0,72	1,07	0,64	0,64	0,64	0,64
		4	0,66	0,65	0,71	0,71	1,06	0,66	0,66	0,66	0,66
		5	0,63	0,62	0,69	0,69	1,06	0,63	0,63	0,63	0,63
		6	0,61	0,63	0,69	0,69	1,06	0,61	0,61	0,61	0,61
		7	0,60	0,60	0,67	0,67	1,04	0,61	0,61	0,61	0,61
		8	0,60	0,61	0,66	0,66	1,04	0,59	0,59	0,59	0,59
		9	0,60	0,60	0,66	0,66	1,04	0,60	0,60	0,60	0,60
		10	0,61	0,59	0,65	0,65	1,04	0,61	0,61	0,61	0,61
	TKNN	1	0,68	0,68	0,68	0,68	0,68	0,68	0,68	0,68	0,68
		2	0,63	0,65	0,64	0,63	0,63	0,64	0,66	0,65	0,63
		3	0,62	0,62	0,64	0,61	0,61	0,63	0,63	0,61	0,62
		4	0,61	0,61	0,63	0,60	0,60	0,61	0,62	0,60	0,61
		5	0,60	0,61	0,61	0,59	0,59	0,60	0,61	0,59	0,60
		6	0,61	0,60	0,62	0,59	0,60	0,60	0,62	0,60	0,60
		7	0,61	0,61	0,62	0,60	0,60	0,60	0,62	0,60	0,61
		8	0,62	0,61	0,63	0,60	0,61	0,61	0,62	0,61	0,62
		9	0,63	0,61	0,63	0,61	0,61	0,61	0,62	0,61	0,62
		10	0,62	0,61	0,62	0,61	0,61	0,62	0,61	0,61	0,62
<i>MachineCPU</i>	KNN	1	39,90	39,90	39,90	39,90	42,44	39,90	39,90	39,90	39,90
		2	34,37	34,58	36,26	36,26	39,96	35,92	35,92	35,92	35,92
		3	35,36	34,71	35,85	35,85	39,90	37,21	37,21	37,21	37,21
		4	35,04	34,18	35,22	35,22	39,54	36,62	36,62	36,62	36,62
		5	34,48	33,43	34,65	34,65	39,52	35,30	35,30	35,30	35,30
		6	35,99	33,83	34,65	34,65	39,83	35,63	35,63	35,63	35,63
		7	36,65	33,89	34,50	34,50	40,05	35,51	35,51	35,51	35,51
		8	37,11	34,00	34,38	34,38	40,23	35,62	35,62	35,62	35,62
		9	37,36	33,93	34,12	34,12	40,29	35,46	35,46	35,46	35,46
		10	38,72	34,32	34,08	34,08	40,57	35,94	35,94	35,94	35,94
	TKNN	1	34,39	34,39	34,39	34,39	34,39	34,39	34,39	34,39	34,39
		2	34,50	33,97	34,09	34,06	33,89	34,18	34,26	34,28	34,15
		3	35,65	34,47	34,45	34,47	34,55	34,50	34,43	34,51	34,58
		4	36,64	34,85	34,66	34,76	35,01	34,65	34,45	34,61	34,94
		5	37,87	35,26	35,01	35,11	35,58	35,31	34,99	35,16	35,71
		6	39,31	35,71	35,35	35,53	36,19	36,22	35,80	35,99	36,70
		7	40,83	36,32	35,88	36,05	36,79	37,20	36,70	36,94	37,71
		8	42,41	36,93	36,46	36,64	37,45	38,15	37,65	37,91	38,74
		9	43,75	37,51	37,02	37,21	38,07	39,00	38,45	38,70	39,55
		10	44,81	37,98	37,51	37,67	38,55	39,72	39,21	39,44	40,25

**EK 9 Aykırı Değer Analizi Yapılan Veri Kümelerinde KNN ve TKNN için RMSE Değerleri (Devamı)**

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Baseball</i>	<b>KNN</b>	<b>1</b>	13,23	13,23	13,23	13,23	14,76	13,23	13,23	13,23	13,23
		<b>2</b>	12,54	12,47	12,47	12,47	14,25	12,63	12,63	12,63	12,65
		<b>3</b>	11,52	11,53	11,83	11,83	13,88	11,46	11,46	11,46	11,46
		<b>4</b>	11,85	11,51	11,59	11,59	13,72	11,48	11,48	11,48	11,47
		<b>5</b>	12,03	11,47	11,43	11,43	13,68	11,51	11,51	11,51	11,51
		<b>6</b>	11,74	11,26	11,15	11,15	13,55	11,40	11,40	11,40	11,40
		<b>7</b>	11,93	11,32	11,06	11,06	13,52	11,52	11,52	11,52	11,52
		<b>8</b>	12,44	11,49	11,08	11,08	13,57	11,77	11,77	11,77	11,77
		<b>9</b>	12,74	11,58	11,07	11,07	13,55	11,81	11,81	11,81	11,81
		<b>10</b>	12,89	11,66	11,04	11,04	13,51	11,89	11,89	11,89	11,89
	<b>TKNN</b>	<b>1</b>	11,68	11,68	11,68	11,68	11,68	11,68	11,68	11,68	11,68
		<b>2</b>	11,68	11,56	11,62	11,51	11,55	11,66	11,70	11,59	11,62
		<b>3</b>	11,74	11,46	11,49	11,38	11,43	11,58	11,65	11,48	11,54
		<b>4</b>	12,00	11,52	11,57	11,43	11,48	11,63	11,72	11,51	11,58
		<b>5</b>	12,40	11,64	11,69	11,56	11,63	11,80	11,90	11,68	11,75
		<b>6</b>	12,81	11,83	11,88	11,75	11,82	12,02	12,08	11,91	11,98
		<b>7</b>	13,11	11,98	12,02	11,88	11,96	12,19	12,24	12,09	12,17
		<b>8</b>	13,30	12,07	12,10	11,97	12,05	12,30	12,35	12,19	12,30
		<b>9</b>	13,48	12,16	12,20	12,04	12,15	12,38	12,44	12,28	12,38
		<b>10</b>	13,67	12,24	12,28	12,15	12,24	12,48	12,52	12,37	12,46
<i>Dee</i>	<b>KNN</b>	<b>1</b>	0,56	0,56	0,56	0,56	0,63	0,56	0,56	0,56	0,56
		<b>2</b>	0,49	0,49	0,53	0,53	0,60	0,48	0,48	0,48	0,48
		<b>3</b>	0,44	0,45	0,50	0,50	0,58	0,44	0,44	0,44	0,44
		<b>4</b>	0,42	0,43	0,49	0,49	0,57	0,43	0,43	0,43	0,43
		<b>5</b>	0,42	0,42	0,47	0,47	0,56	0,42	0,42	0,42	0,42
		<b>6</b>	0,41	0,42	0,47	0,47	0,56	0,42	0,42	0,42	0,42
		<b>7</b>	0,42	0,42	0,46	0,46	0,56	0,42	0,42	0,42	0,42
		<b>8</b>	0,41	0,41	0,46	0,46	0,55	0,42	0,42	0,42	0,42
		<b>9</b>	0,41	0,41	0,45	0,45	0,55	0,42	0,42	0,42	0,42
		<b>10</b>	0,41	0,41	0,45	0,45	0,55	0,41	0,41	0,41	0,41
	<b>TKNN</b>	<b>1</b>	0,48	0,48	0,48	0,48	0,48	0,48	0,48	0,48	0,48
		<b>2</b>	0,43	0,44	0,45	0,44	0,44	0,43	0,44	0,43	0,43
		<b>3</b>	0,42	0,42	0,43	0,42	0,42	0,42	0,42	0,41	0,41
		<b>4</b>	0,41	0,42	0,42	0,42	0,41	0,41	0,41	0,41	0,41
		<b>5</b>	0,41	0,41	0,42	0,41	0,41	0,41	0,41	0,41	0,41
		<b>6</b>	0,41	0,41	0,42	0,41	0,41	0,41	0,41	0,41	0,41
		<b>7</b>	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41
		<b>8</b>	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41
		<b>9</b>	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41
		<b>10</b>	0,42	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41

**EK 9 Aykırı Değer Analizi Yapılan Veri Kümelerinde KNN ve TKNN için RMSE Değerleri (Devamı)**

Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>AutoMpg</i>	KNN	1	836,36	836,36	836,36	836,36	827,61	836,36	836,36	836,36	836,36
		2	773,37	767,80	796,04	796,04	791,84	767,38	767,38	767,38	767,38
		3	729,78	729,30	769,18	769,18	769,41	724,12	724,12	724,12	724,12
		4	718,50	715,37	753,85	753,85	759,65	711,10	711,10	711,10	711,10
		5	701,46	701,25	740,91	740,91	748,73	695,86	695,86	695,86	695,86
		6	706,25	698,90	733,36	733,36	743,16	698,39	698,39	698,39	698,39
		7	710,74	699,52	729,32	729,32	740,10	702,50	702,50	702,50	702,50
		8	710,52	697,52	724,60	724,60	737,40	701,56	701,56	701,56	701,56
		9	711,63	696,61	721,11	721,11	735,11	701,55	701,55	701,55	701,55
		10	713,23	696,73	719,10	719,10	733,89	703,05	703,05	703,05	703,05
	TKNN	1	743,62	743,62	743,62	743,62	743,62	743,62	743,62	743,62	743,62
		2	713,78	716,54	717,77	714,78	715,65	711,59	712,83	709,56	710,71
		3	698,12	700,66	702,39	698,38	699,53	695,22	697,10	692,44	694,06
		4	698,62	696,19	697,93	693,93	695,20	694,20	696,03	691,54	693,19
		5	702,94	695,40	697,10	693,43	694,35	696,68	698,54	694,67	695,56
		6	707,77	695,64	697,34	694,23	694,64	699,86	701,70	698,65	698,78
		7	711,70	696,08	697,61	694,93	695,09	702,56	704,15	701,82	701,59
		8	715,89	696,86	698,35	696,00	695,97	705,66	707,14	705,28	704,72
		9	721,39	698,41	699,86	697,67	697,48	709,82	711,28	709,55	708,79
		10	726,19	699,81	701,23	699,18	698,90	713,55	715,06	713,44	712,54
<i>Laser</i>	KNN	1	3,15	3,15	3,15	3,15	4,26	3,15	3,15	3,15	3,15
		2	2,77	2,74	2,94	2,94	4,15	2,78	2,78	2,78	2,78
		3	2,57	2,57	2,81	2,81	4,09	2,60	2,60	2,60	2,60
		4	2,62	2,57	2,77	2,77	4,07	2,64	2,64	2,64	2,64
		5	2,63	2,56	2,73	2,73	4,05	2,63	2,63	2,63	2,63
		6	2,63	2,55	2,70	2,70	4,03	2,63	2,63	2,63	2,63
		7	2,66	2,55	2,67	2,67	4,03	2,63	2,63	2,63	2,63
		8	2,66	2,55	2,65	2,65	4,02	2,64	2,64	2,64	2,64
		9	2,67	2,54	2,63	2,63	4,01	2,64	2,64	2,64	2,64
		10	2,72	2,57	2,63	2,63	4,01	2,67	2,67	2,67	2,67
	TKNN	1	2,71	2,71	2,71	2,71	2,71	2,71	2,71	2,71	2,71
		2	2,56	2,57	2,57	2,56	2,56	2,58	2,58	2,57	2,57
		3	2,58	2,55	2,55	2,54	2,55	2,58	2,58	2,57	2,58
		4	2,61	2,55	2,56	2,55	2,55	2,59	2,60	2,59	2,59
		5	2,64	2,56	2,56	2,55	2,56	2,61	2,61	2,60	2,61
		6	2,67	2,57	2,57	2,55	2,56	2,63	2,63	2,62	2,63
		7	2,69	2,57	2,57	2,56	2,57	2,64	2,64	2,63	2,65
		8	2,70	2,57	2,57	2,56	2,58	2,65	2,64	2,64	2,65
		9	2,72	2,58	2,58	2,57	2,58	2,65	2,65	2,65	2,66
		10	2,73	2,58	2,58	2,58	2,59	2,66	2,66	2,66	2,67

**EK 9 Aykırı Veri Analizi Yapılan Veri Kümelerinde KNN ve TKNN için RMSE Değerleri (Devamı)**

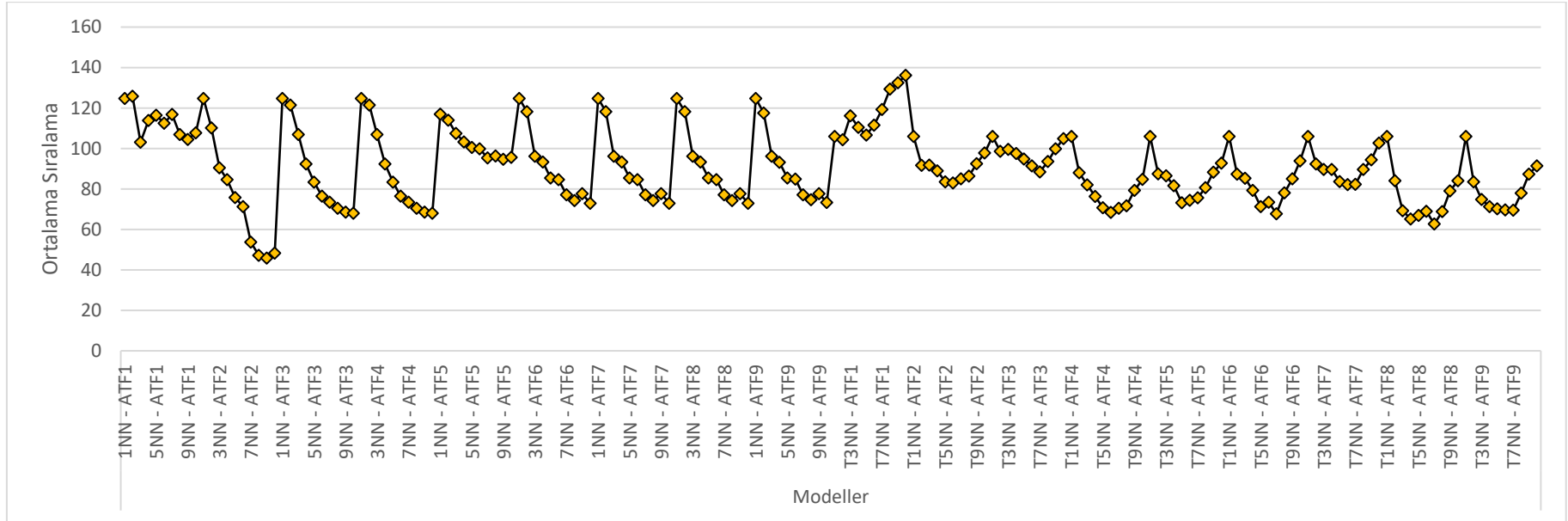
Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9	
<i>Concrete</i>	<b>KNN</b>	<b>1</b>	8,43	8,43	8,43	8,43	9,51	8,43	8,43	8,43	8,43	
		<b>2</b>	7,97	7,60	8,11	8,11	9,05	7,65	7,65	7,65	7,65	
		<b>3</b>	7,90	7,36	7,83	7,83	8,84	7,31	7,31	7,31	7,31	
		<b>4</b>	8,09	7,40	7,75	7,75	8,80	7,37	7,37	7,37	7,37	
		<b>5</b>	8,16	7,42	7,68	7,68	8,77	7,45	7,45	7,45	7,45	
		<b>6</b>	8,26	7,45	7,63	7,63	8,74	7,50	7,50	7,50	7,50	
		<b>7</b>	8,54	7,56	7,63	7,63	8,76	7,59	7,59	7,59	7,59	
		<b>8</b>	8,59	7,60	7,63	7,63	8,77	7,65	7,65	7,65	7,65	
		<b>9</b>	8,75	7,66	7,63	7,63	8,78	7,69	7,69	7,69	7,69	
		<b>10</b>	8,80	7,68	7,63	7,63	8,79	7,74	7,74	7,74	7,74	
	<b>TKNN</b>	<b>1</b>	8,24	8,24	8,24	8,24	8,24	8,24	8,24	8,24	8,24	8,24
		<b>2</b>	8,13	8,03	8,06	7,98	8,02	7,80	7,83	7,77	7,78	
		<b>3</b>	8,30	8,04	8,06	7,98	8,03	7,77	7,80	7,73	7,76	
		<b>4</b>	8,53	8,11	8,13	8,06	8,12	7,85	7,88	7,80	7,84	
		<b>5</b>	8,70	8,18	8,19	8,13	8,19	7,92	7,94	7,87	7,91	
		<b>6</b>	8,82	8,23	8,24	8,18	8,24	7,97	7,98	7,92	7,97	
		<b>7</b>	8,93	8,27	8,28	8,22	8,29	8,02	8,03	7,97	8,03	
		<b>8</b>	9,05	8,32	8,32	8,27	8,34	8,08	8,09	8,03	8,09	
		<b>9</b>	9,17	8,36	8,36	8,32	8,39	8,14	8,15	8,10	8,16	
		<b>10</b>	9,33	8,43	8,42	8,38	8,45	8,23	8,23	8,18	8,24	
<i>Ele2</i>	<b>KNN</b>	<b>1</b>	180,60	180,60	180,60	180,60	213,28	180,60	180,60	180,60	180,60	
		<b>2</b>	165,54	166,38	167,36	167,36	201,49	165,38	165,38	165,38	165,38	
		<b>3</b>	215,72	178,84	202,69	202,69	235,82	164,22	164,22	164,22	164,22	
		<b>4</b>	255,23	195,45	209,45	209,45	243,12	158,80	158,80	158,80	158,80	
		<b>5</b>	287,64	210,94	215,96	215,96	249,20	158,90	158,90	158,90	158,90	
		<b>6</b>	283,26	212,58	216,32	216,32	251,21	158,40	158,40	158,40	158,40	
		<b>7</b>	284,41	214,47	216,24	216,24	252,77	158,92	158,92	158,92	158,92	
		<b>8</b>	298,67	219,80	218,63	218,63	255,80	162,90	162,90	162,90	162,90	
		<b>9</b>	298,62	221,39	219,27	219,27	256,35	162,73	162,73	162,73	162,73	
		<b>10</b>	303,23	224,17	220,60	220,60	258,01	161,23	161,23	161,23	161,23	
	<b>TKNN</b>	<b>1</b>	209,47	209,47	209,47	209,47	209,47	209,47	209,47	209,47	209,47	
		<b>2</b>	275,14	252,54	252,24	251,48	252,91	196,85	196,93	196,71	196,79	
		<b>3</b>	297,68	268,21	267,91	266,94	268,57	192,59	192,73	192,42	192,46	
		<b>4</b>	304,82	272,98	272,79	271,72	273,24	191,78	191,87	191,63	191,71	
		<b>5</b>	311,13	275,98	275,87	274,79	276,16	191,30	191,26	191,16	191,35	
		<b>6</b>	316,40	278,52	278,46	277,38	278,65	190,59	190,48	190,44	190,71	
		<b>7</b>	319,16	279,95	279,92	278,86	280,06	189,01	188,88	188,85	189,17	
		<b>8</b>	321,67	281,31	281,30	280,27	281,41	187,94	187,84	187,74	188,08	
		<b>9</b>	324,08	282,66	282,65	281,65	282,76	187,15	187,08	186,91	187,26	
		<b>10</b>	327,10	284,30	284,28	283,32	284,41	186,79	186,74	186,53	186,88	

**EK 9 Aykırı Değer Analizi Yapılan Veri Kümelerinde KNN ve TKNN için RMSE Değerleri (Devamı)**

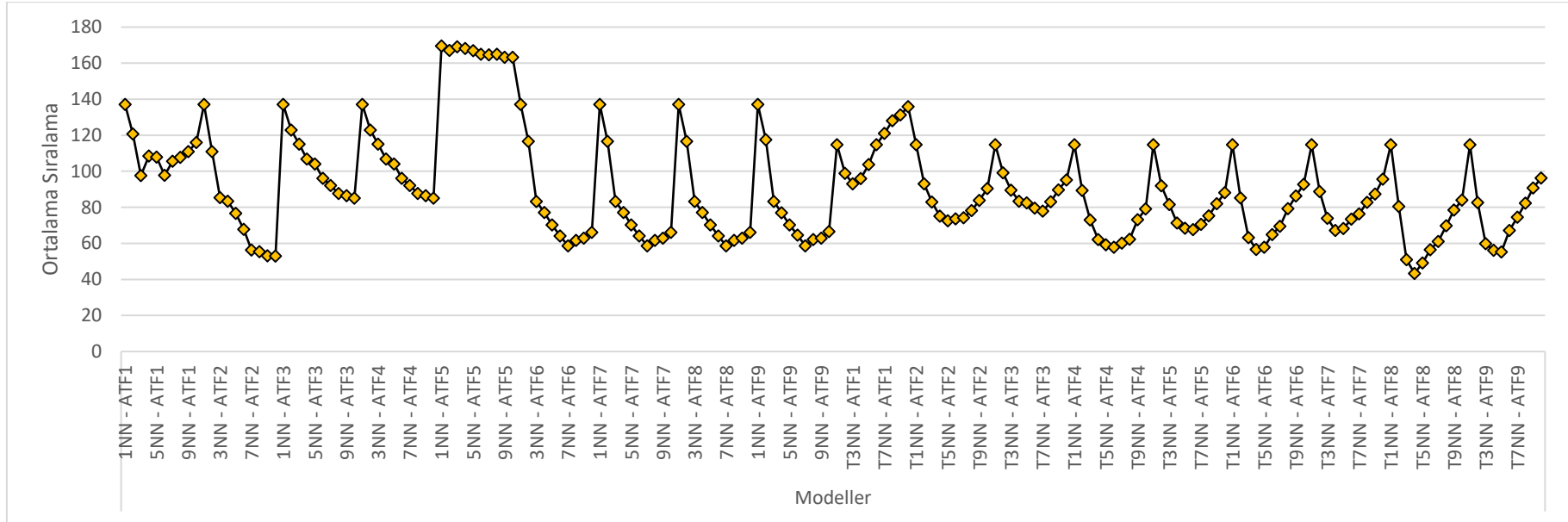
Veri Kümesi	Model	K	ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
<i>Friedman</i>	KNN	1	2,48	2,48	2,48	2,48	2,73	2,48	2,48	2,48	2,48
		2	2,13	2,12	2,28	2,28	2,57	2,11	2,11	2,11	2,11
		3	2,00	1,98	2,17	2,17	2,48	1,98	1,98	1,98	1,98
		4	1,94	1,91	2,10	2,10	2,43	1,91	1,91	1,91	1,91
		5	1,87	1,85	2,04	2,04	2,38	1,84	1,84	1,84	1,84
		6	1,86	1,83	2,01	2,01	2,35	1,82	1,82	1,82	1,82
		7	1,84	1,81	1,98	1,98	2,33	1,81	1,81	1,81	1,81
		8	1,82	1,79	1,95	1,95	2,31	1,79	1,79	1,79	1,79
		9	1,82	1,78	1,94	1,94	2,30	1,78	1,78	1,78	1,78
		10	1,83	1,78	1,93	1,93	2,29	1,79	1,79	1,79	1,79
	TKNN	1	2,04	2,04	2,04	2,04	2,04	2,04	2,04	2,04	2,04
		2	1,89	1,90	1,91	1,90	1,90	1,88	1,89	1,88	1,88
		3	1,84	1,84	1,84	1,84	1,84	1,82	1,82	1,82	1,82
		4	1,81	1,81	1,81	1,81	1,81	1,79	1,79	1,79	1,80
		5	1,81	1,79	1,79	1,79	1,80	1,78	1,78	1,78	1,79
		6	1,81	1,78	1,78	1,78	1,79	1,78	1,78	1,78	1,79
		7	1,83	1,78	1,78	1,78	1,79	1,80	1,79	1,79	1,81
		8	1,85	1,79	1,78	1,79	1,80	1,81	1,80	1,81	1,82
		9	1,87	1,79	1,78	1,79	1,80	1,83	1,81	1,83	1,84
		10	1,89	1,80	1,79	1,80	1,81	1,84	1,83	1,84	1,86



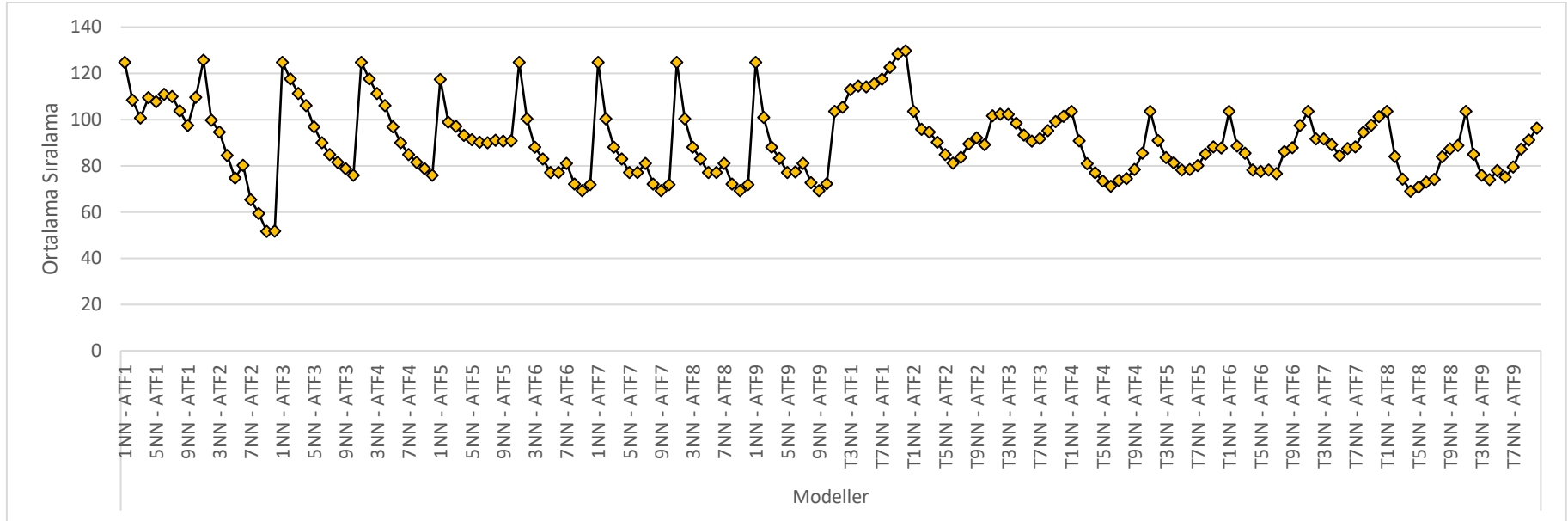
**EK 10 Orijinal Veri Kümelerine Komşuluk Tabanlı Yöntemlerin Uygulanması ile Elde Edilen MAPE Sonuçlarına Göre  $R_j$  Değerleri**



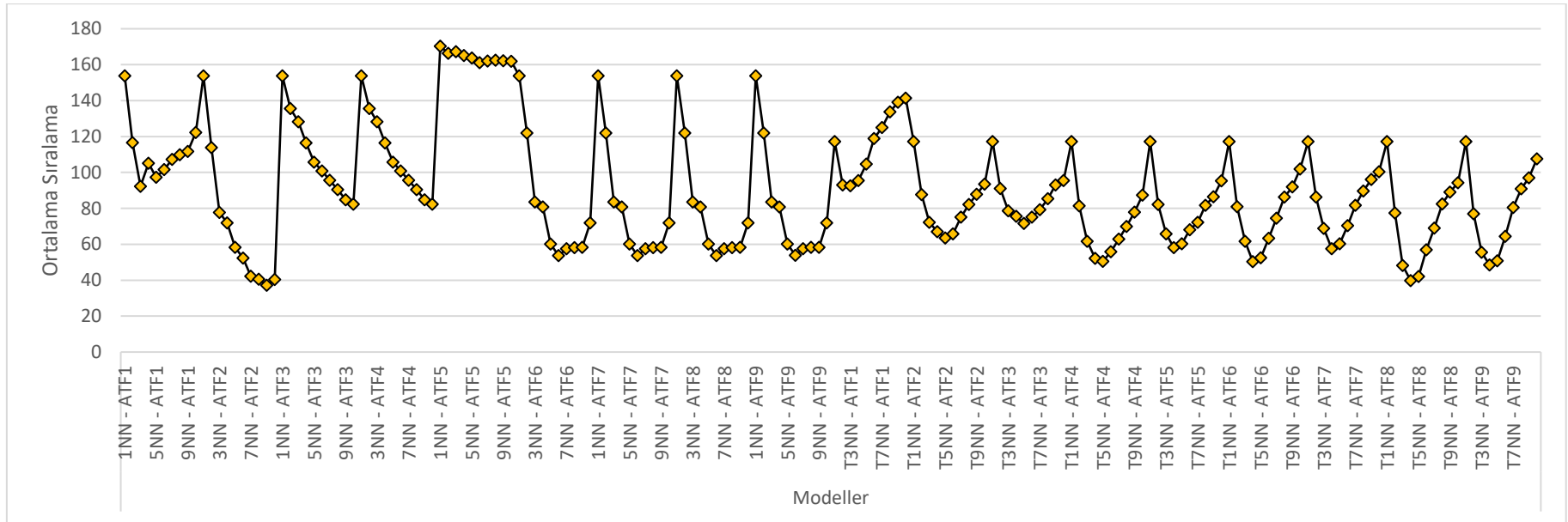
**EK 11 Orijinal Veri Kümelerine Komşuluk Tabanlı Yöntemlerin Uygulanması ile Elde Edilen RMSE Sonuçlarına Göre  $R_j$  Değerleri**



**EK 12 Aykırı Değer Analizi Yapılan Veri Kümelerine Komşuluk Tabanlı Yöntemlerin Uygulanması ile Elde Edilen MAPE**  
**Sonuçlarına Göre  $R_j$  Değerleri**



**EK 13 Aykırı Değer Analizi Yapılan Veri Kümelerine Komşuluk Tabanlı Yöntemlerin Uygulanması ile Elde Edilen RMSE Sonuçlarına Göre  $R_j$  Değerleri**



**EK 14 Kalıp Veri Kümesinin Komşuluk Tabanlı Yöntemler için MAPE Sonuçları  
(%)**

Veri Kümesi	Model	K	Kullanılan Ağırlıklı Tahmin Fonksiyonu								
			ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
Kalıp	KNN	1	31,32	31,32	31,32	31,32	30,46	31,32	31,32	31,32	31,32
		2	30,44	29,43	29,95	29,95	28,83	29,83	29,83	29,83	29,83
		3	30,51	29,72	29,43	29,43	28,23	29,90	29,90	29,90	29,90
		4	30,78	29,61	29,03	29,03	27,88	29,79	29,79	29,79	29,79
		5	31,08	29,98	29,01	29,01	27,72	30,53	30,53	30,53	30,53
		6	30,88	29,62	28,78	28,78	27,58	30,40	30,40	30,40	30,40
		7	31,38	29,80	28,96	28,96	27,79	30,91	30,91	30,91	30,91
		8	32,29	30,14	29,06	29,06	27,90	31,70	31,70	31,70	31,70
		9	31,47	29,74	28,79	28,79	27,62	31,05	31,05	31,05	31,05
		10	31,61	29,64	28,70	28,70	<b>27,50</b>	31,13	31,13	31,13	31,13
	TKNN	1	<b>28,81</b>	<b>28,81</b>	<b>28,81</b>	<b>28,81</b>	<b>28,81</b>	<b>28,81</b>	<b>28,81</b>	<b>28,81</b>	<b>28,81</b>
		2	30,24	29,78	29,76	29,76	29,82	29,80	29,79	29,80	29,85
		3	30,18	29,69	29,64	29,67	29,72	29,74	29,66	29,68	29,78
		4	30,37	29,59	29,53	29,57	29,88	29,65	29,83	29,83	29,94
		5	30,94	29,63	29,58	29,57	30,28	29,68	30,22	30,22	30,37
		6	31,33	29,64	29,60	29,59	30,63	29,70	30,55	30,58	30,72
		7	31,64	29,70	29,65	29,65	30,87	29,76	30,79	30,82	30,95
		8	31,97	29,78	29,74	29,73	31,21	29,83	31,14	31,16	31,28
		9	32,23	29,88	29,83	29,84	31,45	29,93	31,38	31,42	31,53
		10	32,45	29,99	29,94	29,96	31,65	30,05	31,59	31,64	31,72

### EK 15 Kalıp Veri Kümesinin Komşuluk Tabanlı Yöntemler için RMSE Sonuçları

Veri Kümesi	Model	K	Kullanılan Ağırlıklı Tahmin Fonksiyonu								
			ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
Kalıp	KNN	1	1278,57	1278,57	1278,57	1278,57	1278,57	1314,84	1278,57	1278,57	1278,57
		2	1211,30	1185,08	1209,93	1209,93	1203,75	1252,61	1203,75	1203,75	1203,75
		3	1201,90	1165,37	1181,35	1181,35	1187,49	1229,70	1187,49	1187,49	1187,49
		4	1233,19	1160,81	1154,89	1154,89	1175,21	1208,85	1175,21	1175,21	1175,21
		5	1253,76	1172,29	1146,18	1146,18	1216,41	1204,81	1216,41	1216,41	1216,41
		6	1265,19	1178,45	1140,60	1140,60	1224,40	1200,51	1224,40	1224,40	1224,40
		7	1277,00	1183,32	1135,24	1135,24	1242,77	1195,83	1242,77	1242,77	1242,77
		8	1303,77	1195,66	1135,65	1135,65	1265,77	1197,99	1265,77	1265,77	1265,77
		9	1274,10	1182,75	1125,07	1125,07	1238,19	1189,10	1238,19	1238,19	1238,19
		10	1288,82	1184,83	<b>1120,05</b>	<b>1120,05</b>	1239,61	1184,73	1239,61	1239,61	1239,61
	TKNN	1	<b>1153,01</b>	<b>1153,01</b>	<b>1153,01</b>	<b>1153,01</b>	<b>1153,01</b>	<b>1153,01</b>	<b>1153,01</b>	<b>1153,01</b>	<b>1153,01</b>
		2	1175,97	1154,87	1153,98	1154,50	1155,80	1158,60	1157,21	1158,40	1160,09
		3	1212,43	1165,10	1163,51	1163,78	1166,90	1174,18	1171,57	1173,14	1176,98
		4	1242,59	1176,93	1175,16	1175,10	1179,01	1198,96	1196,15	1196,95	1201,99
		5	1259,67	1183,76	1181,65	1181,41	1185,94	1215,57	1212,49	1212,73	1218,80
		6	1268,81	1185,42	1183,47	1183,15	1187,64	1222,01	1219,12	1219,01	1225,04
		7	1275,54	1186,18	1184,41	1183,81	1188,25	1224,54	1222,15	1221,68	1227,17
		8	1287,03	1189,94	1188,50	1187,61	1191,65	1231,80	1230,01	1229,21	1233,86
		9	1296,16	1193,64	1192,21	1191,27	1195,19	1239,63	1238,05	1237,41	1241,46
		10	1305,28	1197,85	1196,69	1195,54	1199,12	1247,76	1246,41	1246,00	1249,25

**EK 16 Öznitelik Seçimi Sonrası Kalıp 2 Veri Kümesinin Komşuluk Tabanlı  
Yöntemler İçin MAPE Sonuçları (%)**

Veri Kümesi	Model	K	Kullanılan Ağırlıklı Tahmin Fonksiyonu								
			ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
Kalıp (öznitelik seçimi)	KNN	1	31,62	31,62	31,62	31,62	30,49	31,62	31,62	31,62	31,62
		2	29,19	28,98	29,91	29,91	28,64	29,31	29,31	29,31	29,31
		3	26,99	27,12	28,60	28,60	27,08	27,42	27,42	27,42	27,42
		4	28,45	27,46	28,02	28,02	26,61	28,22	28,22	28,22	28,22
		5	29,68	28,33	28,35	28,35	26,85	29,22	29,22	29,22	29,22
		6	29,35	28,19	28,03	28,03	26,61	28,97	28,97	28,97	28,97
		7	29,33	27,96	27,80	27,80	26,24	28,59	28,59	28,59	28,59
		8	29,27	27,90	27,68	27,68	25,91	28,61	28,61	28,61	28,61
		9	30,16	28,04	27,65	27,65	25,89	29,19	29,19	29,19	29,19
		10	30,25	28,05	27,58	27,58	<b>25,83</b>	29,34	29,34	29,34	29,34
	TKNN	1	27,91	27,91	27,91	27,91	27,91	27,91	27,91	27,91	27,91
		2	27,69	27,56	27,45	27,52	27,67	27,66	27,53	27,61	27,78
		3	28,11	27,58	<b>27,42</b>	27,53	27,74	27,83	27,65	27,78	28,01
		4	28,77	27,87	27,67	27,84	28,07	28,25	28,05	28,21	28,45
		5	29,28	27,96	27,75	27,92	28,17	28,31	28,10	28,27	28,53
		6	29,64	27,92	27,71	27,88	28,14	28,56	28,37	28,48	28,75
		7	29,93	28,00	27,80	27,93	28,21	28,85	28,66	28,76	29,05
		8	30,10	28,09	27,91	28,02	28,30	29,08	28,88	28,98	29,27
		9	30,15	28,16	27,98	28,06	28,35	29,22	29,02	29,12	29,41
		10	30,25	28,23	28,05	28,13	28,42	29,38	29,18	29,29	29,57

**EK 17 Öznitelik Seçimi Sonrası Kalıp 2 Veri Kümesinin Komşuluk Tabanlı  
Yöntemler için RMSE Sonuçları**

Veri Kümesi	Model	K	Kullanılan Ağırlıklı Tahmin Fonksiyonu								
			ATF 1	ATF 2	ATF 3	ATF 4	ATF 5	ATF 6	ATF 7	ATF 8	ATF 9
Kalıp (öznitelik seçimi)	KNN	1	1299,74	1299,74	1299,74	1299,74	1294,92	1299,74	1299,74	1299,74	1299,74
		2	1212,98	1203,16	1241,39	1241,39	1240,23	1226,53	1226,53	1226,53	1226,53
		3	1111,98	1107,74	1176,63	1176,63	1177,93	1131,67	1131,67	1131,67	1131,67
		4	1113,09	1088,11	1143,69	1143,69	1145,33	1116,64	1116,64	1116,64	1116,64
		5	1148,94	1098,23	1134,46	1134,46	1132,55	1122,11	1122,11	1122,11	1122,11
		6	1177,70	1108,13	1125,83	1125,83	1126,55	1130,39	1130,39	1130,39	1130,39
		7	1197,46	1112,71	1115,25	1115,25	1116,86	1130,45	1130,45	1130,45	1130,45
		8	1199,95	1111,45	1105,38	1105,38	1106,48	1133,90	1133,90	1133,90	1133,90
		9	1238,13	1125,49	1103,12	1103,12	1102,60	1153,88	1153,88	1153,88	1153,88
		10	1229,09	1119,85	1094,54	1094,54	<b>1093,32</b>	1146,67	1146,67	1146,67	1146,67
	TKNN	1	1149,92	1149,92	1149,92	1149,92	1149,92	1149,92	1149,92	1149,92	1149,92
		2	1100,65	1100,08	1098,13	1099,23	1102,13	1101,93	1099,85	1100,92	1104,27
		3	1110,68	1087,82	<b>1084,23</b>	1086,66	1091,71	1090,39	1086,45	1089,36	1094,48
		4	1152,12	1101,25	1096,89	1099,57	1105,82	1104,81	1099,87	1103,61	1110,00
		5	1185,50	1112,27	1108,42	1109,76	1116,57	1117,64	1113,11	1115,64	1122,55
		6	1202,34	1117,29	1113,65	1114,06	1121,40	1125,39	1121,09	1122,39	1130,03
		7	1212,04	1120,44	1116,95	1116,66	1124,40	1130,99	1126,94	1127,52	1135,41
		8	1216,16	1122,55	1119,24	1118,72	1126,38	1135,19	1131,40	1131,61	1139,28
		9	1216,41	1123,30	1120,14	1119,40	1127,03	1137,52	1133,85	1134,07	1141,48
		10	1217,32	1124,09	1121,05	1120,20	1127,76	1141,17	1137,75	1137,94	1144,89



## ÖZGEÇMİŞ

Adı Soyadı : Gözde ESER  
Doğum Yeri ve Tarihi : Edirne 24.11.1992  
Yabancı Dil : İngilizce

Eğitim Durumu  
Lise : Edirne Anadolu Öğretmen Lisesi  
Lisans : Bursa Uludağ Üniversitesi – Endüstri Mühendisliği

Çalıştığı Kurum/Kurumlar : Ermetal Otomotiv ve Eşya Sanayi Tic. A.Ş.

İletişim (e-posta) : ongelengozde@gmail.com

Yayımları :

**Eser, G., İnkaya, T., Ekdik, O. 2019.** Predicting the Production Times in Die Manufacturing Process Using Data Mining Methods. In 10th International Symposium on Intelligent Manufacturing and Service Systems, Sakarya, Turkey, September, 2019.

**Gül, G., Erol, B., Öngelen, G., Eser, S., Çetinkaya, Ç., Özmutlu, H.C., Özmutlu, S., Gökçedağlı, M., Erhuy, C.G. 2016.** Ambar depolama maksimizasyonu, *Endüstri Mühendisliği*, 27(4): 26-38.