



**TOPLULUK ÖĞRENME YÖNTEMLERİ İLE
HANEHALKI TOPLAM ENERJİ
HARCAMALARININ TAHMİN EDİLMESİ**

Esmâ KESRİKLİOĞLU

**Doktora Tezi
Ekonometri Anabilim Dalı
Prof. Dr. Erkan OKTAY
Doç. Dr. Abdulkerim KARAASLAN
2022
Her Hakkı Saklıdır**

**T.C.
ATATÜRK ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANA BİLİM DALI**

Esmâ KESRİKLİOĞLU

**TOPLULUK ÖĞRENME YÖNTEMLERİ İLE
HANEHALKI TOPLAM ENERJİ
HARCAMALARININ TAHMİN EDİLMESİ**

DOKTORA TEZİ

TEZ YÖNETİCİSİ

Prof. Dr. Erkan OKTAY

Ortak Tez Danışmanı

Doç. Dr. Abdulkerim KARAASLAN

ERZURUM – 2022



TEZ BEYAN FORMU

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

BİLDİRİM

Atatürk Üniversitesi Lisansüstü Eğitim ve Öğretim Uygulama Esaslarının ilgili maddelerine göre hazırlamış olduğum "TOPLULUK ÖĞRENME YÖNTEMLERİ İLE HANEHALKI TOPLAM ENERJİ HARCAMALARININ TAHMİN EDİLMESİ " adlı tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kâğıt ve elektronik kopyalarının aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım.

Gereğini bilgilerinize arz ederim *.

- ☐ Tezimin/Raporumun tamamı her yerden erişime açılabilir.
- ☐ Tezimin/Raporumun makale için **altı ay**, patent için **iki yıl** süreyle erişiminin ertelenmesini istiyorum.

14/10/2022

Ash ıslak imzalıdır

Esmâ KESRİKLİOĞLU

*** LİSANSÜSTÜ TEZLERİN ELEKTRONİK ORTAMDA TOPLANMASI, DÜZENLENMESİ VE ERİŞİME AÇILMASINA İLİŞKİN YÖNERGE**

ÜÇÜNCÜ BÖLÜM

Çeşitli ve Son Hükümler

Lisansüstü tezlerin erişime açılmasının ertelenmesi **MADDE 6– (1)** Lisansüstü teze ilgili **patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda**, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu **iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.**

(2) Yeni teknik, materyal ve metotların kullanıldığı, henüz **makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış** ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile **altı ayı aşmamak üzere tezin erişime açılması engellenebilir.**

Gizlilik dereceli tezler MADDE 7– (1) Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.

(2) Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir.



SOSYAL BİLİMLER ENSTİTÜSÜ
Graduate School of Social Sciences

TEZ KABUL TUTANAĞI

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Prof. Dr. Erkan Oktay danışmanlığında, Esma Kesriklioğlu tarafından hazırlanan bu çalışma 14/10/2022 tarihinde aşağıda isimleri yazılı jüri tarafından Ekonometri Anabilim Dalı'nda Doktora Tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Üstün ÖZEN

İmza: Aslı ıslak imzalıdır

Jüri Üyesi : Prof. Dr. Erkan OKTAY

İmza: Aslı ıslak imzalıdır

Jüri Üyesi : Doç. Dr. Abdülkerim KARAASLAN

İmza: Aslı ıslak imzalıdır

Jüri Üyesi : Doç. Dr. Ali Kemal ÇELİK

İmza: Aslı ıslak imzalıdır

Jüri Üyesi : Dr. Öğr. Üyesi Emre YAKUT

İmza: Aslı ıslak imzalıdır

Prof. Dr. Sait UYLAŞ

Enstitü Müdürü

Aslı ıslak imzalıdır

İÇİNDEKİLER

ÖZET.....	V
ABSTRACT	VI
KISALTMALAR DİZİNİ	VII
ŞEKİLLER DİZİNİ	IX
TABLOLAR DİZİNİ	X
ÖNSÖZ.....	XIII
GİRİŞ	1

BİRİNCİ BÖLÜM

GENEL BİLGİLER

1.1. YAPAY ZEKA (ARTIFICIAL INTELLIGENCE- AI)	5
1.2. VERİ MADENCİLİĞİ (DATA MINING - DM)	8
1.2.1. Veri Bilimi.....	8
1.2.2. Veri Tabanlarında Bilgi Keşfi Süreci	9
1.2.3. Veri Madenciliği Kavramı.....	10
1.2.4. Veri Tabanlarında Bilgi Keşfi Sürecinde Kullanılan Metodolojiler	11
1.2.4.1. Veri Tabanlarında Bilgi Keşfi (KDD/ Knowledge Discovery in Databases).....	12
1.2.4.1.1. Veri Seçimi	12
1.2.4.1.2. Veri Ön işleme	13
1.2.4.1.3. Veri Dönüştürme.....	12
1.2.4.1.4. Veri Madenciliği	12
1.2.4.1.5. Yorumlama/ Değerlendirme	12
1.2.4.2. SEMMA.....	13
1.2.4.2.1. Örneklemeye	13
1.2.4.2.2. Keşfetme	14
1.2.4.2.3. Dönüştürme.....	14
1.2.4.2.4. Modelleme	14
1.2.4.2.5. Değerlendirme.....	14
1.2.4.3. CRISP-DM (Cross Industry Standard Process for Data Mining /Çapraz Endüstri Standardı Süreci	14

1.2.4.3.1. İş Süreçlerinin Anlaşılması (Business Understanding).....	15
1.2.4.3.2. Verinin Anlaşılması (Data Understanding)	15
1.2.4.3.3. Veri Ön İşleme (Preprocessing).....	15
1.2.4.3.4. Modelleme (Modelling)	16
1.2.4.3.5. Değerlendirme (Evaluation).....	16
1.2.4.3.6. Uygulama (Deployment)	16
1.2.5. Öznitelik Seçimi (Feature Selection)	17
1.2.5.1. İleriye Doğru Seçim (Forward Selection)	18
1.2.5.2. Geriye Doğru Eleme (Backward Elimination)	18
1.2.5.3. Evrimsel Strateji ile Optimize Etme (Optimize Selection - Evolutionary)	19
1.2.6. Çapraz Geçerlilik (Cross-Validation) Yöntemleri	19

İKİNCİ BÖLÜM

MAKİNE ÖĞRENMESİ ve TOPLULUK ÖĞRENME YÖNTEMLERİ

2.1. MAKİNE ÖĞRENMESİ (MACHINE LEARNING - ML)	21
2.1.1. Makine Öğrenmesi Türleri	24
2.1.1.1. Denetimli Öğrenme (Supervised Learning).....	25
2.1.1.2. Denetimsiz Öğrenme (Unsupervised Learning)	25
2.1.1.3. Yarı Denetimli Öğrenme (Semi-Supervised Learning).....	26
2.1.1.4. Takviyeli/ Pekiştirmeli Öğrenme (Reinforcement Learning).....	27
2.2. SINIFLANDIRMA (CLASSIFICATION)	27
2.2.1. K-en yakın komşu (K-Nearest Neighbor – KNN).....	28
2.2.2. Naive Bayes.....	30
2.2.3. Karar Ağacı (Decision Tree -DT)	30
2.2.4. Random Forest (RF)	32
2.2.5. Gradyan Artırmalı Ağaç (Gradient Boosted Tree - GBT)	33
2.2.6. Derin Sinir Ağları (Deep Neural Networks)	34
2.3. REGRESYON ANALİZİ	36
2.4. KÜMELEME ANALİZİ	36
2.5. TOPLULUK ÖĞRENME (ENSEMBLE LEARNING).....	37
2.5.1. Bagging (Toralama) Topluluk Öğrenme Yöntemi	40

2.5.2. Boosting (Artırma) Topluluk Öğrenme Yöntemi.....	41
2.5.3. Voting (Oylama) Topluluk Öğrenme Yöntemi	43
2.5.4. Stacking (İstifleme) Topluluk Öğrenme Yöntemi.....	44

ÜÇÜNCÜ BÖLÜM

TOPLULUK ÖĞRENME YÖNTEMLERİ İLE HANEHALKI TOPLAM ENERJİ HARCAMALARININ TAHMİN EDİLMESİ

3.1. LİTERATÜR TARAMASI	49
3.1.1. İstatistiksel Yöntemler ve Ekonometrik Modeller Kullanılarak Hanehalkı Enerji Tüketimi Alanında Yapılan Çalışmalar	50
3.1.2. Yapay Zeka Yöntemleri Kullanılarak Hanehalkı Enerji Tüketimi Alanında Yapılan Çalışmalar	53
3.2. ARAŞTIRMANIN VERİ SETİ.....	60
3.3. ARAŞTIRMADA KULLANILAN YÖNTEMLER.....	60
3.4. ARAŞTIRMADA KULLANILAN DEĞİŞKENLER.....	60
3.5. ARAŞTIRMA VERİLERİNE CRISP-DM METODOLOJİSİNİN UYGULANMASI	62
3.5.1. İş Süreçlerinin Anlaşılması (Business Understanding)	63
3.5.2. Verinin Anlaşılması (Data Understanding)	63
3.5.3. Veri Ön İşleme (Preprocessing)	63
3.5.4. Modelleme (Modelling)	66
3.5.5. Değerlendirme (Evaluation)	67
3.5.5.1. Performans Değerlendirme Metrikleri.....	68
3.5.5.1.1. Doğruluk (Accuracy)	68
3.5.5.1.2. Kappa Katsayısı	68
3.5.5.1.3. Duyarlılık (Sensitivity, recall)	69
3.5.5.1.4. Kesinlik (Precision)	69
3.5.5.1.5. Duyarlılık ağırlıklı ortalama (Weighted_mean_recall).....	69
3.5.5.1.6. Kesinlik ağırlıklı ortalama (Weighted_mean_precision).....	69
3.5.6. Uygulama (Deployment)	69
3.6. BULGULAR	70

3.6.1. Makine Öğrenmesi Sınıflandırıcılarının Performanslarının	
Değerlendirilmesi	71
3.6.2. Boosting Topluluk Öğrenme Yöntemi ile Sınıflandırıcıların Performanslarının	
Değerlendirilmesi	74
3.6.3. Bagging Topluluk Öğrenme Yöntemi ile Sınıflandırıcıların Performanslarının	
Değerlendirilmesi	77
3.6.4. Voting Topluluk Öğrenme Yöntemi ile Sınıflandırıcıların Performanslarının	
Değerlendirilmesi	81
3.6.5. Stacking Topluluk Öğrenme Yöntemi ile Sınıflandırıcıların Performanslarının	
Değerlendirilmesi	85
SONUÇ ve TARTIŞMA	91
KAYNAKÇA	96
ÖZGEÇMİŞ.....	114

ÖZET**DOKTORA TEZİ****TOPLULUK ÖĞRENME YÖNTEMLERİ İLE HANEHALKI TOPLAM ENERJİ
HARCAMALARININ TAHMİN EDİLMESİ****Esmâ KESRİKLİOĞLU****Tez Danışmanı: Prof. Dr. Erkan OKTAY****Ortak Tez Danışmanı: Doç. Dr. Abdulkerim KARAASLAN****2022, 114 Sayfa****Jüri: Prof. Dr. Erkan OKTAY****Prof. Dr. Üstün ÖZEN****Doç. Dr. Abdülkerim KARAASLAN****Doç. Dr. Ali Kemal ÇELİK****Dr. Öğr. Üyesi Emre YAKUT**

Bu çalışmada, hanehalkı toplam enerji harcamalarının sınıflandırılması için yüksek performanslı bir topluluk öğrenme modelinin oluşturulması amaçlanmaktadır. Bu amaç doğrultusunda, Türkiye İstatistik Kurumu (TÜİK)'in yayınladığı Hanehalkı Bütçe Araştırması anketi 2019 veri seti kullanılarak, 11.521 hanenin hanehalkı tüketim verileri incelenmiştir. Hanehalkı toplam enerji harcamaları ile doğrudan ilişkili değişkenlere ek olarak, literatür çerçevesinde ve uzman görüşü rehberliğinde yeni değişkenler oluşturulmuştur. Hazırlanan veri, açık kaynak kodlu RapidMiner yazılım programı ile veri ön işleme, öznitellik seçimi, modelleme, tahmin ve performans değerlendirme aşamalarından geçirilmiştir. Makine öğrenmesi ve topluluk öğrenme yöntemlerinin sınıflandırma performansları karşılaştırılmıştır. Çalışmada, K-En Yakın Komşu, Karar ağacı, Naive Bayes, Derin Sinir Ağları, Random Forest, Gradyan Artırmalı Ağaç sınıflandırıcılarının yanında bagging (torbalama), boosting (artırma), voting (oylama) ve stacking (istifleme) topluluk öğrenme yöntemleri kullanılmıştır. Hanehalkı toplam enerji harcamalarının tahmin edilmesinde, en yüksek sınıflama performansını K-En Yakın Komşu, Karar ağacı, Naive Bayes, Derin Sinir Ağları, Random Forest, Gradyan Artırmalı Ağaç sınıflandırıcılarının birleştirilmesinden oluşan ALL model sağlamıştır. Stacking topluluk öğrenme yöntemi ALL modelde ve bagging topluluk öğrenme yöntemi Derin Sinir Ağları sınıflandırıcısında, topluluk öğrenme yöntemleri arasında en yüksek performansı elde etmiştir. Stacking ve bagging topluluk öğrenme yönteminin her ikisinin de doğruluk değeri 0.984, kappa değeri 0.979, duyarlılık ağırlıklı ortalama değeri 0.984 ve kesinlik ağırlıklı ortalama değeri 0.984 olarak elde edilmiştir. Sonuçlar; topluluk öğrenme yöntemlerinin, hanehalkı toplam enerji harcamalarını tahmin etmek için bireysel makine öğrenmesi yöntemlerini önemli ölçüde iyileştirebildiğini göstermektedir.

Anahtar Kelimeler: Topluluk öğrenme, sınıflandırma, hanehalkı toplam enerji harcaması, makine öğrenmesi, yapay zeka

ABSTRACT**Ph. D. DISSERTATION****PREDICTION of HOUSEHOLD TOTAL ENERGY EXPENDITURE with
ENSEMBLE LEARNING METHODS****Esma KESRİKLİOĞLU****Advisor: Prof. Dr. Erkan OKTAY****Co- Advisor: Assoc. Prof. Dr. Abdülkerim KARAASLAN****2022, 114 Sayfa****Jüri: Prof. Dr. Erkan OKTAY****Prof. Dr. Üstün ÖZEN****Assoc. Prof. Dr. Abdülkerim KARAASLAN****Assoc. Prof. Dr. Ali Kemal ÇELİK****Asst. Prof. Dr. Emre YAKUT**

In this study, it is aimed to create a high-performance ensemble learning model for the classification of total household energy expenditure. For this purpose, the household consumption data of 11,521 households were examined by using the Household Budget Survey 2019 data set published by the Turkish Statistical Institute (TUIK). In addition to the variables which directly related to household energy expenditures, new variables were created within the framework of the literature and under the guidance of expert opinion. The prepared data was passed through stages of data preprocessing, feature selection, modeling, prediction and performance evaluation with the open source RapidMiner software program. Classification performances of machine learning and ensemble learning methods were compared. In the study, besides K-Nearest Neighbor, Naive Bayes, Deep Neural Networks, Random Forest, Gradient Boosted Trees classifiers; bagging, boosting, voting and stacking ensemble learning methods are used. In the prediction of household total energy expenditure, the ALL model consisting of combining K-Nearest Neighbor, Decision Tree, Naive Bayes, Deep Neural Networks, Random Forest, Gradient Boosted Trees classifiers provided the highest classification performance. Stacking ensemble learning method in ALL model and bagging ensemble learning method in Deep Neural Networks classifier achieved the highest performance among ensemble learning methods. The accuracy value of both stacking and bagging ensemble learning methods was 0.984, kappa value was 0.979, W_M recall value was 0.984 and W_M precision value was 0.984. The results show that ensemble learning methods can significantly enhance individual machine learning methods for predict household total energy expenditures.

Keywords: Ensemble learning, classification, household energy expenditure, machine learning, artificial intelligence

KISALTMALAR DİZİNİ

AI	: Artificial Intelligence
ALASSO	: Adaptive Lasso
ANN	: Artificial Neural Networks
ARIMA	: Autoregressive Integrated Moving Average
CRISP-DM	: Cross Industry Standard Process for Data Mining
DL	: Deep Learning
DM	: Data Mining
DN	: Doğru Negatif
DP	: Doğru Pozitif
DT	: Decision Tree
ERF	: Extreme Random Forest
GBT	: Gradient Boosted Tree
IoT	: Nesnelerin İnterneti
KDD	: Knowledge Discovery in Databases
KNN	: K-Nearest Neighbor
LOF	: Local Outlier Faktör
LSTM	: Long Short-Term Memory Network
ML	: Machine Learning
NARX	: Non-Linear Autoregressive with Exogenous Input
NB	: Naive Bayes
OLS	: Ordinary Least Squares
RF	: Random Forest
SARIMA	: Seasonal Autoregressive Integrated Moving Average
SAS	: Statistical Analysis Software
SEMMA	: Sample, Explore, Modify, Model, and Assess.
SOM	: Self Organising Maps
SVM	: Support Vector Machine

TÜİK	: Türkiye İstatistik Kurumu
vb.	: Ve benzeri
XGBOOST	: Extreme Gradient Boosting
YN	: Yanlış Negatif
YP	: Yanlış Pozitif



ŞEKİLLER DİZİNİ

Şekil 1. Yapay Zeka, Makine Öğrenmesi ve Veri Bilimi Arasındaki İlişki Diagramı	7
Şekil 2. Bilgi Keşfi Süreci	12
Şekil 3. CRISP-DM Metodolojisi	15
Şekil 4. Makine Öğrenmesi Süreci	23
Şekil 5. Makine Öğrenmesi Türleri	24
Şekil 6. KNN Algoritması	29
Şekil 7. Karar Ağacı Algoritması	31
Şekil 8. Gradyan Artırmalı Ağaç Algoritması	34
Şekil 9. İleri Beslemeli Sinir Ağı Mimarisi	35
Şekil 10. Bagging Topluluk Öğrenme Yöntemi İşlem Aşamaları	41
Şekil 11. Boosting Topluluk Öğrenme Yöntemi İşlem Aşamaları	43
Şekil 12. Voting Topluluk Öğrenme Yöntemi İşlem Aşamaları	44
Şekil 13. Stacking Topluluk Öğrenme Yöntemi İşlem Aşamaları	46
Şekil 14. Hanehalkı Profilleri	59
Şekil 15. Öznitelik Seçimi	65
Şekil 16. 10-Katlı Çapraz Geçerlilik Yöntemi	66
Şekil 17. CRISP-DM Uygulama Süreci	70

TABLOLAR DİZİNİ

Tablo 1. Modellerinin Genel Özellikleri (Sadece Enerji Uygulamaları için Değil).....	48
Tablo 2. Değişkenlerin Listesi ve Tanımı	61
Tablo 3. Karışıklık Matrisi	67
Tablo 4. Knn Sınıflandırıcısının Performans Değerlendirmesi	71
Tablo 5. Naive Bayes Sınıflandırıcısının Performans Değerlendirmesi	72
Tablo 6. Karar Ağacı Sınıflandırıcısının Performans Değerlendirmesi	72
Tablo 7. Derin Sinir Ağları Sınıflandırıcısının Performans Değerlendirmesi.....	72
Tablo 8. Random Forest Sınıflandırıcısının Performans Değerlendirmesi	73
Tablo 9. Gradyan Artırmalı Ağaç Sınıflandırıcısının Performans Değerlendirmesi.....	73
Tablo 10. Makine Öğrenmesi Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması	74
Tablo 11. Boosting Topluluk Öğrenme Yöntemi İle KNN Sınıflandırıcısının Performanslarının Değerlendirilmesi	74
Tablo 12. Boosting Topluluk Öğrenme Yöntemi İle Naive Bayes Sınıflandırıcısının Performanslarının Değerlendirilmesi	75
Tablo 13. Boosting Topluluk Öğrenme Yöntemi İle Karar Ağacı Sınıflandırıcısının Performanslarının Değerlendirilmesi	75
Tablo 14. Boosting Topluluk Öğrenme Yöntemi İle Derin Sinir Ağları Sınıflandırıcısının Performanslarının Değerlendirilmesi	76
Tablo 15. Boosting Topluluk Öğrenme Yöntemi İle Random Forest Sınıflandırıcısının Performanslarının Değerlendirilmesi	76
Tablo 16. Boosting Topluluk Öğrenme Yöntemi İle Gradyan Artırmalı Ağaç Sınıflandırıcısının Performanslarının Değerlendirilmesi	77
Tablo 17. Boosting Topluluk Öğrenme Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması.....	77
Tablo 18. Bagging Topluluk Öğrenme Yöntemi İle KNN Sınıflandırıcısının Performanslarının Değerlendirilmesi	78
Tablo 19. Bagging Topluluk Öğrenme Yöntemi İle Naive Bayes Sınıflandırıcısının Performanslarının Değerlendirilmesi	78

Tablo 20. Bagging Topluluk Öğrenme Yöntemi İle Karar Ağacı Sınıflandırıcısının Performanslarının Değerlendirilmesi	79
Tablo 21. Bagging Topluluk Öğrenme Yöntemi İle Derin Sinir Ağları Sınıflandırıcısının Performanslarının Değerlendirilmesi	79
Tablo 22. Bagging Topluluk Öğrenme Yöntemi İle Random Forest Sınıflandırıcısının Performanslarının Değerlendirilmesi	80
Tablo 23. Bagging Topluluk Öğrenme Yöntemi İle Gradyan Artırmalı Ağaç Sınıflandırıcısının Performanslarının Değerlendirilmesi	80
Tablo 24. Bagging Topluluk Öğrenme Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması.....	81
Tablo 25. Voting Topluluk Öğrenme Yöntemi İle KNN Sınıflandırıcısının Performanslarının Değerlendirilmesi	81
Tablo 26. Voting Topluluk Öğrenme Yöntemi İle Naive Bayes Sınıflandırıcısının Performanslarının Değerlendirilmesi	82
Tablo 27. Voting Topluluk Öğrenme Yöntemi İle Karar Ağacı Sınıflandırıcısının Performanslarının Değerlendirilmesi	82
Tablo 28. Voting Topluluk Öğrenme Yöntemi İle Derin Sinir Ağları Sınıflandırıcısının Performanslarının Değerlendirilmesi	83
Tablo 29. Voting Topluluk Öğrenme Yöntemi İle Random Forest Sınıflandırıcısının Performanslarının Değerlendirilmesi	83
Tablo 30. Voting Topluluk Öğrenme Yöntemi İle Gradyan Artırmalı Ağaç Sınıflandırıcısının Performanslarının Değerlendirilmesi	84
Tablo 31. Voting Topluluk Öğrenme Yöntemi İle ALL Modelin Sınıflandırma Performanslarının Değerlendirilmesi	84
Tablo 32. Voting Topluluk Öğrenme Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması.....	85
Tablo 33. Stacking Topluluk Öğrenme Yöntemi İle KNN Sınıflandırıcısının Performanslarının Değerlendirilmesi	85
Tablo 34. Stacking Topluluk Öğrenme Yöntemi İle Naive Bayes Sınıflandırıcısının Performanslarının Değerlendirilmesi	86
Tablo 35. Stacking Topluluk Öğrenme Yöntemi İle Karar Ağacı Sınıflandırıcısının Performanslarının Değerlendirilmesi	86

Tablo 36. Stacking Topluluk Öğrenme Yöntemi İle Derin Sinir Ağları	
Sınıflandırıcısının Performanslarının Değerlendirilmesi	87
Tablo 37. Stacking Topluluk Öğrenme Yöntemi İle Random Forest	
Sınıflandırıcısının Performanslarının Değerlendirilmesi	87
Tablo 38. Stacking Topluluk Öğrenme Yöntemi İle Gradyan Artırmalı Ağaç	
Sınıflandırıcısının Performanslarının Değerlendirilmesi	88
Tablo 39. Stacking Topluluk Öğrenme Yöntemi İle ALL Modelin Sınıflandırma	
Performanslarının Değerlendirilmesi	88
Tablo 40. Stacking Topluluk Öğrenme Yöntemlerinin Sınıflandırma	
Performanslarının	89
Tablo 41. Topluluk Öğrenme Yöntemlerinin Sınıflandırma Performanslarının	
Karşılaştırılması	90

ÖNSÖZ

Doktora eğitimim süresince bilgi ve deneyimleriyle bana yol gösteren kıymetli danışmanım Sayın Prof. Dr. Erkan OKTAY'a, tezin uzun ve zorlu yazım sürecinde desteğini ve emeğini esirgemeyen ikinci danışmanım Sayın Doç. Dr. Abdulkerim KARAASLAN'a, tez izleme aşamalarındaki değerli görüşleriyle beni motive eden Sayın Prof. Dr. Üstün ÖZEN 'e, tez çalışma konumu belirlememde birikimiyle ufkumu genişleten ve tez yazım sürecimde de eşsiz katkıları bulunan Prof. Dr. Cemil ÇOLAK'a teşekkür ederim ve saygılarımı sunarım.

Hayatımın her anında olduğu gibi doktora eğitimim ve bilhassa tez çalışmam süresince varlıklarını daima yanımda hissettiğim kıymetli aileme emekleri ve bana olan güvenleri için sonsuz teşekkür ederim.

Erzurum-2022

Esmâ KESRİKLİOĞLU

GİRİŞ

Ülkelerin kalkınma hızı ve ekonomik gelişmelere bağlı olarak tüketim alışkanlıklarının yelpazesi kontrolsüz bir şekilde genişlemektedir. Son yıllarda yükselen yaşam standartları ile birlikte hanehalkı enerji tüketimi artmakta ve kaçınılmaz olarak, enerji talebinin de önemli ölçüde artmasına neden olmaktadır. Yükselen maliyetler, sınırlı doğal kaynaklar, hanehalkının verimli enerji tüketimi konusunda bilinçsiz olması da enerji tüketimini arttıran nedenler arasındadır. Ayrıca, tüm dünyayı etkisi altına almaya başlayan iklim krizi, nüfus artışı, Covid-19 pandemisi ile birlikte değişen alışkanlıklar, hanehalklarının enerji taleplerindeki artışı da etkilemektedir. Bu kontrolsüz artışlar, enerji kullanımı konusunda birçok yönden dünyamızın varlığını tehdit eden bir etken olarak karşımıza çıkmaktadır. Buna paralel olarak, dünya çapında önemli bir enerji kullanıcısı olan hanehalklarının enerji kullanımı hakkında artan bir endişe bulunmaktadır.

Enerji, modern ekonomilerin tüm sektörlerinin üzerine kurulduğu yapı taşı olarak hizmet ettiği için tüm ekonomik faaliyetlerin temelini oluşturur (Atems ve Hotaling, 2018). Enerji tüketimi ise, ülkelerin ekonomik ve sosyal kalkınmışlıklarının, gelişmişliklerinin ve refah seviyelerinin bir göstergesi olarak kabul edilmektedir. Bu bağlamda, tüketiciler için etkili bir enerji politikası geliştirilmesi açısından, bazı hanelerin neden diğerlerinden daha fazla enerjiye ihtiyaç duyduğunu bilmek önemlidir (Vringer, Aalbers ve Blok, 2007).

Hanehalkları, gelecekteki enerji tasarrufu planlamaları için büyük bir potansiyeli temsil etmektedir. Dolayısıyla, hanehalkı enerji tüketimini azaltmanın uygun yollarını bulmak, enerji verimliliği programlarını etkin bir şekilde planlamak ve yürütmek politika yapıcıların temel hedefleri arasındadır (Salari ve Javid, 2017).

2019 yılında Türkiye'de, toplam enerji tüketiminde en fazla payı; ilk sırada %35,61 ile konut ve hizmet sektörü, ikinci sırada %34,30 ile sanayi sektörü oluşturmaktadır. Bunları %27,69 ile ulaştırma sektörü ve %4,71 ile tarım-hayvancılık sektörü izlemektedir. Enerji dışı tüketimin payı ise %7,08'dir (ETKB, 2021). Enerji ve Tabii Kaynaklar Bakanlığı'nın yayınladığı bu verilerden anlaşıldığı üzere; Türkiye'de hanehalkı enerji tüketimi, toplam enerji tüketimi içinde oldukça yüksek bir orana sahiptir. Bu nedenle,

hanehalkı enerji harcamalarının yeni yaklaşımlar kullanılarak değerlendirilmesi son derece önemlidir.

Bugün, büyük miktarda ve yüksek boyutlu verilerin var olması nedeniyle hanehalkı toplam enerji tüketimini doğru tahmin etmek zor olabilmektedir. Bu nedenle, verilen bir problemi çözmek için, makinenin öğrenme algoritmalarını kullanarak ve problem ortamından elde edilen verileri inceleyerek çözüme ulaşmak hedeflenmektedir. Burada amaç; çok büyük boyutlu verileri analiz etmek ve bu verileri en doğru şekilde değerlendirerek farklı uygulamalar için kullanılabilecek kalıpları ortaya çıkarmaktır. Temel olarak uygulanan makine öğrenmesi teknikleri, veri kalıplarının analizi sonucunda sınıflandırma, tahmin ve regresyon modellerinde kullanılabilir. Bu bağlamda, yapay zeka (Artificial Intelligence -AI), özellikle de makine öğrenmesi (Machine Learning - ML) çağımızın en önemli teknolojisi durumundadır.

Veri kümelerinde ilk bakışta anlamsız görünen gizli bilgileri ortaya çıkarmaya çalışan makine öğrenmesi yöntemleri ile analiz yapmak, hanehalkı toplam enerji tüketim modellerinin anlaşılmasına yardımcı olmaktadır. Uygun makine öğrenmesi teknikleri, enerji tüketimini iyileştirmek ve doğru stratejilere ulaşmak için bilinmeyen kalıpların keşfedilmesini sağlamaktadır (Nazeriye, Haeri, Haghighat ve Panchabikesan, 2021).

Hanehalkı verilerini kullanarak hanehalkı enerji tüketimi profillerini belirlemek için klasik istatistiksel yöntemler ve makine öğrenmesi yöntemlerinden yararlanılmaktadır. Makine öğrenmesi dışındaki istatistiksel yöntemler çeşitli varsayımların bağımlılığı ile sınırlıdır. Değişkenlerin doğasına ilişkin varsayımlar gerektirmeyen, çeşitli uygulamalar için giderek daha fazla kabul gören makine öğrenmesi yöntemleri genellikle klasik yöntemlerden daha iyi genelleme yeteneği sağladığı için tercih edilmektedir.

Hanehalkı enerji tüketim profilini analiz etme süreci, çok sayıda parametrenin tahmin edilmesini içermektedir. Bu durumda, makine öğrenmesi yaklaşımları; veri ön işleme (preprocessing) süreçleri, çapraz geçerlilik (cross validation) yöntemleri ile modeli eğitme ve model karşılaştırma yoluyla tahminleri etkinleştirme avantajına sahiptir (Burnett ve Kiesling, 2022).

Makine öğrenimi araştırmaları son yıllarda önemli ölçüde ilerlemiştir. Ancak tek bir makine öğrenmesi ile yapılan analiz sonuçları bazı veri setleri için istenilen yüksek

performansı sağlayamamaktadır. Topluluk öğrenme yöntemleri ise, tek makine öğrenmesi yapılarından çok daha istikrarlı performans göstermektedir. Topluluk öğrenme yaklaşımları, daha iyi bir sınıflandırma tahmini yapmak için bireysel makine öğrenimi yöntemlerini birleştirmektedir. Bu yaklaşım, makine öğrenimi sonuçlarını iyileştirmektedir ve tek bir modelin tahmininden daha iyi tahminler sağlamaktadır (Gürsaka, 2018).

Çalışmada, hanehalklarının toplam enerji harcama profillerini analiz etmek için TÜİK'in her yıl düzenli olarak yayınladığı Hanehalkı Bütçe Araştırması anketi 2019 veri seti kullanılmıştır. Bu çalışma, hanelerin toplam enerji harcamalarının makine öğrenmesi ve topluluk öğrenme yöntemleri ile karşılaştırmalı olarak incelenmesi neticesinde, önceki çalışmaları bir adım ileriye taşımaktadır. Burada amaç, yapay zeka tabanlı topluluk öğrenme yöntemleri ile daha kesin bir bilgi ortaya koymaktır.

Topluluk öğrenme yöntemleri; makine öğrenmesi ile analizi yapılan bireysel modellerin yanında bireysel modelleri birleştirip daha iyi bir sınıflama performansı elde etmek için kullanılmaktadır. Topluluk temelli karar verme, bireyler olarak bizim için yeni bir şey değildir; bu tür sistemleri günlük hayatımızda çok sık kullanırız. Bazı önemli konularda herhangi bir karar vermemize yardımcı olması için genellikle farklı uzmanların görüşlerini alırız. Büyük bir tıbbi operasyonu kabul etmeden önce birkaç doktora danışmak, bir ürünü satın almadan önce kullanıcı incelemelerini okumak, potansiyel bir iş başvurusunda bulunmadan önce referansları aramak, hatta bu makalenin yayınlanmadan önce akran incelemesi bile topluluğa dayalı karar verme sistemi örnekleridir (Zhang ve Ma, 2012).

Topluluk öğrenme yöntemleri, aynı sorunu çözmek için birden çok yöntemi eğitir. Bir topluluğun genelleme yeteneği, genellikle temel makine öğrenmesi yöntemlerinden çok daha güçlüdür (Murphy, 2012). Sonuç olarak, topluluk öğrenme yöntemleri, verilere en uygun yaklaşımları sağlayabilecek çok esnek yöntemler olduğu için tek modelleme çerçevelerinden çok daha istikrarlı bir performans ortaya koymaktadır.

Çalışmada; “ısınma ve mutfak harcama”, “aydınlatma harcama”, “yakıt harcama” değerlerinin toplanarak hanelerin “toplam enerji harcama” değerlerinin elde edilmesi, bu alana yeni bir bakış açısı sunmaktadır. Bu bakış açısı ile hanehalkı “toplam” enerji

harcamalarını bir bütün olarak ele alan bu çalışma, büyüyen literatüre yeni bir yaklaşım eklemeyi amaçlamaktadır.

Karar vericiler, enerji verimliliği ve enerji kullanım davranışları birbiriyle ilişkili olan hanehalklarının; enerji tüketiminde azalma, enerji verimliliğinde artış ve davranış değişikliklerinin meydana gelmesini hedeflemektedir. Bu bağlamda, bu çalışma ilk olarak; hanehalkı enerji tüketimini kontrol etmeyi ve enerji hizmetlerinin kalitesini iyileştirmeyi amaçlayan işletmelere yol gösterici olacaktır. İkinci olarak: politika yapıcılara, düzenleyici politikalar için daha etkin kararlar almalarında yardımcı olacak ve aynı zamanda literatüre titiz bir katkı sağlayacaktır.

Bu tez çalışmasının birinci bölümünde, tezin sonraki bölümlerinin temelini oluşturan genel bilgiler açıklanmıştır. Bu bölümde yapay zeka ve veri madenciliği konuları hakkında genel bir çerçeve çizilmiştir. Veri bilimi ile ilgili tanımlara, veri tabanlarında bilgi keşfi süreci metodolojilerine, öznitelik seçim yöntemlerine ve çapraz geçerlilik yöntemlerine detaylı bir şekilde birinci bölümde yer verilmiştir. İkinci bölümde, tezin ana konusu olan makine öğrenmesi ve topluluk öğrenme yöntemleri açıklanmıştır. Üçüncü bölümde, modeller üzerinden detaylı bir literatür taraması anlatılmış, araştırmada kullanılan veri setinden, yöntemlerden ve değişkenlerden bahsedilmiştir. Sonrasında, araştırma verilerine crisp-dm metodolojisinin uygulanma aşamaları ayrıntılı bir şekilde sunulmuştur. Ayrıca araştırmanın bulgularına bu bölümde yer verilmiştir ve bulgular Tablolar ile kapsamlı bir şekilde anlatılmıştır.

BİRİNCİ BÖLÜM

GENEL BİLGİLER

1.1. YAPAY ZEKA (ARTIFICIAL INTELLIGENCE- AI)

Yapay zeka; akıllı makineler, özellikle de akıllı bilgisayar programları yapma bilimi ve mühendisliğidir (Mccarthy, 2004). Yapay zeka çalışmalarında, insanın zekası ve davranışları baz alınarak bu bilgileri yapay bir ortama aktarmak amaçlanır. Burada istenilen; insan zekasını anlamak için bilgisayarları kullanmak ve bilgisayarın insan gibi düşünebilen algoritmalar geliştirmesini sağlamaktır. (Çoban, 2018). Yapay zeka, manuel görevleri otomatikleştirmek yerine yüksek hacimli ve bilgisayar kullanımı gerektiren görevleri gerçekleştirir. Bunu güvenilir bir şekilde ve yorulmadan yapar. Bu sayede, veriler aracılığıyla tekrarlayan öğrenmeyi ve keşfi otomatikleştirir.

Yapay zeka çalışmalarında ilk adımların; 1206 yılında El Cezeri'nin yaptığı makineler, 1623 yılında hesap makinesinin icadı, 1672 yılında ikili sayma sisteminin geliştirilmesi olduğu düşünülmektedir. 1900'lü yılların ilk yarısında insan beyni ile ilgili çalışmalar artmış ve bilgisayar biliminde gelişmeler yaşanmıştır. İkinci Dünya Savaşı'ndan sonra, birkaç kişi bu alanda bağımsız olarak çalışmaya başlamıştır. 1947'de akıllı makineler üzerine bir konferans veren İngiliz matematikçi Alan Turing, bu bağımsız çalışanların ilki olmuştur. Alan Turing'in 1950'lerde "Makineler düşünebilir mi?" diye sorması ile birlikte, yapay zeka çalışmaları ivme kazanmıştır. Alan Turing'in 1950 yılında yayınladığı "Bilgisayar Makineleri ve Zeka" adlı makalede, bir makinenin akıllı olduğunu düşünmenin koşulları tartışılmıştır (Mccarthy, 2004). Alan Turing, bu makalede, bir makinenin insan davranışı sergileme yeteneğine sahip olup olmadığını belirlemek için bir test önermiştir. Bu test, ünlü Torino Testi olarak bilinir.

1956 yılına gelindiğinde, Allen Newell ve Herbert A. Simon tarafından "Logic Theorist" adlı ilk yapay zeka programı geliştirilmiştir. Bu program, 52 matematik teoreminin 38'ini doğrulamış ve bunların birçoğu için yeni ve daha zarif ispatlar keşfetmiştir. Aynı yıl, "Yapay Zeka" kelimesi ilk olarak, Amerikalı matematik profesörü John McCarthy tarafından konuyla ilgili Dartmouth'da düzenlenen konferansta

kullanılmıştır. Yapay Zeka'ya yönelik ilgi, bu yıldan sonra hızla artmıştır. 1959 yılında IBM'de çalışan Arthur Samuel, makine öğrenmesi terimini ortaya atmıştır.

1960'lara gelindiğinde endüstriyel alanda robotlar üretilmiş, ilk makine öğrenmesi kitabı yayınlanmış ve John McCarthy, Stanford'da bir Yapay Zeka Laboratuvarı kurmuştur. 1966 yılında, Joseph Weizenbaum tarafından MIT'in yapay zeka laboratuvarlarında ELIZA programı geliştirilmiş, ilk hareketli robot üretilmiş ve yapay zeka konusunda ilk uluslararası konferans düzenlenmiştir. 1980'lerde yapay zeka endüstride kullanılmaya başlanmış ve bu alanda kâr elde edilmiştir. 1997 yılında IBM tarafından üretilen Deep Blue (Derin Mavi) adlı programın, dünya satranç şampiyonu Garry Kasparov'u satranç maçında yenmesi büyük ses getirmiştir (Gülleroğlu, 2021).

2002 yılında elektrikli süpürgelerin kullanılmaya başlanması, yapay zekanın evlere girmesini sağlamış, 2005 yılında Amerikan ordusu otonom robotlara yatırım yapmaya başlamıştır. 2006 yılında Facebook, Google, Twitter, Netflix gibi şirketler yapay zeka kullanmaya başlamıştır. 2012 yılında Andrew Ng, derin sinir ağları algoritmalarını kullanmış ve sinir ağı, bir kedinin ne olduğu konusunda bilgilendirilmeden bir kediye tanımayı öğrenmiştir. Bu durum derin sinir ağları ve sinir ağları alanlarında yeni bir çağın başlangıcı olmuştur.

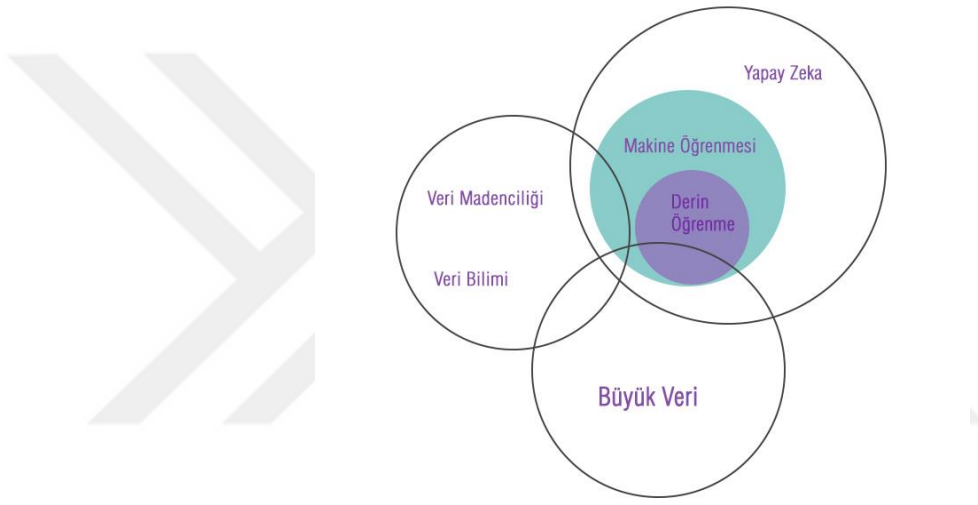
2014 yılında Google, sürüş testini geçen ilk sürücüsüz arabayı yapmıştır. 2016 yılında Hanson Robotics, yüz tanıma, sözlü konuşma ve yüz duygusu yapabilen, insansı bir robot olan ilk "robot vatandaş" Sophia'yı tanıtmıştır. 2000'li yıllarda başlayıp günümüze kadar devam eden robot teknolojilerindeki gelişmeler ve 2017 yılına AlphaGo yazılımının, strateji oyunu olan GO' da dünya şampiyonunu yenmesi, yapay zeka çalışmalarını farklı boyutlara taşımıştır (Çoban, 2018).

2020 yılına gelindiğinde; SARS-CoV-2 pandemisinin ilk aşamalarında Baidu, bir aşı oluşturmak isteyen bilim ve tıp ekiplerinin kullanımına, oluşturduğu yapay zeka algoritmasını sunmuştur. Bu algoritma sayesinde, önceki yöntemlerden 120 kat daha hızlı olan virüsün RNA dizisi sadece 27 saniyede tahmin edilmiştir.

Yapay zeka, genel anlamda; insan zekası gerektiren görevleri yerine getirebilen akıllı makineler oluşturmakla ilgilenen bilgisayar biliminin geniş kapsamlı bir dalıdır. Günümüzde endüstriyel amaçlı kullanımın devam etmesiyle birlikte yapay zeka tek bir

teknolojiyi tanımlamamaktadır. Bir dizi teknoloji ve yöntemi tanımlamak için bir şemsiye terim olarak kullanılmaktadır. Bu yönüyle yapay zeka; veri bilimi, makine öğrenmesi, büyük veri, derin sinir ağları, doğal dil işleme, nesnelerin interneti (IoT), veri madenciliği, sinir ağları gibi kavramlarla birlikte açıklanmaktadır (Zawacki-Richter, Marín, Bond ve Gouverneur, 2019).

Yapay zeka, makine öğrenmesi ve veri bilimi arasındaki ilişki Şekil 1'deki gibi ifade edilebilir.



(Gürsaka, 2018)

Şekil 1. Yapay Zeka, Makine Öğrenmesi ve Veri Bilimi Arasındaki İlişki Diagramı

Yapay zeka algoritmaları, genellikle gerçek zamanlı verileri kullanarak kararlar almak için tasarlanmaktadır. Yalnızca mekanik veya önceden belirlenmiş tepkiler verebilen pasif makinelerden farklılık göstermektedirler. Sensörler, dijital veriler veya uzak girişler kullanarak çeşitli farklı kaynaklardan gelen bilgileri birleştirmekte, materyali anında analiz etmekte ve bu verilerden elde edilen içgörülere göre hareket etmektedirler. Depolama sistemlerindeki, işlem hızlarındaki ve analitik tekniklerdeki muazzam gelişmelerle birlikte, analiz ve karar vermede muazzam bir gelişmişlik yeteneğine sahip olduğu bilinmektedir. Güncel yapay zeka teknolojileri, insan hayatını kolaylaştırmak için birçok farklı disiplinde kullanılmaktadır. Yapay zeka sistemleri,

çoğunlukla şirketler tarafından süreç verimliliklerini iyileştirmek, dolandırıcılık tespit sistemi kurmak, sağlık hizmetinde insan hatalarını en aza indirmek, ağır görevleri otomatikleştirmek, büyük e-ticaret kullanıcılarının ilgilenebilecekleri ürün önerileri yapmak ve içgüdülerden ziyade somut verilere dayalı iş tahminleri yapmak için geliştirilmektedir.

1.2. VERİ MADENCİLİĞİ (DATA MINING - DM)

1.2.1. Veri Bilimi

Veri bilimini açıklamadan önce, veri, bilgi ve özbilgi kavramlarının tanımlarından bahsetmek gerekmektedir. Hesaplamalı sistemlerde kodlanmış değişmezler, veri (data) olarak tanımlanır. İnsan söyleminde ise veri, bilgi verenler tarafından belirtilenlerdir. Veri kelimesi genellikle bilgisayarda kullanılmak üzere kodlanmış kayıtları ifade etmek için kullanılır. Veriler, gözlem yoluyla elde edilen sayısal değerler veya diğer bilgilerin temel bireysel öğeleridir; ama kendi içlerinde bağlamdan yoksundurlar (Zins, 2007).

Bilgi (Information), alıcı için anlamlı olan bir formda işlenmiş verilerdir (Zins, 2007). Bu anlam yararlı olabilir fakat yararlı olması zorunlu değildir. Yararlı olması için işlenen veriler; "kim", "ne", "nerede" ve "ne zaman" sorularına cevaplar sağlar. Bilgisayar dilinde ilişkisel bir veritabanı, içindeki veri depolarından bilgi sağlar. Bilgi, hem bir süreç hem de bir ürün olarak bir farkındalık (bilinç) durumunu temsil eder (Zins, 2007).

Necmi Gürsakal'ın 2001 de Türkçesini özbilgi olarak açıkladığı knowledge kavramı; insan tarafından anlaşılan ve değerlendirilen şeydir. Özbilgi, öngörülmek üzere yeni bağlamlar sağlayan, deneyimle bütünleştirilmiş birikimden elde edilen, anlayış ve farkındalıktır. Özbilgi, bilişsel/duygusal bir anlam ve anlayışta tanımlanan bir durumdur, bu nedenle veri ve bilgiye kıyasla daha öznel ve soyuttur. Özbilgi, "nasıl" ve "neden" sorularına bilgi sağlar (Zins, 2007).

Veri bilimi; görünmeyen kalıpları bulmak, anlamlı bilgiler elde etmek ve iş kararları vermek için modern araçlar ve teknikler kullanarak büyük hacimli verilerle ilgilenen çalışma alanıdır.

Veri bilimi, görünüşte yapılandırılmamış veya bağlantısız verilerdeki kalıpları tespit ederek sonuçların bulunmasına ve tahminlerin yapılmasına izin verir. Yapay zeka, bir bilgisayarın insan gibi davranmasını/düşünmesini sağlar. Veri bilimi, verilerden anlam elde etmek, tahmine dayalı modeller oluşturmak için karmaşık makine öğrenmesi algoritmalarını kullanır. Makine öğrenmesi, bilgisayarlara sağlanan verilerden bir şeyler öğrenmeyi öğreten, veri yöntemleri, bilimsel analiz ve istatistiklerle ilgilenen yapay zekanın bir alt kümesidir.

1.2.2. Veri Tabanlarında Bilgi Keşfi Süreci

1960 yılında Veri Tarama (Data Fishing/ Data Dredging), 1980'lerde Veritabanı Madenciliği (Database Mining) olarak ifade edilen Veri Madenciliği kavramı, 1989 yılında Veri Tabanında Bilgi Keşfi Süreci (Knowledge Discovery in Databases - KDD) olarak tanımlanmıştır.

Veri madenciliği ve veri tabanında bilgi keşfi süreci kavramları birlikte kullanılmaktadır ancak genel görüş, bu iki kavramın ayrı ayrı değerlendirilmesi gerektiğidir. Buradan hareketle 1995'teki Uluslararası bir konferansta veri madenciliği ve veri tabanında bilgi keşfi süreci ilk olarak ayrı ayrı ele alınmıştır. 1997 yılına gelindiğinde Veri Madenciliği ve Bilgi Keşfi dergisi yayınlanmıştır. 2007 yılında Tahmini Analitik (Predictive Analytics) olarak ifade edilen veri madenciliği kavramı 2011 yılına gelindiğinde Veri Bilimi (Data Science) olarak en genel şeklini almıştır.

Veri bilimi, verilerin organizasyonu, özellikleri ve analizi hakkında sistematik bir çalışmaya odaklanır. Veri bilimi teriminin kullanımı, büyük veri kavramı ile birlikte giderek daha yaygın hale gelmektedir. Büyük ve karmaşık veri kümelerini anlama ihtiyacı tüm iş, bilim ve mühendislik alanları için tipik bir problemdir. Yazılım mühendisliği ve geliştirilen yıllar boyunca, uygulamalar oluşturmak için her zaman doğru araçlara ihtiyaç duyulmuştur (Dhar, 2013).

Veri bilimi, genel olarak veritabanları ve bilgisayar bilimi gibi alanlarla güçlü şekilde bağlantılıdır, daha spesifik olarak, makine öğrenmesi ve istatistiklere dayanmaktadır (Kantardzic, 2011). Bu anlamda veri tabanı teknolojisi multidisipliner bir yaklaşıma ihtiyaç duymaktadır.

Veri bilimi; bilgisayar bilimleri, istatistik, yapay zeka, makine öğrenmesi, matematiksel modelleme, örüntü tanıma, yapay sinir ağları, yönetim bilişim sistemleri, verilerin görselleştirilmesi, veritabanı mantığı ve yönetimi, örüntü tanıma, algoritmalar, yüksek bilgi işlem performansı ve uzaysal veri analizi gibi farklı disiplinlerde yer alan tekniklerin bir birleşimini içerir. Bugün dünyada kullanılan teknoloji, büyük miktarlarda veriyi toplama fırsatı sağlar. Sorun, verileri bulmak değil, toplanan verilerin nasıl analiz edileceği ve ondan yararlı bilgilerin nasıl çıkarılacağıdır. Bu süreç genellikle veri madenciliği olarak bilinir.

1.2.3. Veri Madenciliği Kavramı

Veri madenciliği; örüntüler ve faydalı bilgiler elde etmek için büyük miktarda veriyi analiz etme işlemidir (Gullo, 2015). Veri madenciliği, yaygın uygulama alanı nedeniyle araştırmacılar ve uzmanlar için popüler bir araştırma alanıdır. Hemen hemen her sektörde büyük verilerin birikmesi ve tüm iş ortamlarında rekabet gücünün artması ile birlikte, ham verileri işlemek, verileri sınıflandırmak ve kümelemek, hipotez oluşturmak ve kararlar vermek için veri madenciliği önemli bir veri konuşma aracı haline gelmiştir.

Veri madenciliği yeni bir kavram değildir ancak uygulama alanlarının popülaritesi günden güne artmaktadır. Kuruluşların ve şirketlerin verilerini depolarda saklamaya başlamaları sonucunda, kaydedilen verilerin hacmi büyük sayılar haline gelmiş ve geleneksel yaklaşımlar bu devasa verileri işlemek için yeterli olamamıştır. Bu nedenle, veri madenciliği ve uygulamaları büyük veriler ile başa çıkmak ayrıca profesyonellerin rekabet ortamında karar destek sistemleri oluşturması amacıyla stratejik kararlar almalarına yardımcı olmak için geliştirilmiştir.

Veri madenciliği değişkenler arasındaki ilginç, genellikle beklenmedik, önceden bilinmeyen ilişkileri ve yapıları ortaya çıkarmak amacıyla büyük veri setlerini işlemek için tasarlanmıştır. Bu bağlamda, veri tabanlarında, veri ambarlarında veya diğer bilgi havuzlarında depolanan büyük miktardaki verilerden değerli bilgileri keşfetme süreci olarak tanımlanabilir. Bu değerli bilgiler; örüntüler, çağrışımlar, değişiklikler, anomaliler ve önemli yapılar olabilir. Keşif genellikle birden fazla senaryo ve olasılığın denendiği

yinelemeli bir süreçle gerçekleştirilir (Fayyad, Piatetsky-Shapiro ve Smyth, 1996; Frawley, Piatetsky-Shapiro ve Matheus, 1992; Zhao ve Luan, 2006).

Bu süreçte bilgiyi elde etmek için makine öğrenmesi algoritmaları kullanılır. Makine öğrenmesi veriden öğrenmedir ve veri madenciliği ise veriden öğrenme sürecidir. Bu iki kavram birbirinin bir parçasıdır. Kısaca, veri madenciliği görevidir ve makine öğrenmesi araçtır. Bu görevi çözmek için devasa veri kümelerinden faydalı bilgiyi çıkarma ve bu bilgiyi kullanma yeteneği günümüzün rekabetçi dünyasında önemli hale gelmiştir (Kantardzic, 2011).

Veri kümelerinin boyutu büyümeye ve daha karmaşık hale gelmeye devam etmekte; otomatik ve akıllı veri analizine sahip yazılım araçlarına olan ihtiyaç artmaktadır. Bu nedenle, makine öğrenmesine ilgi son yıllarda artış göstermektedir. Makine öğrenmesi geliştirmek karmaşık bir iştir; veri yakalama, veri temizleme ve temizlenmiş bir veri kümesiyle eğitmek için bir model oluşturma dahil olmak üzere birkaç adımdan oluşur. Bu adımlardan biri olan veri tabanlarında bilgi keşfi süreci metodolojileri mevcut kaynaklara bağlı olarak farklı şekillerde ifade edilebilir.

1.2.4. Veri Tabanlarında Bilgi Keşfi Sürecinde Kullanılan Metodolojiler

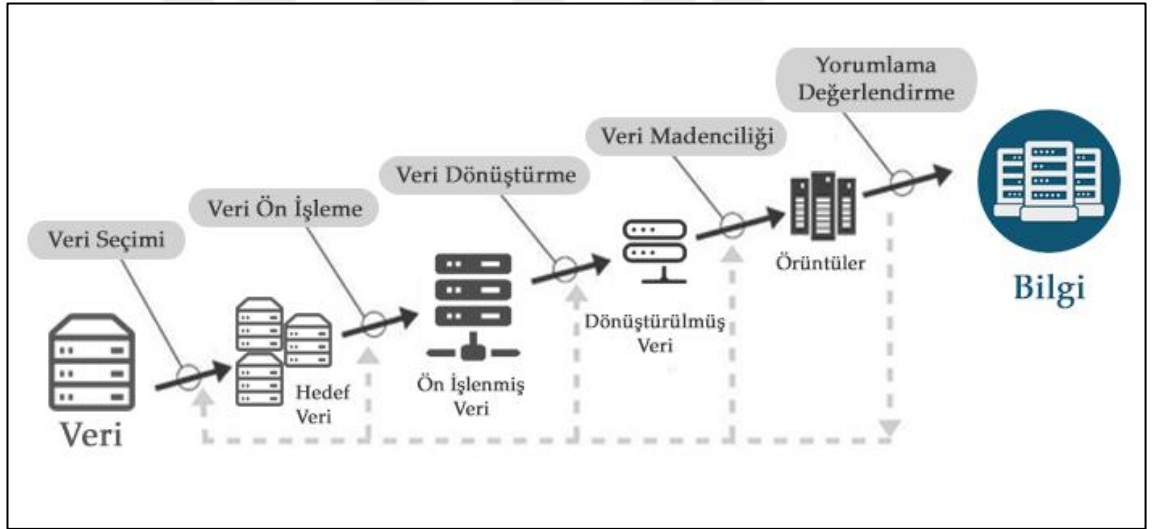
Bilimsel araştırma yapılırken metodoloji önemlidir, projenin amacına bağlı olarak metodolojinin uygun olması gerekir. Metodoloji seçmenin temel amacı, hedefe ve istenen sonuca ulaşmada art arda katkı sağlayan bir sistem bulmaktır (Håkansson, 2013).

Makine öğrenmesi, bilgisayar bilimi içinde karmaşık bir alandır ve makine öğrenmesini doğru ve verimli bir şekilde uygulamak için, geliştiricilere süreç boyunca rehberlik edecek bir metodolojiye ihtiyaç vardır. Bugün, tüm makine öğrenmesi yöntemlerine uygun bir metodoloji yoktur, ancak bazıları diğerlerinden daha yaygın olarak kullanılmaktadır. CRISP-DM, SEMMA ve KDD yaygın olarak kullanılan süreç metodolojileridir. CRISP-DM, SEMMA ve KDD oldukça benzer görünmekle birlikte detaylarda farklılaşan metodolojilerdir.

1.2.4.1. Veri Tabanlarında Bilgi Keşfi (KDD/ Knowledge Discovery in Databases)

Veritabanlarında bilgi keşfi veya kısaca KDD terimi, 1989 yılında verilerde değerli bilgi bulma sürecini ifade etmek için geniş kapsamlı bir kavram olarak kullanılmıştır (Fayyad vd., 1996). KDD, büyük ve karmaşık veri kümelerindeki kritik öneme sahip bilgileri organize etmenin yanı sıra, büyük veri havuzlarının modellenmesini otomatik olarak keşfeden bir analizdir.

KDD süreci, veri madenciliğinde kritik bir görevdir. Bu süreç, önceden bilinmeyen kalıpların keşfi, veri keşif görevine uygun algoritmaların seçimi ve modelin geliştirilmesi süreçlerini içerir. Model, genel olarak verileri tahmin etmek, gruplamak, ilişkilendirmek, görselleştirmek ve analiz etmek için kullanılır (Getahun, 2020). KDD süreci Şekil 2’de gösterilen beş aşamadan oluşmaktadır.



Şekil 2. Bilgi Keşfi Süreci

1.2.4.1.1. Veri Seçimi

İlk aşama, uygulama alanının geliştirilmesi ve anlaşılması aşamasıdır. KDD sürecinde araştırmacının bakış açısıyla hedeflerin tanımlandığı, araştırmanın uygulama alanı ve ön bilgilerin hakkında anlayış geliştirilmesi için kullanıldığı aşamadır. Bu aşama, bir hedef veri seti oluşturmayı veya bir değişken alt kümesine odaklanmayı içerir.

1.2.4.1.2. Veri Ön işleme

İkinci aşama veri ön işlemedir. Bu aşama, tutarlı veri elde etmek için ön işlemlerden geçirilen verinin temizlenmesi sonucunda hedef veri setinin oluşturulmasıdır.

1.2.4.1.3. Veri Dönüştürme

Veri dönüştürme aşaması, üçüncü aşamadır. Bu aşama, boyutsallık azaltma veya dönüştürme yöntemleri kullanılarak verilerin dönüştürülmesinden oluşur.

1.2.4.1.4. Veri Madenciliği

Dördüncü aşama olan veri madenciliğinde, hedefe bağlı olarak belirli bir alandaki ilgi kalıpları aranır.

1.2.4.1.5. Yorumlama/ Değerlendirme

Beşinci aşama olan değerlendirme aşaması, mayınlı kalıpların yorumlanması ve değerlendirilmesinden oluşur (Azevedo ve Santos, 2008).

1.2.4.2. SEMMA

SAS enstitüsü tarafından geliştirilmiştir. SEMMA kısaltması, Sample (Örnek), Explore (Keşfet), Modify (Dönüştür), Model (Modelle), Assess (Değerlendir) kelimelerinin baş harfleridir. SEMMA, bir veri madenciliği projesi yürütme sürecini ifade eden beş aşamalı bir döngüden oluşmaktadır (Azevedo ve Santos, 2008).

1.2.4.2.1. Örnekleme

Bu aşama, büyük bir veri kümesinin bir kısmını çıkararak önemli bilgileri içerecek kadar büyük ancak hızlı bir şekilde manipüle edilebilecek kadar küçük verilerin örneklenmesinden oluşur. Bu aşama, verilerin örneklenmesine odaklanan SEMMA sürecinin ilk ve isteğe bağlı yapılan aşamasıdır.

1.2.4.2.2. Keşfetme

Bu aşama, beklenmeyen eğilimleri ve anomalileri arayarak verilerin araştırılmasından oluşur. Bu, verilerin keşfine odaklanan SEMMA sürecinin ikinci aşamasıdır.

1.2.4.2.3. Dönüştürme

Bu aşama, verilerin oluşturulması, seçilmesi ve model seçim sürecine odaklanmak için değişkenleri dönüştürme işlemlerinin yapıldığı SEMMA'nın üçüncü aşamasıdır.

1.2.4.2.4. Modelleme

Bu aşama, istenilen sonucu güvenilir bir şekilde tahmin eden bir veri kombinasyonu için otomatik olarak yazılımın arama yapmasına izin vererek verilerin modellenmesinden oluşur.

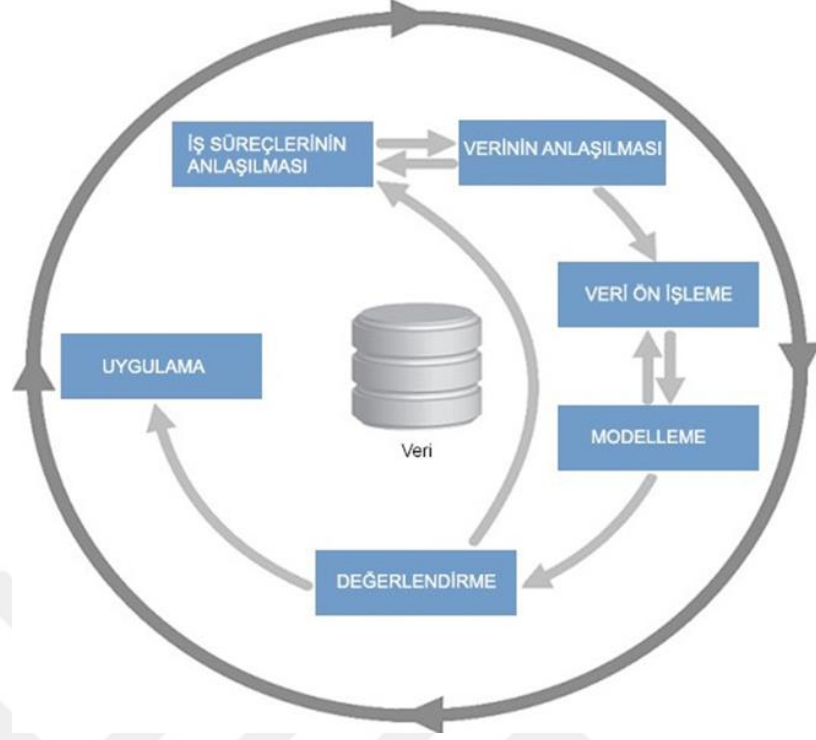
1.2.4.2.5. Değerlendirme

Bu aşama, veri madenciliği sürecinden elde edilen bulguların yararlılığını ve güvenilirliğini ele alarak değerlendirilmesinden oluşur. Ayrıca bu bulguların ne kadar iyi performans gösterdiği bu aşamada tahmin edilir (Azevedo ve Santos, 2008).

1.2.4.3. CRISP-DM (Cross Industry Standard Process for Data Mining /Çapraz Endüstri Standardı Süreci)

Sosyal bilimlerde yaygın olarak kullanılan CRISP-DM metodolojisi 90'lı yılların ortalarında geliştirilmiştir. CRISP-DM; uygulanması kolay, endüstriden, araçtan ve uygulamadan bağımsız, yinelemeli bir yöntemdir.

CRISP-DM ayrıca iyi tanımlanmış aşamalarla yapılandırılmıştır. (Dåderman vd., 2018). Bu sağlam ve kanıtlanmış metodoloji döngüsel olarak tasarlanmış altı aşamadan oluşmaktadır. Aşamalar farklı bir sırayla gerçekleştirilebilir. Ayrıca, analiz sürecinde önceki aşamalara geri dönmek ve belirli eylemleri tekrarlamak gerekebilir (Nazeriye vd., 2021). CRISP-DM metodolojisinin aşamaları ve genel süreç Şekil 3'te gösterilmiştir.



Şekil 3. CRISP-DM Metodolojisi

1.2.4.3.1. İş Süreçlerinin Anlaşılması (Business Understanding)

İlk aşama olan iş süreçlerinin anlaşılması aşamasında; problem tanımı yapılır, konunun amacı belirlenir. Bu aşama temel olarak hedefleri, sorunları ve gereksinimleri ortaya koymaya odaklanır.

1.2.4.3.2. Verinin Anlaşılması (Data Understanding)

Verinin anlaşılması aşaması; başlangıç verilerinin toplanması, verinin tanımlanması, verinin keşfi ve veri kalitesinin doğrulanması maddelerine odaklanan CRISP-DM sürecinin ikinci aşamasıdır (Cios, Pedrycz, Swiniarski ve Kurgan, 2007).

1.2.4.3.3. Veri Ön İşleme (Preprocessing)

Üçüncü aşama olan veri ön işleme aşaması, nihai veri kümesini oluşturmak için gerekli tüm faaliyetleri kapsar (Cios vd., 2007). Bu aşamada, mevcut veri kaynakları belirlendikten sonra verinin istenen formda inşa edilmesi ve biçimlendirilmesi gerekir.

Burada, veri üzerinde yapılacak işlemlere ve bu işlemlerin hangi yöntemlerle yapılacağına karar verilir (Olson ve Delen, 2008). Verilerin hazırlanması sırasında veri kalitesini etkileyen tutarsızlıklar ortadan kaldırılır. Bu aşama, veri temizleme, filtreleme, öznitelik seçimi, aykırı (uç) değerlerin ve anormalliklerin giderilmesi gibi yapıcı veri hazırlığı işlemlerini içerir (Wirth ve Hipp, 2000).

1.2.4.3.4. Modelleme (Modelling)

Modelleme aşaması çeşitli modelleme tekniklerinin uygulandığı dördüncü aşamadır. Bu aşamada farklı parametreler belirlenir ve aynı veri madenciliği problemi için farklı modeller oluşturulur (Shafique ve Qaiser, 2014). Veri hazırlama ve modelleme arasında yakın bir bağlantı vardır. Çoğu zaman, modelleme sırasında problemler yaşanır, bu durumlarda modelleme aşaması, yeni duruma uygun olarak tekrar düzenlenir (Wirth ve Hipp, 2000).

1.2.4.3.5. Değerlendirme (Evaluation)

Değerlendirme aşaması elde edilen modellerin değerlendirilmesine ve sonuçların nasıl kullanılacağına karar verilmesine odaklanır (Shafique ve Qaiser, 2014). Sonucu oluşturmak için yürütülen aşamaların kapsamlı bir şekilde gözden geçirilip değerlendirildiği ve bu aşamanın sonunda ilk aşamadaki hedeflere ne kadar ulaşıldığına ilişkin bir karara varıldığı aşamadır (Dåderman vd., 2018). Verilen karara göre bir sonraki aşamaya geçilebilir veya bütün aşamaların tekrar gözden geçirilmesi için ilk aşamaya geri dönülebilir (Şeker, 2018).

1.2.4.3.6. Uygulama (Deployment)

Son aşama olan uygulama aşaması CRISP-DM sürecinin altıncı ve son aşamasıdır. Elde edilen bilgi ve sonuçların kullanımının belirlenmesine odaklanır. Bu aşama ayrıca gerektiğinde kazanılan bilgileri organize etmeye, raporlamaya ve sunmaya odaklanır (Shafique ve Qaiser, 2014).

CRISP-DM'in tüm aşamaları usulüne uygun olarak organize edilmiş, kolayca anlaşılır veya gözden geçirilir şekilde yapılandırılmış ve tanımlanmıştır. Şekilde gösterilen aşamalar arasında bazı durumlarda döngü oluşur. Bu döngülerden ilki, iş

süreçlerinin anlaşılması ile verinin anlaşılması aşamaları arasındadır. Burada verinin elde edilmesi aşamasında yaşanan zorluklara veya fırsatlara göre araştırmanın tanımlanması aşamasına dönülebilir. Bu durumda, iş süreçlerinin anlaşılması aşaması yeniden gözden geçirilir. Benzer bir döngü de modelleme ile veri ön işleme aşamaları arasında vardır. Bunun sebebi de model seçimine göre bazı veri ön işleme aşamalarının tekrar ele alınması gerektiğidir. Bu durum; bazı modellerin eksik, kategorik veya nümerik veriler ile çalışmamasından kaynaklanır. Son olarak; yöntemin en önemli döngüsü, değerlendirme aşamasındaki sonuçlara göre sistemin en başa dönerek bütün aşamaların üzerinden geçilebilmesi anlamına gelen ana döngüdür (Şeker, 2018).

Günümüzde endüstride kullanılan bu üç metodoloji arasında SEMMA, çoğunlukla veri yönetimine ve veri madenciliğinin model yönlerine odaklanır. SEMMA, CRISP-DM'den farklı olarak, problem tanımını içermez, sorunu iş perspektifinden anlamakla başlar yani endüstrideki çoğu uygulamada değişen problemler ve bu problemler için verinin farklı kullanımları SEMMA'da bulunmaz. SEMMA, verinin farklı amaçlara yönelik olarak bir seferde işlenmesi amacıyla geliştirilmiştir.

KDD aşamalarında ise, ön işleme ve dönüştürme süreçleri ayrılmıştır. CRISP-DM bu iki dönüşümü tek bir aşama olarak ele alır. Bununla birlikte KDD, eksik verilerin tamamlanması, kirli ve gürültülü verilerin çözülmesi gibi aşamaları ön işlemede ele alır (Azevedo ve Santos, 2008).

1.2.5. Öznitelik Seçimi (Feature Selection)

Öznitelik seçimi, verilen örnek kümesi içinde hedef analiz için en uygun öznitelikleri seçen yöntemlerin genel adıdır (Panthong ve Srivihok, 2015). Öznitelik seçimi, yani sınıflandırma veya regresyon problemleri için en uygun öznitelikleri bulma sorusu, ana veri madenciliği görevlerinden biridir. Öznitelik seçiminin odak noktası, girdi verilerini doğru bir şekilde karakterize edebilecek aykırı ve alakasız değişkenlerin etkilerini azaltırken yüksek tahmin sonuçları sağlayan bir değişken alt kümesi bulmaktır.

Öznitelik seçimi öğrenilen sonuçların karmaşıklığını azaltmada, öğrenme verimliliğini ve tahmin doğruluğunu artırmada etkili olmak için hem teoride hem de pratikte kullanılan kanıtlanmış bir uygulamadır. Öznitelik seçimi, istatistik, makine

öğrenmesi, örüntü tanıma gibi pek çok alanda veri bilimi uygulamalarında kullanılmaktadır.

Her alandaki verilerin hacim, hız ve çeşitlilik bakımından muazzam artışı nedeniyle öznitelik seçim sürecinin kaçınılmaz olduğu fark edilmektedir. Öznitelik seçimi sürecinde dikkat edilmesi gereken nokta, verilerin değerinin ve kalitesinin süreçten etkilenmemesini sağlamaktır (Arunadevi ve Nithya, 2016). Literatürde öznitelik seçimi için çeşitli yöntemler kullanılmıştır (Arunadevi ve Nithya, 2016; Blum ve Langley, 1997; Chandrashekar ve Sahin, 2014; Ghasemi, Neysiani ve Nematbakhsh, 2020; Guyon, research ve 2003, 2003; Lee, Schowe ve Sivakumar, 2011).

Bu yöntemler arasından; İleriye Doğru Seçim (Forward Selection), Geriye Doğru Eleme (Backward Elimination), Optimize Edilmiş Seçim –Evrimsel (Optimize Selection-Evolutionary) öznitelik seçim yöntemlerinin detayları aşağıda açıklanmıştır.

1.2.5.1. İleriye Doğru Seçim (Forward Selection)

Bu strateji başlangıçta yalnızca bir öznitelik olan öznitelik alt kümelerini kullanır. Daha sonra ek nitelikler, bir nitelik ekleyerek daha fazla performans kazancı kalmayana kadar buluşsal olarak eklenir. İleri seçim operatörü, boş bir nitelik seçimi ile başlar ve her turda, verilen örnek setine kullanılmayan her öz niteliği ekler. Eklenen her öznitelik için çapraz geçerlilik gibi operatörler kullanılarak performans tahmin edilir. Yalnızca en yüksek performans artışını sağlayan öznitelik, seçime eklenir. Daha sonra değiştirilen seçimle yeni bir tur başlatılır.

Bu uygulama, orijinal olarak verileri depolamak için kullanılan bellek ve operatörlerin uygulanması için gerekli olabilecek bellek dışında herhangi bir ek bellek tüketimini önler. Performansta herhangi bir artış olduğu sürece yineleme çalışır. Artış, belirtilen düzeyde önemli olmadığında yineleme durur (Arunadevi ve Nithya, 2016).

1.2.5.2. Geriye Doğru Eleme (Backward Elimination)

İleriye doğru seçim stratejisinin aksine geriye doğru eleme stratejisi, tüm öznitelik kümesini ilk alt küme olarak başlatır. Her turda verilen örnek setinden kalan her öz niteliği kaldırır. Kaldırılan her öznitelik için çapraz geçerlilik gibi operatörler kullanılarak

performans tahmin edilir. Yalnızca performansta en az düşüşe neden olan nitelik seçimden kaldırılır. Ardından değiştirilen seçimle yeni bir tur başlatılır.

Bu uygulama, orijinal olarak verileri depolamak için kullanılan bellek ve operatörlerin uygulanması için gerekli olabilecek bellek dışında herhangi bir ek bellek tüketimini önler. Yineleme, performansta herhangi bir artış olduğu sürece çalışır. Azalma belirtilen düzeyden az olduğu sürece yineleme çalışır. Azalma, belirtilen düzeyde önemli bir eşikte olduğu anda yineleme durur (Arunadevi ve Nithya, 216).

1.2.5.3. Evrimsel Strateji ile Optimize Etme (Optimize Selection - Evolutionary)

Bir optimize edilmiş öznelik alt kümesi, evrimsel bir strateji tarafından da bulunabilir. Bu nedenle, her öznelik alt kümesi bir birey olarak kabul edilir. Evrimsel bir algoritma, bu tür bireylerin bir popülasyonu üzerinde çalışır. Evrimsel bir öznelik seçim stratejisi, genetik algoritma kullanılarak uygulanır. Genetik algoritmalar, optimizasyona çözümler üreten daha geniş evrimsel algoritmalar sınıfına aittir (Arunadevi ve Nithya, 2016).

Genetik algoritma, doğal evrim sürecini taklit eden bir arama buluşsal yöntemdir. Bu buluşsal yöntem, optimizasyon ve arama problemlerine faydalı çözümler üretmek için rutin olarak kullanılır. Genetik algoritmalar problemlere tek bir çözüm üretmek yerine farklı çözümlerden oluşan bir çözüm kümesi üretir. Genetik algoritmalar, kalıtım, mutasyon, seçim ve çaprazlama gibi doğal evrimden ilham alan teknikleri kullanarak optimizasyon problemlerine çözümler üreten daha geniş evrimsel algoritmalar sınıfına ait olarak çalışır (Arunadevi ve Nithya, 2016).

1.2.6. Çapraz Geçerlilik (Cross-Validation) Yöntemleri

Çapraz geçerlilik yöntemi, ilk olarak 1948 yılında Kurtz tarafından sunulmuştur. Modelin doğruluğu test etmek için kullanılan çapraz geçerlilik yöntemlerinde veri seti başlangıçta test ve eğitim verisi olarak tesadüfi olarak ayrılır. Eğitim verisi modelin eğitimi aşamasında kullanılır. Eğitim verisinden öğrenilen model test verisine uygulanır. Test verisi model kurulumunda kullanılmaz, modelin doğruluğunu test etmek için kullanılır. Basit çapraz geçerlilik yönteminde; eğitim ve test verisi genellikle (0.90 – 0.10), (0.85 – 0.15), (0.80 – 0.20), (0.75 – 0.25) ve (0.70 – 0.30) oranlarında

ayrılmaktadır. Bu oranlardan büyük olan oran eğitim verisi ve küçük olan oran test verisi olarak kullanılmaktadır.

1951 yılında Moiser bu yöntemi geliştirerek çift çapraz geçerlilik yöntemini ileri sürmüştür. Çift çapraz geçerlilik yöntemi, veri setinin sınırlı sayıda olması durumunda kullanılmaktadır. Bu yöntem, modelleme aşamasında bütün veri setinin kullanımına imkân sağlamaktadır. Veri seti, bir bölümü eğitim verisi diğer bölümü test verisi olmak üzere tesadüfî olarak iki eşit parçaya ayrılır ve bu şekilde bir doğruluk hesaplaması yapılır. Ardından test verisi ile eğitim verisi yer değiştirilerek yeniden bir doğruluk hesaplaması yapılır ve doğruluk değerlerinin ortalaması alınarak modelin doğruluk oranı hesaplanır.

1982 yılında da Krus ve Fuller tarafından çoklu çapraz geçerlilik teknikleri geliştirilmiştir. Bu yöntemler genel olarak k-katlı çapraz geçerlilik yöntemleri olarak ifade edilmektedir. K-katlı çapraz geçerlilik yöntemi, model seçiminde en sık kullanılan veri bölme yöntemidir. Bu yöntemde veri k farklı parçaya ayrılır. Bu eşit parçalardan (k -1) tanesi eğitim verisi, bir tanesi ise test verisi olarak kullanılır. Model, (k -1) parça üzerinde eğitilir ve test verisi için tahmin performansı kaydedilir. Bu işlem k kez tekrarlanır, böylece her parça bir kez test verisi olarak kullanılır. Kaydedilen tahmin performanslarının ortalaması alındıktan sonra optimal model parametresi en iyi ortalama tahmin performansına sahip olan olarak belirlenir (Temel, Erdoğan ve Ankaralı, 2012; Xu ve Goodacre, 2018).

İKİNCİ BÖLÜM

MAKİNE ÖĞRENMESİ ve TOPLULUK ÖĞRENME YÖNTEMLERİ

2.1. MAKİNE ÖĞRENMESİ (MACHINE LEARNING - ML)

Makine öğrenmesi, ilk olarak 1959 yılında Arthur Samuel tarafından ortaya atılmış, anlamlı bilgilerin otomatik olarak algılanmasını ifade eden bir kavramdır. Arthur Samuel'e göre makine öğrenmesi; bilgisayarlara açıkça programlanmadan öğrenme yeteneği veren çalışma alanı şekilde tanımlanır (Mahesh, 2018). Genel olarak makine öğrenmesi; bilgisayarlara sağlanan verilerden bir şeyler öğrenmeyi öğreten yapay zekanın bir alt alanı olarak tanımlanmaktadır. Makine öğrenmesi, bir makineye geçmiş deneyimlere dayanarak nasıl çıkarımlarda bulunacağını ve karar vereceğini öğretir. İnsan deneyimini dahil etmek zorunda kalmadan olası bir sonuca varmaya çalışır. Veri noktalarından anlam çıkarmak için kalıpları tanımlar ve geçmiş verileri analiz eder. Bu anlamda makine öğrenmesi algoritmaları; kalıpları tanımlar, tahminlerde bulunur ve yeni keşfedilen kalıpları ilk tasarımlarında yer almayan veya kapsanmayan durumlara uygulayarak yazılımları içeren bir sistem oluşturur (Popenici ve Kerr, 2017).

Makine öğrenmesi, bilgisayar bilimi ve istatistiğin kesişiminin doğal bir sonucu olarak ortaya çıkmıştır. Aşağıda yer alan birbiriyle ilişkili iki soru üzerine odaklanan bir disiplindir.

- Deneyim yoluyla otomatik olarak gelişen bilgisayar sistemleri nasıl inşa edilebilir/ nasıl kurulabilir?
- Bilgisayarlar, insanlar, organizasyonlar dahil tüm öğrenme sistemlerini yöneten temel istatistik hesaplama bilgi teorik yasaları nelerdir?

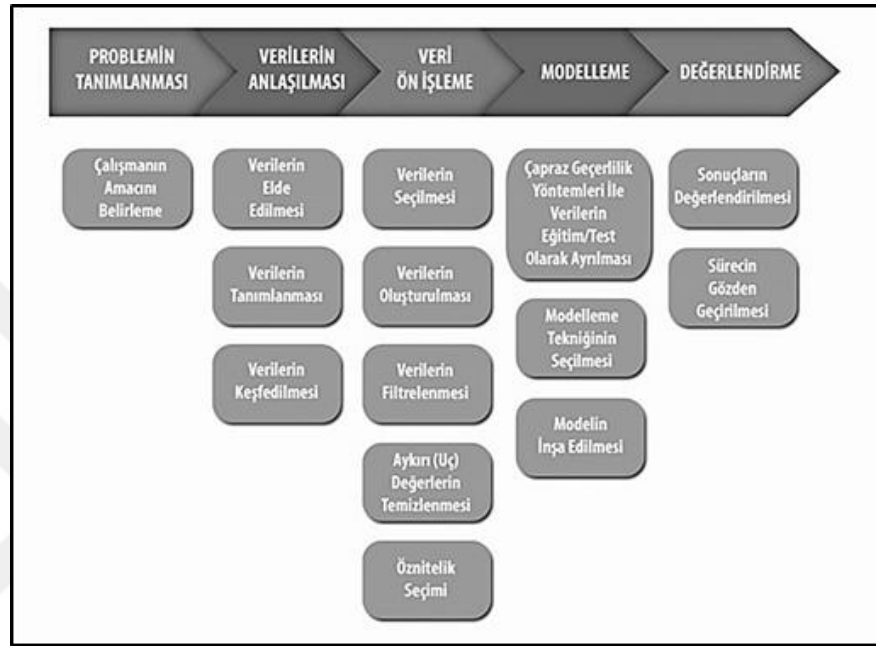
Bu soruların cevabı için, makine öğrenmesi ile istatistik arasındaki farkın ortaya konulması gerekmektedir. İstatistik öncelikle, verilerden hangi sonuçların çıkartılabileceğine odaklanır. Makine öğrenmesi ise bu veriyi en etkili şekilde ele geçirmek, depolamak, endekslemek, yeniden düzenlemek, birleştirmek için hangi hesaplama mimarilerinin ve algoritmalarının kullanılabileceği, birden fazla öğrenme alt görevinin büyük sistemlerde nasıl düzenleneceği hakkında ek soruları da içerir (Mitchell, 2006).

Makine öğrenmesi, veriyi eğiterek zekaya dayalı yapay bilgi üretimi sağlar. Bu yapay akıllı sistem, bilinen (eğitim) verilerden öğrenir ve edindiği bilgileri bilinmeyen (test) verilere uygular. Makine öğrenmesinin birincil amacı, yeni ve görünmeyen veriler üzerinde mümkün olduğunca doğru bir tahmin yapmak için bir eğitim veri setinden öğrenmektir (Dåderman ve Skolan, 2018). Sonuçta öğrenme, deneyimi uzmanlığa dönüştürme veya karmaşık duyuşal verilerdeki anlamlı kalıpları tespit etme yeteneği olduğundan, insan zekasının temel taşıdır. Ancak, geleneksel yapay zekanın aksine, makine öğrenmesi akıllı davranışın otomatik olarak taklidi değildir. Makine öğrenmesi algoritmaları, bilgisayarların güçlü ve özel yeteneklerini insan zekasını tamamlamak için kullanan ve genellikle insan yeteneklerinin çok ötesine geçen görevleri yerine getiren algoritmalar (Shalev-Shwartz ve Ben-David, 2014).

Makinenin öğrenip öğrenmediğini anlamak için en kritik faktör eğitim ve test sürecidir. Bu süreçte veri eğitim ve test olmak üzere parçalara ayrılır. Eğitim verisi bir model oluşturmak için kullanılırken test verisi bu modelin geçerliliğini test etmek için kullanılır. Etkili bir eğitim süreci, geliştirilen sistemin kalitesini artırır. Bu anlamda makine öğrenmesi; verileri anlamlandırmak, karar vermek ve tahmin etmek için kullanılan yöntemlerin genel adıdır. Makine öğrenmesi modeli eğitim verilerine göre eğitildikten sonra, eğitim verileri kullanılarak test edilmektedir. Bunun amacı veriden ne kadar öğrenildiğini belirlemektir (Uçar, Nour, Sindi ve Polat, 2020).

Makine öğrenmesi beş ana adıma ayrılır. İlk adımda; araştırma problemi tanımlanır. Çalışmanın amacı ve kapsamı belirlenir. İkinci adımda; veri toplanır, tanımlanır ve anlaşılır. Makineler başlangıçta onlara verilen verilerden öğrenir. Bu nedenle; makine öğrenmesinde, modelin doğru kalıpları bulabilmesi için güvenilir veriler toplamak son derece önemlidir. Makineye verilen verilerin kalitesi, modelin ne kadar doğru olduğunu belirler. Yanlış veya güncel olmayan veriler ile konuyla ilgili olmayan yanlış sonuçlara veya tahminlere sahip olunur. Üçüncü aşamada; verilerin hazırlanır. Bu adım veri ön işleme yöntemlerinden oluşur. İstenmeyen veriler, eksik değerler, veri dönüştürme, öznitelik seçimi v.b ön işleme yöntemleri kullanılır. Bu adımda veri kümesini çalışmanın amacına uygun olarak yeniden yapılandırmak gerekir. Dördüncü adım modelin seçildiği ve inşa edildiği adımdır. Bu adımda veri eğitim ve test verisi olarak ikiye ayrılır. Eğitim, makine öğreniminde en önemli adımdır. Eğitimde, kalıpları bulmak ve tahminler yapmak için hazırlanan veriler, makine öğrenmesi modeline iletilir. Model eğitildikten sonra, nasıl

performans gösterdiğini kontrol etmek gerekir. Bu, bilinmeyen veriler üzerinde modelin performansını test ederek yapılır. Beşinci adımda; test verileri, modelin nasıl performans gösterdiğine ve hızına ilişkin doğru bir ölçüm verir. Neticede, doğru tahminler yapmak için, model bilinmeyen veriler üzerinde kullanılır ve sonuçlar değerlendirilir. Şekil 4'te makine öğrenmesi süreci ifade edilmiştir.



Şekil 4. Makine Öğrenmesi Süreci

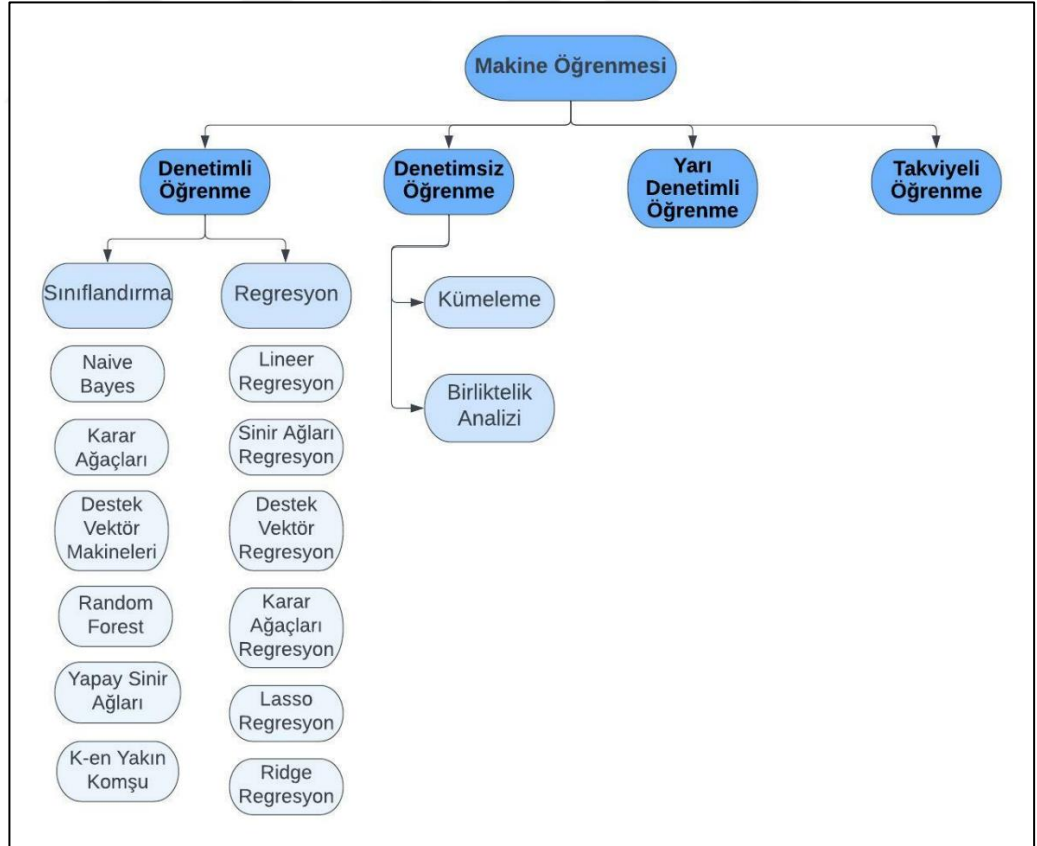
Makine öğrenmesi, gelecekteki bilinmeyen sonuçları tahmin etmek için daha çok girdi verileri ve algoritmalar kullanmaktadır. Günümüzde, birçok yapay zeka sistemi geliştiricisi, istenen giriş-çıkış davranışı örneklerini göstererek bir sistemi eğitmenin pek çok uygulama için oldukça kolay olduğunu kabul etmiştir. Makine öğrenmesi çalışmaları hem temel bilimsel mühendislik soruları için hem de son derece pratik bilgisayar yazılımları için önemlidir. Makine öğrenmesinin kapsamı, tahmine dayalı analizin kalitesini, daha hızlı karar vermeyi, daha sağlam, bilişsel analizleri, robotların yükselişini ve iyileştirilmiş tıbbi hizmetleri geliştirmektir. Makine öğrenmesinin uygulama alanları genel olarak; veri analizi, optimizasyon analizi, yapısal analiz, performans tahmini, pratik yazılım geliştirme, bilgisayarla görme, konuşma, tanıma ve doğal dil işlemedir (Liu vd., 2019).

Makine öğrenmesinin etkisi; bilgisayar bilimi, yoğun veri ile ilgili sektörler ile birlikte tüketici hizmetleri, karmaşık sistemlerde arıza tespiti ve lojistiğin kontrolü ile

ilgili endüstrilerde de geniş çapta hissedilmektedir. Makine öğrenmesi, deneysel verilerin yüksek çıktılarını yeni yollarla analiz etmek için geliştirildiğinden benzer şekilde; biyolojiden kozmolojiye, sosyal bilimlerden tıp bilimine kadar deneysel bilimlerde yaygın bir etkiye sahiptir (Mitchell, 2006).

2.1.1. Makine Öğrenmesi Türleri

Makine öğrenmesi, veri sorunlarını çözmek için farklı algoritmalara dayanır. Makine öğrenmesi türleri, istenen sonuca bağlı olarak düzenlenen algoritmalarlardır. Makine öğrenmesi, genellikle bir algoritmanın tahminlerinde daha doğru olmayı nasıl öğrendiğine göre sınıflandırılır. Bu sınıflandırma için dört temel yaklaşım vardır. Bunlar; denetimli öğrenme, denetimsiz öğrenme, yarı denetimli öğrenme ve pekiştirmeli öğrenmedir. Bilim adamlarının kullanmayı seçtikleri algoritma türü, tahmin etmek istedikleri veri türüne bağlıdır. Makine öğrenmesi türleri Şekil 5'te ifade edilmiştir.



Şekil 5. Makine Öğrenmesi Türleri

2.1.1.1. Denetimli Öğrenme (Supervised Learning)

Denetimli öğrenme, girdi-çıkı örneğine dayalı olarak bir girdiyi bir çıktıya eşleyen bir fonksiyon öğrenme görevidir (Mahesh, 2020). Denetimli öğrenmede, makineye örnek olarak girdiler ve onların çıktıları verilmektedir. Bu durum literatürde etiketlendirme olarak adlandırılmaktadır. Burada amaç girdiler ile çıktılar arasındaki ilişkiyi öğrenen genel bir kural bulmaktır (Gürsakar, 2018). Başka bir deyişle, modelde sonuçları bilinen eğitim verileri kullanılarak bir fonksiyon üretilmesi amaçlanmaktadır. Bu şekilde, etiketli verilerden oluşturulan modele dayanarak veri kümesinde etiket bilgisi olmayan verilerin sonuçlarının tahmin edilmesi hedeflenmektedir (Eraslan, 2019). Denetimli öğrenme algoritmaları, etiketlenmiş eğitim verilerini analiz ederek bir fonksiyon oluşturmaktadır ve daha sonra bu fonksiyon etiketlenmemiş olan yeni verilerin analizinde kullanılmaktadır (Gürsakar, 2018). Denetimli öğrenmenin görevi, sınıf etiketlerinin dağılımına ilişkin özlü bir model oluşturmaktır. Denetimli öğrenme algoritmaları; verileri iki kategoriye ayırarak ikili sınıflandırma, ikiden fazla hedef değişkene ayırarak çoklu sınıflandırma görevlerini yürütmektedir. Ayrıca, regresyon modelleri ile tahmin üretmek ve daha doğru bir tahmin üretmek için birden çok makine öğrenmesi modelinin tahminlerini birleştirmek için iyi düzenlenmiş sistemlerdir.

Denetimli öğrenme yapısına sahip birçok makine öğrenmesi yöntemi vardır. Bunları genel olarak iki başlık altında toplayabiliriz. Bunlar sınıflandırma ve regresyon yöntemleridir. Karar ağaçları, k-en yakın komşu, naive bayes, destek vektör makineleri, sinir ağları, denetimli öğrenme alanında en sık kullanılan makine öğrenmesi yöntemleridir (Gürsakar, 2018).

2.1.1.2. Denetimsiz Öğrenme (Unsupervised Learning)

Denetimsiz öğrenme, etiketlenmemiş veriler üzerindeki bilinmeyen bir yapıyı tahmin etmek için işlemler uygulayan bir makine öğrenmesi tekniğidir. Denetimsiz öğrenmede öğrenme algoritmasına etiketler verilmez. Algoritma kendi başına girdilerin yapısını belirler. Etiketler olmadığı için de bu algoritmalar çıktıların doğruluğunu değerlendiremez (Gürsakar, 2018). Bu teknikte, girdilerin ait olduğu sınıflar belirsiz olduğundan veri kümesindeki bileşenlere göre gizli ilişkileri veya grupları ortaya koymak amaçlanmaktadır.

Denetimsiz öğrenme, verilerdeki gizli kalıpları veya içsel yapıları bulur. Etiketli yanıtlar olmadan girdi verilerinden oluşan veri kümelerinden çıkarımlar yapmak için kullanılır. Bu bölümdeki algoritmalar, verileri gruplayarak yeni verileri en uygun gruba atamaya çalışır.

Denetimsiz öğrenmede eğitim verisi yoktur ve sonuçlar bilinmez. Temelde makine öğrenmesinde denetimsiz öğrenme, makinenin problemi adeta gözü kapalı çözmeye çalışmasına benzemektedir. Makine, bu makine öğrenmesi türünde problemi sadece girdi verileri ile çözmeye çalışmaktadır (Gürsakal, 2018). Eğitim verisi olmadığı için uygulanması nispeten kolaydır. Ancak, zor sorunlarda doğru sonuçlar elde edilemeyebilir (Eraslan, 2019).

Denetimsiz öğrenme algoritmaları; veri kümesini benzerliğe göre gruplara ayırarak kümeleme analizi, bir veri setindeki olağandışı veri noktalarını belirleyerek anormallik tespiti yapabilmektedir. Ayrıca, bir veri setinde sıklıkla birlikte meydana gelen öğe setlerinin belirlenmesi yoluyla birliktelik analizi ve bir veri kümesindeki değişkenlerin sayısını azaltarak boyut azaltma görevlerini gerçekleştirir.

2.1.1.3. Yarı Denetimli Öğrenme (Semi-Supervised Learning)

Yarı denetimli öğrenme; denetimli ve denetimsiz öğrenme yöntemlerinin karışımıdır. Yarı denetimli öğrenme, az miktarda etiketlenmiş eğitim verisi ile oluşturulan bir algoritma ile çalışır. Algoritma daha sonra, yeni, etiketlenmemiş veri kümesinin boyutlarını öğrenir. Algoritmaların performansı, etiketli veri kümeleri üzerinde eğitildiklerinde genellikle iyileşir. Ancak verileri etiketlemek zaman alıcı ve pahalı olabilir.

Yarı denetimli öğrenme, denetimli öğrenmenin performansı ile denetimsiz öğrenmenin verimliliği arasında bir orta yol bulur. Birçok uygulamada önemli insan emeği gerektirdiği için etiketleme maliyeti yüksek olur. Bu tür durumlarda eğer bazı örnekler için etiketleme yapılmışsa, yarı denetimli öğrenme uygulanır (Gürsakal, 2018). Yarı denetimli öğrenme algoritmaları; kelimelerin sözlükteki tam karşılığından daha azına dayalı olarak dili çevirmek için algoritmaları makineye öğretmek için kullanılır. Ayrıca, yalnızca birkaç olumlu örnek olduğunda dolandırıcılık vakalarını belirlemek ve

küçük veri kümeleri üzerinde eğitilen algoritmaların daha büyük kümelere otomatik olarak uygulanmasını sağlayarak veri etiketlemek için kullanılır.

2.1.1.4. Takviyeli/ Pekiştirmeli Öğrenme (Reinforcement Learning)

Takviyeli öğrenme, bir aracının kendi eylemlerinden ve deneyimlerinden gelen geri bildirimleri kullanarak deneme yanılma yoluyla etkileşimli bir ortamda öğrenmesini sağlayan bir tür makine öğrenmesi tekniğidir.

Denetimsiz öğrenmeye kıyasla, takviyeli öğrenme hedefler açısından farklıdır. Denetimsiz öğrenmede amaç, veri noktaları arasındaki benzerlikleri ve farklılıkları bulmak iken, takviyeli öğrenme, farklı bir hedefi olan bir algoritma ve bu hedefe ulaşmak için önceden belirlenmiş bir dizi kural programlayarak çalışır. Algoritma; nihai amaca yönelik faydalı bir eylem gerçekleştirdiğinde aldığı olumlu ödülleri aramaya ve nihai hedefinden daha uzağa götüren bir eylem gerçekleştirdiğinde aldığı cezalardan kaçınmaya programlanır. Takviyeli öğrenme algoritmaları genellikle, robotik teknolojiler ve video oyunları alanlarında kullanılır. Ayrıca, işletmelerin kaynakları nasıl tahsis edeceklerini planlamasına yardımcı olan kaynak yönetimi görevlerini yürütür.

Takviyeli öğrenme, çevre ile etkileşime dayalı amaca yönelik bir öğrenme yöntemidir ve bu yöntemde bilgisayarda bir ajan (gerçek hayatta öğrenecek kişi) ödülleri maksimize edecek şekilde, deneme- yanılma ile hareket eder. Ajanın bilgisayardaki her doğru ve yanlış hareketinde bilgisayar bir değer hesaplar ve bu değerleri büyük bir Tabloda saklar. Ajan öğrendikçe, bu değerler sürekli olarak güncellenir ve süreç sürekli olarak gelecek olasılıkların son değerlerden hesaplandığı bir süreç olarak düşünülebilir. Bu yöntemin bilgisayarın görmesi ile ilişkilendirilmiş şekline ise “derin takviyeli öğrenme” adı verilmektedir (Gürsaka, 2018).

2.2. SINIFLANDIRMA (CLASSIFICATION)

Makine öğrenmesinde sınıflandırma, bir dizi etiketli veriye dayalı olarak sınıf etiketlerini tahmin eden, denetimli öğrenmede kullanılan tahmine dayalı bir modelleme yaklaşımıdır. Sınıflandırma, önceden belirlenmiş hedefler ile eğitim veri setlerinden elde edilen sonuçlardan yararlanarak test veri setleri için uygun bir sınıflandırıcı bulma işleminin genel adıdır. Sınıflandırmada amaç, önceden belirlenen niteliklere göre

girdilerin sınıflara atanması işlemidir. Bu işleme göre eldeki verilerin bir sınıfa dahil olup olamayacağı ya da hangi sınıfa dahil olacağı belirlenir. Sınıflandırmada girdi; gözlem ve örneklerden oluşmuş eğitim setidir. Çıktı ise modelin gözlem sonucunda belirlenmiş olduğu sınıf etiketidir.

Sınıflandırma algoritması iki aşamadan oluşur. Bu aşamalardan ilki eğitim aşaması ikincisi test aşamasıdır. Eğitim aşaması (training), eğitim kümesinin kullanılarak modelin inşa edilmesi yani modelin öğrenmesidir. Test (testing) aşaması ise eğitim aşamasında elde edilen modelin tahmin için kullanılmasıdır. Test aşaması modelin doğruluğunun kontrol edilmesine karşılık gelir. Modelin doğruluğunu kontrol etmek için, test örneklerinden sınıfı iyi bilinen model, tahmin edilen sonuç ile karşılaştırılır. Girdilere göre çıktıları üreten model, daha sonra sınıf etiketi bilinmeyen ya da kayıp olan örneklerin sınıfının tahmin edilmesinde kullanılır (Taşkın, 2005).

Sınıflandırma, makine öğrenmesi yöntemleri arasında en çok kullanılan yöntemlerden biridir. Sınıflandırma işlemi; tüketici profili belirleme, tıbbi teşhiste bulunma, sahtekarlık tespiti, kredi onaylanması, sigorta miktarının hesaplanması, pazarlama, web sayfası kategorizasyonu çalışmaları gibi birçok alanda kullanılmaktadır.

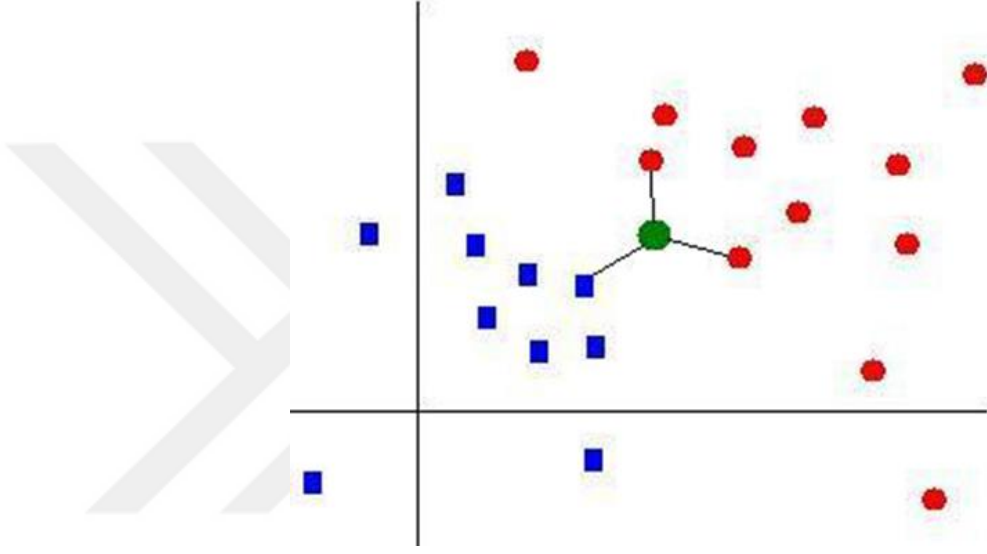
Denetimli makine öğrenmesinde, sınıflandırma algoritmaları çeşitlilik içerir. Karar ağaçları, KNN, lojistik regreyon, naive bayes, random forest, destek vektör makineleri, yapay sinir ağları sınıflandırma için yaygın olarak kullanılan yöntemlerdir. Bu çalışma; KNN, karar ağaçları, gradyan artırmalı ağaç, random forest, naive bayes, ve derin sinir ağları yöntemlerine odaklanmıştır.

2.2.1. K-en yakın komşu (K-Nearest Neighbor – KNN)

K-en yakın komşu yöntemi, Cover ve Hart (1967) tarafından sınıflama ve regresyon problemlerine çözüm olarak öne sürülmüştür. Amaç; girdi ve çıktı değişkenleri arasındaki ilişkiyi modelleyen, KNN regresyon fonksiyonu olarak da bilinen fonksiyonu tahmin etmektir.

KNN algoritması, eğitim veri seti olarak tüm veriyi kullanır. KNN algoritması eğitim aşamasında sadece veri setini saklar ve yeni veri aldığında bu veriyi yeni veriye çok benzeyen bir kategoride sınıflandırır. Yani, tüm veriler bir örüntü uzayında saklanır.

Algoritma, bilinmeyen bir verinin hangi sınıfa dahil olduğunu belirlemek için örüntü uzayını araştırarak bilinmeyen örnekleme en yakın olan en yakın komşuyu bulur. Yakınlık öklid uzaklığı ile tanımlanır. Daha sonra, bilinmeyen veri, k en yakın komşu içinden en çok benzediği sınıfa atanır. Sonra eğitim veri setinin k-en yakın komşuluk verileri elde edilir. Bu verilere dayanarak her yeni veri için tahmin yapılır (Ayık vd., 2007; Poloczek, Treiber ve Kramer, 2014). KNN algoritması Şekil 6’da gösterilmiştir.



(Şeker, 2008)

Şekil 6. KNN Algoritması

KNN, temel veriler üzerinde herhangi bir varsayımda bulunmaz. Bu nedenle, tembel öğrenen algoritması olarak da adlandırılır çünkü eğitim kümesinden hemen öğrenmez, bunun yerine veri kümesini depolar ve sınıflandırma anında veri kümesi üzerinde bir işlem gerçekleştirir. KNN algoritması eğitim aşamasında sadece veri setini saklar ve yeni veri aldığı anda bu veriyi yeni veriye çok benzeyen bir kategoride sınıflandırır.

KNN, denetimli öğrenme tekniğine dayalı en basit makine öğrenmesi algoritmalarından biridir. Bununla birlikte; uygulanması basit, gürültülü eğitim verilerine karşı dayanıklı ve eğitim verilerinin büyük olması halinde daha etkili olduğundan bu teknik birçok uygulama için tercih edilir.

2.2.2. Naive Bayes

Naive Bayes, İngiliz matematikçi Thomas Bayes tarafından 1763 yılında geliştirilen Bayes Teoremini temel alan denetimli bir makine öğrenme algoritmasıdır. Naive Bayes algoritması, Naive ve Bayes olmak üzere iki kelimeden oluşur. İlk kelime, belirli bir özelliğin ortaya çıkmasının diğer özelliklerin ortaya çıkmasından bağımsız olduğunu varsaydığı için Naive olarak adlandırılır. İkinci kelime, Bayes Teoremi ilkesine bağlı olduğu için Bayes olarak adlandırılır.

Bayes teoremi, sınıf değişkeninin değeri göz önüne alındığında tüm değişkenlerin bağımsız olduğunu varsayan bir teoremdir. Naive Bayes ise Bayes kuralını, niteliklerin sınıfa göre koşullu olarak bağımsız olduğu güçlü bir varsayım ile birlikte kullanan basit bir öğrenme algoritmasıdır. Bayes kuralını güçlü bir öğrenme yöntemiyle birlikte kullanan sınıflandırıcı, varsayımların ve parametre tahminlerinin öğrenme sürecini nasıl basitleştirdiğinin klasik bir gösterimidir. Naive Bayes algoritması, belirli bir veri kümesindeki değerlerin kombinasyonlarının ve frekanslarının olasılığını hesaplayan basit bir olasılık sınıflandırıcısıdır (Shalev-Shwartz ve Ben-David, 2014).

Naive Bayes algoritması, Bayesyen olasılık ve istatistik yöntemleri aracılığıyla regresyon ve sınıflandırma problemlerinin çözümünde kullanılır. Olasılıksal bir sınıflandırıcıdır, yani bir nesnenin olasılığı temelinde tahmin yapar. Naive Bayes kullanmanın avantajı, sınıflandırma sürecinde ihtiyaç duyulan tahminleri belirlemek için yalnızca küçük bir miktar eğitim verisinin yeterli olmasıdır. Naive Bayes hem eğitim hem de sınıflandırma aşamalarında az depolama alanı gerektirdiğinden, çoğu karmaşık durumda genellikle beklenenden daha iyi performans gösterir (Jet ve O, 2017).

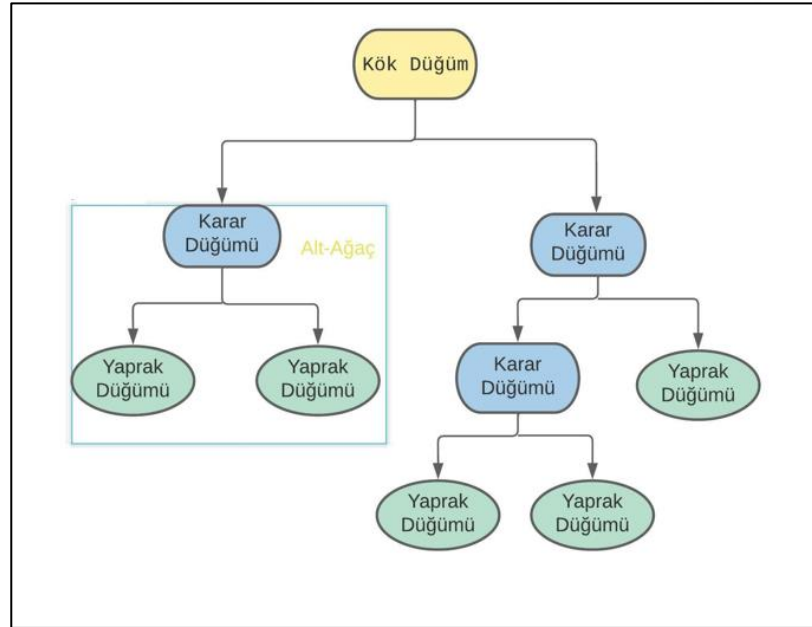
2.2.3. Karar Ağacı (Decision Tree -DT)

Karar ağacı algoritması, ilk olarak Morgan ve Sonquist (1963) tarafından geliştirilmiştir. Karar ağaçları, regresyon ve sınıflandırma için kullanılan genel amaçlı bir tahmin yöntemidir. Ancak çoğunlukla sınıflandırma problemlerinin çözümünde tercih edilen bir denetimli öğrenme tekniğidir. Karar ağaçları; güçlü, esnek, kullanımı kolay ve eksik verilerin varlığında, eksik verileri yeni modellere dahil etmenin birden çok yolunu

sunan yöntemlerdir. Bu nedenle, bu yöntem yüksek kaliteli sonuçların üretilmesini sağlamaktadır (de Ville, 2013).

Karar ağacı algoritmaları, çok aşamalı ve ağaç benzeri bir yapıya dayanır. Kurulan yapıda; her ağaç, düğümler ve dallardan oluşur. Karar ağacı algoritmalarının amacı; yapraklar veya düğümler olarak adlandırılan alt kümeler oluşturarak hedef veri alanını tahmin etmektir. Ağacın dallarında tahmin elde edilmiyorsa bu dalda düğüm oluşur. Buna karşın dallarda tahmin elde ediliyorsa dalın sonunda yaprak oluşur. Bu yaprak, veri üzerinden elde edilen tahminlerden biridir (Ayık, Özdemir ve Yavuz, 2007). Her alt küme için aynı işlem bir yaprak oluşana kadar ardışık olarak yürütülür. Burada temel görev; karmaşık bir kararı birkaç basit karara dönüştürmektir ve yorumlanması daha kolay bir çözüme ulaşmaktır (Zhang vd., 2005).

Karar ağacı sınıflandırıcıları, akış şemaları gibi çalışır. Bir karar ağacının her bir düğümü, iki yaprak düğüme ayrılan bir karar noktasını temsil eder. Bu düğümlerin her biri kararın sonucunu temsil eder ve kararların her biri aynı zamanda karar düğümlerine dönüşebilir. Sonunda, farklı kararlar nihai bir sınıflandırmaya yol açacaktır. Karar ağacı algoritması Şekil 7’de gösterilmiştir.



Şekil 7. Karar Ağacı Algoritması

Karar ağacı algoritmalarında veriler, işlev görmesi için karar ağaçları tarafından bir dizi ikili karara bölünerek çalışır. Ağaçtan aşağı doğru hareket etme yeteneğiniz bu

seçimlere bağlıdır. Tahmin edilen sınıflandırma yakalanana ve bir yaprak düğüm ile sonuçlanıncaya kadar, kararlar arasında ilerlemeye devam edilir. Bir karar ağacı algoritması, hangi kararları vereceğine veri kümesini bir dizi karara bölmek için bir dizi farklı yol kullanarak karar verir.

Bir karar ağacını uygularken, temel sorun, kök düğüm ve alt düğümler için en iyi özniteliğin nasıl seçileceğidir. Dolayısıyla, bu tür problemleri çözmek için nitelik seçim ölçüsü (attribute selection measure) olarak adlandırılan bir teknik vardır. Bu ölçüm ile ağacın düğümleri için en iyi öznitelik seçilir. Bu seçim için kullanılan yollardan biri gini İndeks; rastgele bir değişkenin yeni bir örneğinin, veri kümesindeki sınıf etiketlerinin dağılımına göre rastgele sınıflandırılmış olması durumunda, bu örneğin yanlış sınıflandırılma olasılığının bir ölçümünü ifade eder. Gini indeksi, karar ağacı oluşturulurken kullanılan safsızlık veya saflık ölçüsüdür. Yüksek gini indeksine kıyasla düşük gini indeksine sahip bir nitelik tercih edilir.

Bir diğer nitelik seçim ölçüsü bilgi kazancı (information gain) dir. Bilgi kazancı, bir özelliğin bize bir sınıf hakkında ne kadar bilgi sağladığını hesaplar. Bilgi kazancının değerine göre düğümü böler ve karar ağacını oluşturur. Bir karar ağacı algoritması her zaman bilgi kazancının değerini maksimize etmeye çalışır ve en yüksek bilgi kazancına sahip bir düğüm önce bölünür.

2.2.4. Random Forest (RF)

Random Forest modeli, Breiman (2001) tarafından geliştirilen, farklı örnek alt kümelerinde birkaç rastgele ağaçtan oluşan tahmin edicilerin bir kombinasyonudur. Random Forest algoritması, ağaç yapılı tahmincilerin gelişmiş bir sürümü olarak ifade edilir (Akar ve Güngör, 2012).

Random forest yöntemi tahmin için bir dizi ağaç kullanır. Tahmin için bir dizi ağaç kullanmanın ardındaki mantık, birden fazla ağacın tahminini birleştirerek her ağacın kararsızlık sorununu azaltmaktır (Wang vd., 2018). Veri setinden farklı alt kümeler oluşturur ve sonra bu alt kümeler, karar ağacı topluluğundaki her bir karar ağacı tarafından eğitilir. Her bir düğümde rastgele seçilen değişkenler arasından en iyisi kullanılarak her bir düğüm dallara ayrılır. Her düğümü bölmek için rastgele bir özellik seçimi kullanılarak ağaçlar geliştirilir. Geliştirilen ağaçlar daha fazla bölünme mümkün

olmayana kadar büyütülür ve budanmaz (Archer ve Kimes, 2008). Ortaya çıkan model, tüm bu ağaçların oylanmasına dayanmaktadır. Bu fark nedeniyle, aşırı öğrenmeye daha az meyillidir. Güçlü bir tahmin modeli olan random forest algoritması, sınıflandırma ve regresyon tahminlerinde yüksek performans gösterir.

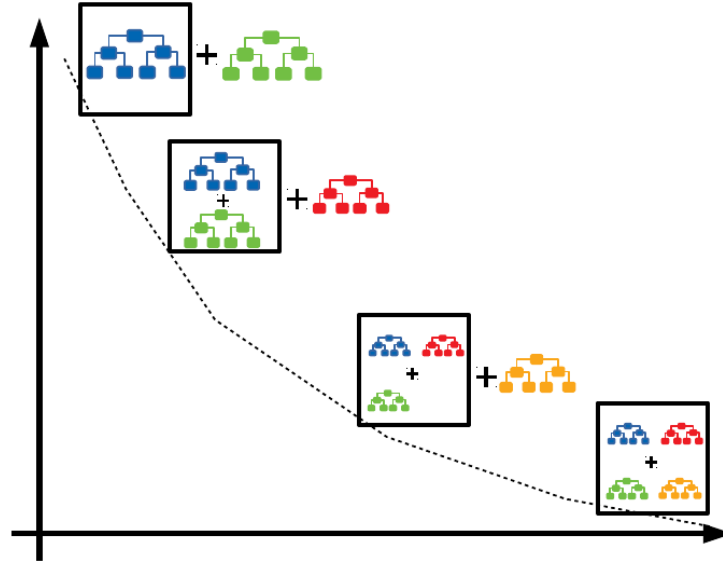
2.2.5. Gradyan Artırmalı Ağaç (Gradient Boosted Tree - GBT)

Gradyan artırmalı ağaç, Friedman (2001) tarafından geliştirilmiştir. Gradient Boosted Tree, “Gradient descent” ve “boosting” kelimelerinin bir araya gelmesi ile oluşan bir karar ağacı algoritmasıdır.

Gradyan artırmalı ağaç algoritmaları; regresyon veya sınıflandırma ağacı modellerinin her ikisi için de kademeli olarak geliştirilmiş tahminler yoluyla tahmine dayalı sonuçlar elde eden, ileriye dönük öğrenme topluluğu yöntemleridir. Artırma (boosting), ağaçların doğruluğunu artırmaya yardımcı olan esnek, doğrusal olmayan bir regresyon prosedürüdür. Artımlı olarak değiştirilen verilere sıralı olarak zayıf sınıflandırma algoritmaları uygulayarak, bir dizi zayıf tahmin modelleri üreten bir dizi karar ağacı oluşturulur.

Gradyan artırma işlemi, ağaçların doğruluklarını artırırken, hızı ve insan yorumlanabilirliğini de azaltır. Gradyan artırma yöntemi, bu sorunları en aza indirmek için ağaç artırmayı genelleştirir. Bu süreç sonucunda, zayıf olarak adlandırılan ağaçlardan daha sağlam tahmin modelleri oluşturur (Gilbert ve ark., 2020). Bu yöntem, modeli diğer artırma yöntemleri gibi aşamalar halinde meydana getirir ancak bunları keyfi bir türevlenebilir kayıp fonksiyonu optimize ederek genelleştirir (Flores ve Keith, 2019).

Gradyan artırma yoluyla kayıp fonksiyonları azaltırken, karar ağaçlarını birleştirerek tahmin hatalarını azaltır. En güçlü algoritmalarından biridir, hızlı çalışır ve çok iyi çözümler verebilir. Regresyon ve sınıflandırma problemlerinde kullanılır (Olu-Ajayi, Alaka, Sulaimon, Sunmola ve Ajayi 2022). Yaygın olarak kullanılan gradyan artırmalı ağaç modelleri: Scikit-learn, XGBoost, LightGBM, Catboost dır. Şekil 8’de gradyan artırma algoritması gösterilmiştir.



(Pal, 2020)

Şekil 8. Gradyan Artırmalı Ağaç Algoritması

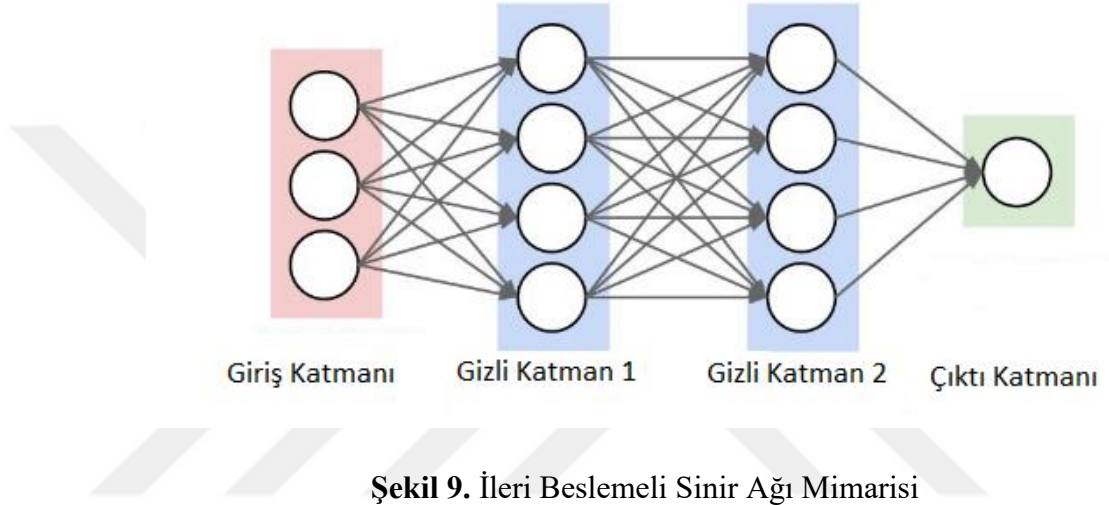
2.2.6. Derin Sinir Ağları (Deep Neural Networks)

Yapay sinir ağları, denetimli öğrenme ve takviyeli öğrenme problemlerinde yaygın olarak kullanılır. Bu ağlar birbirine bağlı bir dizi katmana dayanır. Derin sinir ağı, giriş ve çıkış katmanları arasında birden çok gizli katmana sahip bir yapay sinir ağı algoritmasıdır.

Yapay sinir ağları, kabaca insan beynindeki nöronların nasıl çalıştığından ilham alan bir makine öğrenmesi disiplini. Sinir Ağları, insan sinir hücrelerine benzer prensipler ile çalışır. Bunlar, çeşitli temel değişkenler arasındaki ilişkiyi yakalayan ve verileri insan beyninin yaptığı gibi işleyen bir dizi algoritmadır. Karmaşık sinir ağlarını bazı alanlarda başarılı bir şekilde eğitmek çok fazla veri ve işlem kapasitesi gerektirir. Bu nedenle farklı uygulamalar için çeşitli sinir ağları türleri vardır (Jain vd., 1996).

Çalışmamızda, örüntü sınıflandırma, kümeleme, regresyon, ilişkilendirme, optimizasyon, kontrol ve tahminde başarılı bir şekilde uygulanan ileri beslemeli sinir ağı (feed-forward artificial neural network, FNN) kullanılmıştır. Kullanılan derin sinir ağı algoritması, geri yayılım kullanılarak stokastik gradyan inişi ile eğitilmiş çok katmanlı bir ileri beslemeli yapay sinir ağına dayanır.

İleri Beslemeli Sinir Ağı, düğümlerin dairesel olarak bağlandığı yapay bir sinir ağıdır. Bazı rotaların döngülediği ileri beslemeli bir sinir ağı, tekrarlayan sinir ağı (Recurrent Neural Network- RNN)'nın tam tersidir. İleri beslemeli modelde, girdi yalnızca bir yönde işlendiğinden, temel sinir ağı mimarisinin bir türüdür. Bu tür sinir ağı, tüm bilgileri yalnızca ilettiği için çok katmanlı sinir ağı olarak da bilinir. Şekil 9'da ileri beslemeli sinir ağı mimarisi gösterilmiştir.



Şekil 9. İleri Beslemeli Sinir Ağı Mimarisi

En temel biçiminde, İleri Beslemeli Sinir Ağı, tek katmanlı bir algılayıcıdır. Bir dizi girdi katmana girer ve bu modeldeki ağırlıklarla çarpılır. Ağırlıklı girdi değerleri daha sonra bir toplam oluşturmak üzere toplanır. Değerlerin toplamı, normalde sıfıra ayarlanan önceden belirlenmiş bir eşikten fazlaysa, çıkış değeri genellikle 1'dir ve toplam, eşikten küçükse, çıkış değeri genellikle -1'dir. Tek katmanlı algılayıcı, sınıflandırma için sıklıkla kullanılan popüler bir ileri beslemeli sinir ağı modelidir.

Tek katmanlı algılayıcılar, makine öğrenmesi özelliklerini de içerebilir. Sinir ağı, delta kuralı olarak bilinen bir özelliği kullanarak düğümlerinin çıktılarını istenen değerlerle karşılaştırabilir ve ağı daha doğru çıktı değerleri oluşturmak için eğitim yoluyla ağırlıklarını değiştirmesine izin verir. Bu eğitim ve öğrenme prosedürü, gradyan inişle sonuçlanır. Çok katmanlı algılayıcılarda ağırlıkları güncelleme tekniği hemen hemen aynıdır, ancak süreç geri yayılım olarak adlandırılır. Bu gibi durumlarda, son katman tarafından sağlanan çıkış değerleri, ağ içindeki her bir gizli katmanı değiştirmek için kullanılır.

2.3. REGRESYON ANALİZİ

Regresyon, değişkenler arasındaki korelasyonu bulmaya yardımcı olan ve bir veya daha fazla tahmin değişkenine dayalı olarak sürekli çıktı değişkenini tahmin etmemizi sağlayan denetimli bir öğrenme tekniğidir. Regresyon analizi, bağımlı (hedef) ve bağımsız (yordayıcı) değişkenler arasındaki ilişkiyi bir veya daha fazla bağımsız değişkenle modellemek için kullanılan istatistiksel bir yöntemdir. Daha spesifik olarak, Regresyon analizi, diğer bağımsız değişkenler sabit tutulduğunda bağımsız bir değişkene karşılık gelen bağımlı değişkenin değerinin nasıl değiştiğini anlamamıza yardımcı olur.

Regresyon, hedef değişken ile bağımsız değişken arasındaki ilişkiyi tahmin ederek verilerdeki eğilimleri bulmak için kullanılır. Regresyon analizi yaparak, en önemli faktör, en az önemli faktör ve her bir faktörün diğer faktörleri nasıl etkilediği belirlenir. Sınıflandırma verilerin ait olduğu kategoriyi tahmin ederken regresyon, önceden gözlemlenen verilere dayalı sayısal bir değer öngörür.

Veri bilimi ve makine öğrenmesinde kullanılan çeşitli regresyon türleri vardır. Her türün farklı senaryolarda kendi önemi vardır, ancak özünde, tüm regresyon yöntemleri bağımsız değişkenin bağımlı değişkenler üzerindeki etkisini analiz eder.

2.4. KÜMELEME ANALİZİ (CLUSTER ANALYSIS)

Kümeleme, temel olarak bir tür denetimsiz öğrenme yöntemidir. Denetimsiz öğrenme yöntemi, etiketli yanıtlar olmadan girdi verilerinden oluşan veri kümelerinden referanslar alınarak çalışan bir yöntemdir. Genellikle, bir dizi örnekte bulunan anlamlı yapıyı; açıklayıcı temel süreçleri, üretken özellikleri ve gruplamaları bulmak için bir süreç olarak kullanılır.

Kümeleme, veri noktalarını, aynı gruptaki veri noktalarının aynı gruptaki diğer veri noktalarına daha benzer ve diğer gruptaki veri noktalarından farklı olacak şekilde birkaç gruba ayırma görevidir. Temel olarak, aralarındaki benzerlik ve farklılık temelinde bir nesneler topluluğudur.

Kümeleme, benzer verileri kümeler halinde gruplandırır ve bir veri kümesi için uygun grupları bulmak için kullanılır. Sınıflandırmadan farklı olarak, kümeleme bir tür

yönlendirilmemiş bilgi keşfidir yani, hedef alan yoktur ve veriler arasındaki ilişki yaklaşımlarla tanımlanır (Getahun ve To, 2020).

Kümeleme analizinde amaç verileri birbirleriyle benzer alt kümelere ayırmaktır. Verilerin hangi kümelere ayrılacağı, hatta kaç değişik küme oluşturulacağı verilerin birbirine olan benzerliğine ve uzaklığına göre belirlenmektedir.

Verilerin benzerliği ile kastedilen şey ise aralarındaki mesafenin ölçülmesi ve değerlendirilmesidir. Bu veri tabanında bulunan her bir verinin diğer bir veriyle olan benzerliği ya da diğer verilere uzaklığı olduğu gibi, oluşturulan gerçek ve aday kümeler arasındaki benzerlik ve uzaklığı da içermektedir. En çok kullanılan uzaklık ölçütleri Öklid, Manhattan, Minkowski; benzerlik ölçütleri ise Dice, Jaccard, Cosine, Overlap'dir.

Makine öğrenmesi uygulamalarında verileri benzerliklerine göre kümelemek için yaygın olarak kullanılan bazı kümeleme yöntemleri hiyerarşik kümeleme, k-means, k-medoids, kohonen ağları, bulanık c-ortalamlardır (Erdoğan, 2019).

2.5. TOPLULUK ÖĞRENME (ENSEMBLE LEARNING)

Topluluk öğrenme, aynı problemi çözmek için birden fazla sınıflandırıcının eğitildiği bir makine öğrenme algoritmasıdır. Eğitim verilerinden bir hipotez öğrenmeye çalışan sıradan makine öğrenmesi yaklaşımlarının aksine, topluluk öğrenme yöntemleri bir dizi hipotez oluşturmaya ve bunları kullanmak için birleştirmeye çalışır (Zhou, 2009).

Herhangi bir makine öğrenmesi probleminin amacı, istediğimiz sonucu en iyi tahmin edecek tek bir model bulmaktır. Tek bir model yapıp bu modelin yapabileceğimiz en iyi ve en doğru tahmin edici olduğunu ummak yerine, topluluk yöntemleri sayısız modeli hesaba katar ve bu modellerin ortalamasını alarak tek bir nihai model üretir. Topluluk öğrenme yöntemlerinin amacı, tek bir tahminci üzerinde genellenebilirliği ve sağlamlığı geliştirmek için belirli bir öğrenme algoritması ile oluşturulmuş birkaç temel tahmincinin tahminlerini birleştirmektir.

Dasarathy ve Sheela'nın 1979 yılında yayınlanan çalışması, topluluk öğrenme üzerine yapılan en eski çalışma olarak kabul edilmektedir (G. Wang, Hao, Ma ve Jiang, 2011). Bu çalışma, çoklu sınıflandırıcılar kullanarak özellik uzayını bölümlendirmeyi tartışmıştır (Dasarathy ve Sheela, 1979).

1990 yılında, Hansen ve Salamon, benzer şekilde yapılandırılmış sinir ağlarından oluşan bir topluluk kullanarak bir sinir ağının genelleme performansının geliştirilebileceğini göstermiştir (Hansen ve Salamon, 1990). Aynı yıl Schapire, PAC-öğrenme modeli ile zayıf sınıflandırıcıları birleştirerek güçlü bir sınıflandırıcı oluşturulabileceğini kanıtlamıştır (Schapire, 1990). Freund ve Schapire, yeni bir boosting algoritması yaklaşımı ile zayıf öğrenme algoritmalarının performanslarını iyileştirmek için topluluk öğrenme çalışmalarını sürdürmüştür (Freund ve Schapire, 1997).

2004 yılında Windeatt ve Ardeshir, iyi bir topluluk elde etmek için doğruluk, verimlilik ve çeşitlilik olmak üzere üç gerekli koşulun yerine getirilmesi gerektiğini savunmuştur (Windeatt ve Ardeshir, 2004). Bagui 2005 yılında yayınladığı kitabında sınıflandırıcıları birleştirme felsefesini tartışmıştır (Bagui, 2005).

2014 yılına gelindiğinde Kuncheva kitabında "Sınıflandırıcıları neden birleştirmeliyiz?" ve "Topluluğu nasıl eğitiriz?" gibi sınıflandırıcıların birleştirilmesiyle ilgili bazı felsefi soruları tartışmıştır (Kuncheva, 2014).

Günümüze gelindiğinde, topluluk öğrenme yöntemleri, günlük hayatımızın bir parçası olan karar verme mekanizmamızdan hareketle ortaya çıkan bir alan olarak tanımlanabilir. Bizler günlük hayatta, bir konu hakkında nihai bir karara varmadan önce çeşitli fikirleri tartarak ve birleştirerek doğru kararı veririz. Her karar vericinin bir kusuru vardır. Bu nedenle birden fazla karar verici ile istişare edilerek verilen kararlarda hata payı en aza indirgenmiş olur (Polikar, 2012). Bu bağlamda, topluluk öğrenme, iki veya daha fazla modelin göze çarpan özelliklerini birleştirerek tahminlerde fikir birliğine varma yöntemidir. Son topluluk öğrenme çerçevesi, tahmin hatalarındaki varyansı azalttığından topluluğu oluşturan bireysel modellerden daha sağlamdır.

Topluluk öğrenme algoritması, farklı katkıda bulunan modellerinden tamamlayıcı bilgileri yakalamaya çalışır; yani, katkıda bulunan modeller istatistiksel olarak farklı olduğunda bir topluluk çerçevesi başarılı olur. Başka bir deyişle, aynı veri kümesi üzerinde değerlendirildiğinde performans değişikliği gösteren modeller, bir topluluk oluşturmak için daha uygundur.

Topluluk sınıflandırma modelleri, mükemmel performans elde edebilen ve yeni, görünmeyen veri kümelerine iyi genelleme yapabilen güçlü makine öğrenmesi

araçlarıdır. Topluluk öğrenme teorisinde, birkaçını birleştirerek daha karmaşık modeller tasarlamak için yapı taşları olarak kullanılabilecek zayıf öğrenen temel modeller vardır. Çoğu zaman, bu temel modeller ya yüksek bir yanlılığa sahip olduklarından ya da güçlü bir varyansa sahip olduklarından kendi başlarına çok iyi performans gösteremezler. Topluluk yöntemleri fikri, daha iyi performanslar elde eden güçlü bir topluluk modeli oluşturmak için modellerin birkaçını bir araya getirerek bu tür zayıf modellerin yanlılık ve varyansını azaltmaya çalışmak olarak tanımlanabilir.

Topluluk öğrenme yaklaşımları, bireysel makine öğrenmesi modellerini birleştirip daha iyi bir sınıflama tahmini yapmayı amaçlar. Bu yaklaşım, makine öğrenmesi sonuçlarını iyileştirir ve tek bir modelin tahmininden daha iyi tahminler üretilmesini sağlar (Gürsakar, 2018). Topluluk öğrenme denetimli bir öğrenme algoritmasıdır, çünkü eğitilebilir ve daha sonra tahminlerde bulunmak için kullanılabilir (Kabari ve Onwuka, 2019).

Topluluk öğrenme, sınıflandırıcılar veya uzmanlar gibi birden çok modelin stratejik olarak oluşturulduğu ve belirli bir hesaplama zekasını çözmek için birleştirildiği süreçtir. Topluluk öğrenme algoritmaları bir modelin (sınıflandırma, tahmin, fonksiyon yaklaşımı, vb.) performansını iyileştirmek veya kötü bir modelin talihsiz bir seçim olasılığını azaltmak için kullanılır (Polikar, 2012).

Bir topluluk öğrenme yöntemi oluşturmak için öncelikle toplanacak temel modellerimizi seçmemiz gerekir. Çoğu zaman farklı şekillerde eğitilmiş, homojen zayıf öğrencilere sahip olmak için tek tabanlı bir öğrenme algoritması kullanılır. Bununla birlikte, farklı türde bazı heterojen zayıf öğrencilere sahip temel öğrenme algoritmaları kullanan bazı yöntemler de vardır. Burada önemli olan nokta, zayıf öğrenenleri seçmenin bu modelleri toplama şekli ile tutarlı olması gerektiğidir. Düşük yanlılığa ancak yüksek varyansa sahip temel modeller seçilirse, varyansı azaltma eğiliminde olan bir topluluk öğrenme yöntemi seçilmelidir. Düşük varyanslı ancak yüksek yanlılığa sahip temel modeller seçilirse, yanlılığı azaltma eğiliminde olan bir topluluk öğrenme yöntemi seçilmelidir.

Topluluk öğrenme yönteminde, uygun birleştirme yönteminin seçimi oldukça önemli bir konudur. Topluluk üyelerini birleştirmek için bir dizi farklı strateji kullanılabilir. Sağladıkları belirli teorik garantiler nedeniyle bagging (torbalama),

boosting (artırma), voting (oylama) ve stacking (istifleme) yaygın olarak kullanılan topluluk öğrenme yöntemleridir.

2.5.1. Bagging (Torbalama) Topluluk Öğrenme Yöntemi

Breiman (Breiman, 1996) tarafından geliştirilen bagging algoritması topluluk tabanlı algoritmalar arasında en eski, en basit ancak etkili algoritmalarından biridir. Bagging algoritmasında, her model ayrı ayrı eğitilir ve bir ortalama alma süreci kullanılarak birleştirilir. Bagging yönteminin birincil odak noktası, herhangi bir modelin bireysel olarak sahip olduğundan daha az varyans elde etmektir. Bootstrap Aggregation'ın kısaltması olarak ifade edilen bu yöntemde eğitim veri seti içerisinde her seferinde farklı eğitim verisi alt kümeleri, tüm eğitim verilerinden rastgele (değiştirilerek) çekilir. Her eğitim verisi alt kümesi, aynı türden farklı bir temel öğrenmeyi eğitmek için kullanılır (Polikar, 2012; Wang vd., 2011). Bir sınıflandırıcı oluşturulan her eğitim verisi alt kümesini eğitir. Tüm sınıflayıcılar, aynı anda farklı alt eğitim verilerini sınıflandırır. Bagging topluluk öğrenme yöntemi genellikle homojen zayıf öğrencileri dikkate alır, bunları birbirinden bağımsız olarak paralel olarak öğrenir ve bir tür ortalama alma sürecini izleyerek birleştirir.

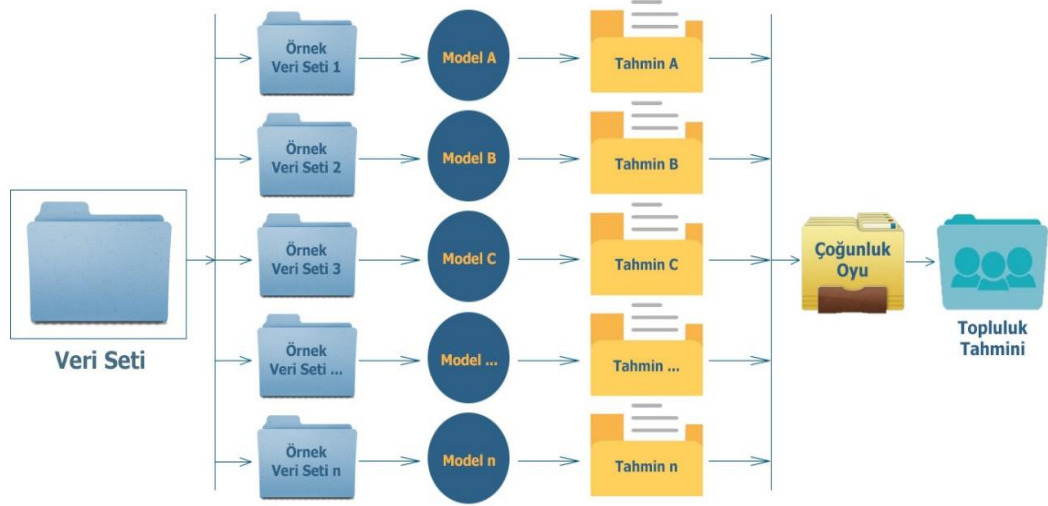
Topluluk algoritmalarında, bagging yöntemleri, orijinal eğitim kümesinin rastgele alt kümeleri üzerinde bir kara kutu tahmincisinin birkaç örneğini oluşturan ve daha sonra nihai bir tahmin oluşturmak için bireysel tahminlerini toplayan bir algoritma sınıfı oluşturur. Bu yöntemler, bir temel tahmincinin varyansını, yapım prosedürüne rasgeleleştirmeyi dahil ederek ve daha sonra bir topluluk oluşturarak azaltmanın bir yolu olarak kullanılır. Bagging yöntemleri çoğu durumda, temel algoritmayı uyarlamayı gerekli kılmadan, tek bir modele göre iyileştirmenin çok basit bir yolunu oluşturur. Bu yöntemler aşırı uyumu (overfitting) azaltmak için bir yol sağladıklarından, bagging yöntemleri, genellikle zayıf modellerle iyi çalışır.

Sınıflandırıcıların tahminlerini birleştirmek amacıyla bagging yöntemi çoğunluk oyu stratejisini kullanır. Bu stratejide; her model, her test örneği için bir tahminde bulunur (oy kullanır) ve nihai çıktı tahmini, oyların yarısından fazlasını alan tahmindir. Tahminlerin hiçbirisi oyların yarısından fazlasını alamazsa, topluluk öğrenme yönteminin bu örnek için istikrarlı bir tahmin yapamadığı söylenir. Bu yaygın olarak kullanılan bir

teknik olmasına rağmen, en çok oy alan tahmin (oyların yarısından az olsa bile) nihai tahmin olarak denenebilir. Bu yöntem “çoğulcu oylama” olarak adlandırılır (Doğaner, 2020).

Basit olduğu kadar uygulaması kolay olan bu strateji, temel öğrenme stratejileriyle birleştirildiğinde varyansı azaltabilir. Bagging algoritması, özellikle sınıflandırıcılar kararsız olduğunda ve aşırı öğrenme (overfitting) eğiliminde olduğunda işe yarar, yani eğitim verilerindeki küçük değişiklikler, tahmin edilen çıktıda büyük değişikliklere yol açar. Farklı standart sapmalar, ortalamalar vb. gibi farklı istatistiksel özelliklerden oluşan bireysel sınıflandırıcıları bir araya getirerek varyansı etkili bir şekilde azaltır. Bagging algoritması, Karar Ağaçları gibi yüksek varyanslı modeller ile birlikte iyi çalışır. Doğrusal regresyon gibi düşük varyanslı modellerle kullanıldığında öğrenme sürecini pek etkilemez.

Bagging topluluk öğrenme yönteminin süreci Şekil 10’da gösterilmiştir.



Şekil 10. Bagging Topluluk Öğrenme Yöntemi İşlem Aşamaları

2.5.2. Boosting (Artırma) Topluluk Öğrenme Yöntemi

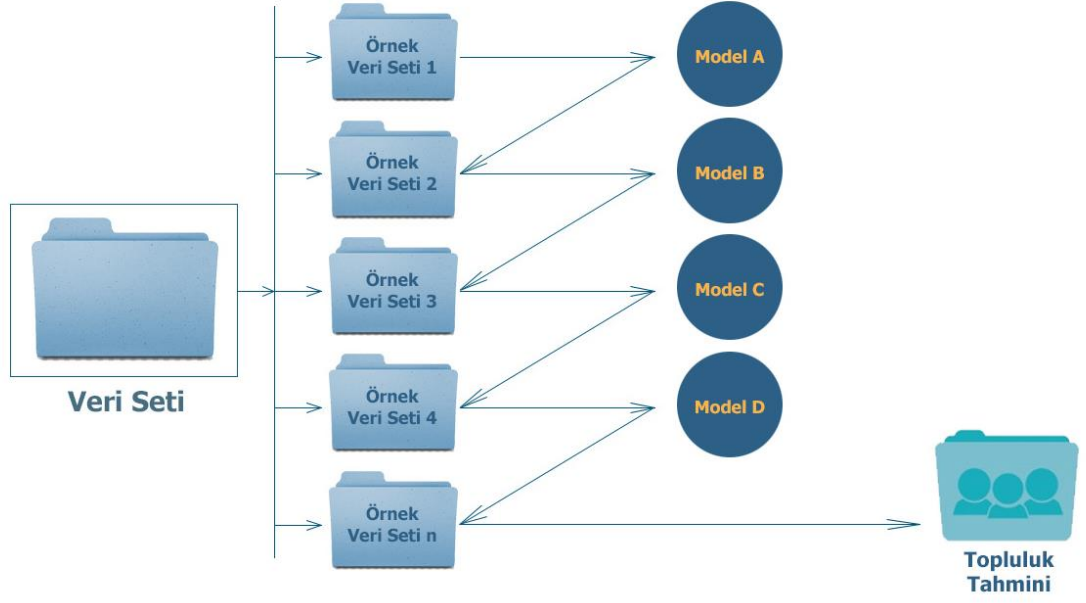
Boosting topluluk öğrenme yöntemi, Freund ve Schapire (Freund ve Schapire, 1996) tarafından zayıf öğrenmeleri birleştirerek daha güçlü bir sınıflandırıcı oluşturmak

için geliştirilen yinelemeli bir yaklaşımdır. Boosting, genellikle homojen zayıf öğrenmeleri dikkate alır, onları uyarlanabilir bir şekilde sırayla öğrenir ve bir strateji izleyerek bunları birleştirir.

Boosting yöntemleri, daha iyi performans gösteren güçlü bir öğrenici elde etmek için bir araya getirilen bir model ailesi oluşturma mantığı ekseninde bagging yöntemleriyle benzer şekilde çalışır. Bununla birlikte, temelde varyansı azaltmayı amaçlayan bagging yönteminden farklı olarak, boosting, çok sayıda zayıf öğreniciyi, çok uyarlanabilir bir şekilde sıralı olarak yerleştirmeyi içeren bir tekniktir. Sezgisel olarak, her yeni model çabalarını en zor gözlemlere odaklar. Böylece sürecin sonunda, boosting algoritmasının varyansı azaltma etkisine sahip olabileceğini fark edebilsek bile daha düşük yanlılığa sahip güçlü bir öğrenici elde edilir. Boosting, bagging gibi, sınıflandırma problemleri için olduğu kadar regresyon için de kullanılabilir.

Boosting algoritması üç sınıflayıcıdan oluşmaktadır. İlk sınıflayıcı, eğitim veri seti içerisinde rastgele seçilen bir veri setini sınıflandırmaktadır. İkinci sınıflayıcı, yarısı birinci sınıflayıcı tarafından yanlış sınıflandırılmış, diğer yarısı doğru sınıflandırılmış veri seti üzerine eğitilmektedir. Son sınıflayıcı daha önce iki sınıflayıcının uyum göstermediği veriler üzerine eğitilmiştir. Her sınıflayıcı tahminler üretmektedir. Üç sınıflandırıcı, üçlü çoğunluk oyu ile birleştirilir. Boosting topluluk öğrenme yönteminde çoğunluk oyunu alan sınıflandırıcının tahminleri, topluluk tahmini olarak kabul edilir (Rokach, 2010). Boosting algoritması tarafından elde edilen son model, kendi performanslarına göre ağırlıklandırılmış birkaç temel öğrenmelerin doğrusal bir birleşimidir (Wang vd., 2011). Boosting algoritmalarının birkaç versiyonu olmasına rağmen, en yaygın kullanılanı Freund ve Schapire (1996) tarafından önerilen AdaBoost olarak bilinmektedir.

Boosting topluluk öğrenme yönteminin süreci Şekil 11’de gösterilmiştir.



Şekil 11. Boosting Topluluk Öğrenme Yöntemi İşlem Aşamaları

2.5.3. Voting (Oylama) Topluluk Öğrenme Yöntemi

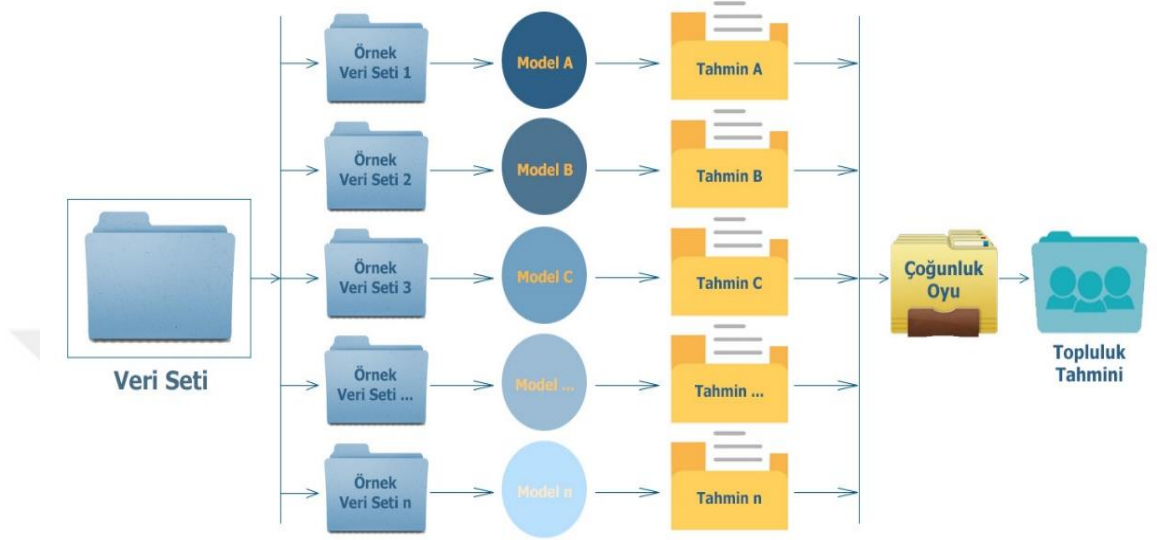
Voting, sınıflandırma veya regresyon problemlerinde kullanılan basit topluluk öğrenme algoritmasıdır ve genellikle çok etkilidir. Voting yöntemi, iki veya daha fazla alt model oluşturarak çalışır. Her bir alt model tahminler yapar. Tahminlerin ortalamasını veya modunu alarak, her bir alt modelin oy kullanmasına izin vererek tahminler birleştirilir (Kabari ve Onwuka, 2019).

Voting topluluk öğrenme yöntemi, majority voting (çoğunluk oyu) tekniğinden yararlanarak birleştirme işlemini yapmaktadır. Çoğunluk oylamasının üç versiyonu vardır. Birinci versiyonda topluluk, tüm sınıflandırıcıların hemfikir olduğu sınıfı seçer (oybirliği ile oylama). İkinci versiyonda sınıflandırıcı sayısının yarısından en az bir fazlası tarafından tahmin edilen çoğunluk (basit çoğunluk) geçerli olur. Üçüncü versiyonda, bu oyların toplamı %50'yi aşıyor olsun ya da olmasın en fazla oyu alan (çoğunluk veya sadece çoğunluk oylaması) sınıf seçilir (Polikar, 2006).

Voting regresyon problemlerinde biraz farklı çalışır. En yüksek frekansa sahip tahmini bulmak yerine, oylama ile oluşturulan regresyon modelleri, her modelin tahminlerini alır ve nihai bir tahmin elde etmek için ortalama değerlerini hesaplar. Genel

anlamda, sınıflandırma veya regresyon problemlerinde voting, tahmin performansını artırmanın bir yolu olarak hizmet eder.

Voting topluluk öğrenme yönteminin süreci Şekil 12’de gösterilmiştir.



Şekil 12. Voting Topluluk Öğrenme Yöntemi İşlem Aşamaları

2.5.4. Stacking (İstifleme) Topluluk Öğrenme Yöntemi

Wolpert tarafından geliştirilen stacking (istifleme) topluluk öğrenme yöntemi bir öğrenmenin bireysel öğrenmeleri birleştirmek için eğitildiği genel bir prosedür olarak tanımlanmaktadır (Wolpert, 1992).

Stacking, iki katmanlı tahmin edicilerden oluşan sınıflandırma veya regresyon modellerini birleştirmenin bir yoludur. İlk katman, test veri kümelerindeki çıktıları tahmin etmek için kullanılan tüm temel modellerden oluşur. İkinci katman, temel modellerin tüm tahminlerini girdi olarak alan ve yeni tahminler üreten meta sınıflandırıcı veya regresörden oluşur. Burada, bireysel öğrenmeler birinci seviye öğrenmeler olarak adlandırılır, bu öğrenmeleri birleştiren birleştiriciye ikinci seviye öğrenme veya meta-öğrenme denir.

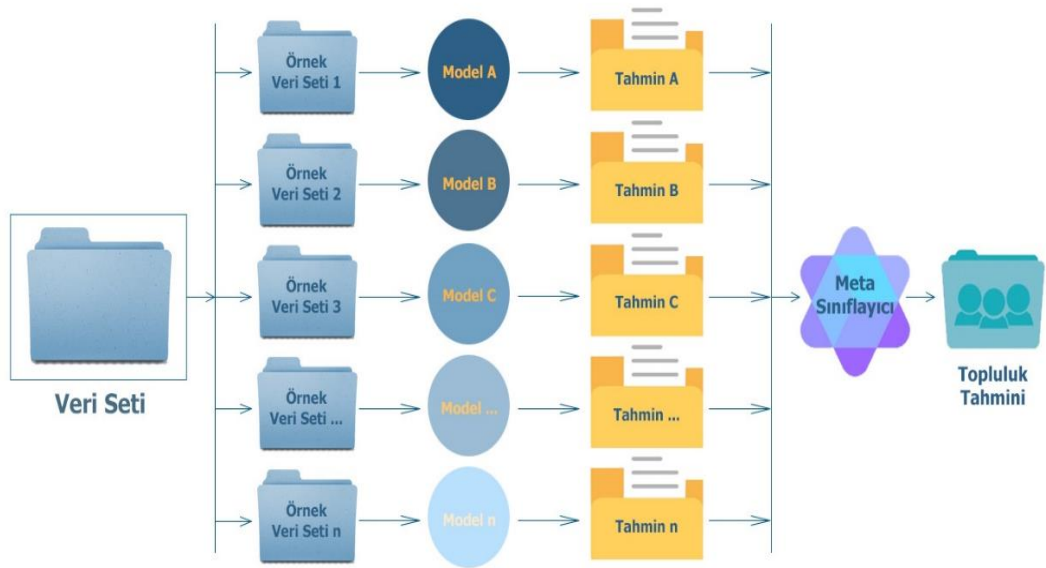
Temel fikir, orijinal eğitim veri setini kullanarak birinci seviyedeki öğrenmeleri eğitmek ve ardından birinci seviye öğrenmelerin çıktılarını girdi olarak kabul eden ikinci seviyeyi eğitmek için yeni bir veri seti oluşturmaktır (Zhou, 2019).

Bagging ve boosting'in aksine, stacking normalde aynı türdeki temel öğrenmeleri birleştirmek için kullanılmaz. Bunun yerine, farklı öğrenme algoritmaları tarafından oluşturulan temel sınıflayıcılara uygulanır (Wang vd., 2011).

Stacking, çoklu sınıflandırmaları veya regresyon modelini birleştirmenin bir yoludur. Bagging, varyansı azaltmak için yüksek varyansa sahip birden fazla benzer modelin ortalamasının alınmasına izin verirken Boosting, varyansı küçük tutarak yanlılığı azaltmak için birden fazla artımlı model oluşturur.

Stacking farklı bir paradigmadır. Bu yöntem farklı tür sınıflayıcıların tahminlerini meta sınıflayıcı için girdi kabul ederek bu tahminlerden daha yüksek performanslı bir tahmin üretme esasına dayanmaktadır (Doğaner, 2020).

Stacking algoritmasının amacı, aynı problem için farklı modellerden oluşan bir uzayı keşfetmektir. Buradaki fikir, bir öğrenme problemine, problemin tüm alanını değil, problemin bir kısmını öğrenebilen farklı tipte modellerle saldırabilmeyi sağlamaktır. Böylece, birden fazla farklı öğrenmeler oluşturabilir ve bunları, öğrenilen her model için bir tahmin olmak üzere bir ara tahmin oluşturmak için kullanır. Ardından, aynı hedefi ara tahminlerden öğrenen yeni bir model ekleyerek devam ettirir. Bu son modelin diğerlerinin üzerine yığıldığı söylenebilir. Böylece genel performans iyileşir ve genellikle herhangi bir ara modelden daha iyi bir model elde edilir. Bununla birlikte, herhangi bir makine öğrenmesi tekniğinde olduğu gibi, herhangi bir garanti vermediğine dikkat etmek gerekir. Stacking topluluk öğrenme yönteminin süreci Şekil 13'te gösterilmiştir.



Şekil 13. Stacking Topluluk Öğrenme Yöntemi İşlem Aşamaları

Sonuç olarak; klasik istatistiksel modeller, birkaç düzine girdi değişkeni ve günümüzde küçük ila orta düzeyde kabul edilebilecek örneklem büyüklükleri olan veriler için tasarlanmıştır. Buna karşılık, makine öğrenmesi yöntemleri genellikle, daha büyük, zengin ve hantal verilerdeki kalıpları bulmak için genel amaçlı öğrenme algoritmalarını kullanarak tahmine odaklanır. İstatistiksel modellerde, model sistemin gözlemlenmeyen yönlerini doldurur. Ancak, girdi değişkenlerinin sayısı ve aralarındaki olası ilişkiler arttıkça, bu ilişkileri yakalayan model daha karmaşık hale gelir. Bu durumda, istatistiksel çıkarımlar daha az kesin hale gelir ve istatistiksel ve makine öğrenmesi yaklaşımları arasındaki sınır daha belirsiz hale gelir (İj, 2018).

Makine öğrenmesi yöntemleri, bir dizi tahmin değişkeninin tamamını veya bir kısmını bir sonuçla ilişkilendirmek için hesaplama algoritmalarından oluşur. Modeli tahmin etmek için, en uygun olanı ya rastgele ya da deterministik olarak arar. Bu arama işlemi, farklı algoritmalar arasında farklılık gösterir. Bununla birlikte, bu arama yoluyla, her algoritma iki rakip çıkarı dengelemeye çalışır: yanlılık ve varyans. Makine öğrenmesi bağlamında, yanlılık, uygun tahminlerin gerçek değerlere ne ölçüde karşılık geldiğidir yani model, popülasyondaki 'gerçek' ölüm riskini ne kadar doğru tahmin ediyor? Varyans tahminlerin giriş verilerindeki bozulmalara duyarlılığı nedir? Sorularının karşılığıdır. Bir modelin yanlılığını ve varyansını ayrı ayrı ölçmek mümkün olmasa da, bu iki değer kayıp

fonksiyonları ile birlikte özetlenir. Amacımız hem yanlılığı hem de varyansı azaltmak olsa da, bu iki hedef genellikle çatışır: azaltılmış yanlılık, varyansı artırabilir ve bunun tersi de gerçekleşebilir. Örneğin, veri kümemizdeki tüm ölümleri doğru bir şekilde tahmin eden bir algoritma oluşturabiliriz. Bununla birlikte, bu model, esasen 'istatistiksel gürültüyü' modelleyerek, veri setimizin bireysel karmaşıklıklarına çok özel olarak bağlı olacak şekilde yapılandırılabilir. Bu durumda model, bir doğrulama veri kümesinde uygulandığında (yani yüksek varyansa sahip olduğunda) düşük performans gösterecektir. Bu aynı zamanda bir aşırı öğrenme (overfit) modeli olarak da adlandırılır.

Yanlılık ve varyansı dengelemek için farklı yaklaşımlar kullanılır, topluluk öğrenme yöntemleri bu yaklaşımlardandır (Goldstein, Navar ve Carter, 2017).

Özetle; prensipte istatistik ve makine öğrenmesi alanında birçok yöntem, hem tahmin hem de çıkarım için kullanılabilir ancak bu yöntemlerin zayıf yanlarının iyileştirilmesi için topluluk öğrenmesi yaklaşımlarına başvurmak gerekir.

Modellerin genel özellikleri karşılaştırmalı olarak Tablo 1’de ifade edilmiştir.

Tablo 1. Modellerinin Genel Özellikleri (Sadece Enerji Uygulamaları için Değil)

Genel Grup	Katkılar ve Avantajlar	Zayıf Yönler ve Dezavantajlar
Makine Öğrenmesi Dışındaki Yöntemler	-Geniş literatürde yaygın ve köklü bir yer edinmesi -Farklı zamansal çözümlemelerde zaman serisi tahminlerinde iyi performans göstermesi -Tahminin istatistiksel önemini sağlaması -Genellikle elde edilen tahminlerdeki belirsizliği ölçmesi -Herhangi bir vaka çalışmasına hızlı uygulanması	-Kısa vadeli karmaşık zaman dizilerini tahmin etmede zayıf performans göstermesi -Çok güvenilir olmayan çeşitli varsayımlara bağımlılık -Tahmin modeline eklenen değişkenleri değerlendirmede sınırlı kabiliyet -Açıklayıcılar arasındaki korelasyonlara duyarlılık -Boyutluluğun laneti
Makine Öğrenmesi Yöntemleri	-Değişkenlerin doğası hakkında varsayımlar gerektirmemesi -Literatürdeki çeşitli uygulamalar için artan şekilde kabul edilmiş metotlar olması -Karmaşık zaman dizilerine daha iyi uyabilen daha esnek metotlar olması -Zaman serileri tahmininde farklı değişkenleri barındırması -Sıklıkla klasik metotlardan daha iyi tümevarım kabiliyeti sağlaması	-Sayısal olarak klasik metotlardan daha pahalı olabilir olması -Zaman dizileri tahmininde, çoğunlukla tahminlerin istatistiksel birlikte çalışabilirliğinden ziyade eğri uydurma amaçlar için kullanılabilir olması -Pek çok metotun uydurma davranışının hala çok az anlaşılmış olması -Boyutluluğun laneti -Bir vaka çalışmasını öğrenmede doğal istikrarsızlık
Topluluk Öğrenme Yöntemleri	-Değişkenlerin doğası hakkında varsayım gerektirmemesi -En uygun yaklaşımları sağlayabilen çok esnek metotlar olması -Tek modelleme yapılarından çok daha istikrarlı performansın keyfini çıkarması -Belirsizlik hakkında bilgi sağlayabilmesi -Boyutluluğun etkisini önemli ölçüde azaltabilir; yüksek ölçülü sistemlerle performans üzerinde önemli etki olmaksızın daha iyi başa çıkılabilir olması	-Nispeten yeni öğrenme yapıları -Seri öğrenmede sayısal olarak pahalı metotlar oluşturabilmesi -Çoğunlukla regresyondan ziyade sınıflandırma problemleri için uygun olması -Çeşitlilik kavramı, onun tümevarım kabiliyetine katkısının, genellikle yaygın grup modellerinin pek çoğunda açık bir biçimde ele alınamaması -Yaygın öğrenme yapıları, belli bir vaka çalışmasına ve belirli bir alana uygulandığı zaman dikkatli değerlendirme gerektirmesi

Kaynak: (Alobaidi, Chebana ve Meguid, 2018)

ÜÇÜNCÜ BÖLÜM

TOPLULUK ÖĞRENME YÖNTEMLERİ İLE HANEHALKI TOPLAM ENERJİ HARCAMALARININ TAHMİN EDİLMESİ

3.1. LİTERATÜR TARAMASI

Hanehalkı enerji tüketiminin miktarı, dünya çapında toplam enerji tüketiminin önemli bir kısmını oluşturmaktadır. Farklı hanelerin enerji tüketim kalıpları ve enerji kullanım kararlarının yüksek olması nedeniyle hanehalkı enerji kullanımı çeşitli faktörlerden etkilenmektedir. Bununla birlikte, hanehalkı enerji tüketimi büyük bir tasarruf potansiyeline sahiptir. Hanehalklarının enerji tüketimi davranışlarını anlamak ve değiştirmek, enerji verimliliğini artırmanın ve enerji tasarrufunu teşvik etmenin etkili yolları olarak kabul edilmektedir. İnsanların enerji kullanım davranışlarındaki değişiklikleri teşvik etmeyi amaçlayan müdahale stratejileri, politika yapıcılar için her zaman geliştirilmesi gereken bir konudur. Artan enerji fiyatları, küresel kaynakların tükenmesi ve iklim değişikliği gibi ciddi durumlarla başa çıkmak için hanehalkı enerji verimliliğinin iyileştirilmesi, enerji tasarrufu, temiz enerjiye ulaşımın sağlanması ve sürdürülebilirlik konuları önem kazanmaktadır.

Hanehalkları, hanelerindeki faaliyetler için enerjiyi çok yönlü olarak kullanmaktadır. Hanelerde talep edilen enerji, elektrik kullanımından su kullanımına, yemek pişirmeden yerleşim yerlerinde aydınlatmaya kadar çeşitli amaçlar için kullanılmaktadır. Hanehalkının enerji kullanımını doğrudan etkileyen bazı temel değişkenlerin yanı sıra dolaylı olarak etkileyen değişkenler de bulunmaktadır. Bu kullanım hem doğrudan elektrik ve doğalgaz olarak düşünülebildiği gibi aynı zamanda dolaylı olarak üretim, ulaşım ve diğer mal ve hizmetlerin ticaretini içermektedir. Hanehalklarının özellikleri, birçok çalışmada enerji harcamalarını etkileyen ana faktörlerden biri olarak tanımlanmaktadır.

Hanelerin enerji kullanım profilleri açısından enerjiyi hangi demografik özelliklere ve hangi sosyo-ekonomik alt yapıya sahip olarak kullandığı önem taşımaktadır. Günümüzde yapılan çalışmalarda, hanehalklarının enerji harcama değişkenlerini belirleyen birçok sosyoekonomik faktör yer almaktadır. Bu anlamda, pek çok bilim insanı

ve akademisyen enerji tüketim davranışları hakkında sosyodemografik değişkenler perspektifinden birçok çalışma yürütmektedir (Ding, Wang, Liu ve Long, 2017). Geniş bir literatür, farklı istatistiksel yaklaşımlar ve ekonometrik modeller kullanarak; büyüyen bir literatür, farklı makine öğrenmesi yöntemleri kullanarak hanelerdeki enerji kullanım profillerini keşfetmeye odaklanmıştır (Adusah-Poku ve Takeuchi, 2019). İstatistiksel veya ekonometrik yöntemler ile yapılan çalışmaların yanında makine öğrenmesi yöntemleri ile yapılan çalışmalar aşağıda özetlenmiştir.

3.1.1. İstatistiksel Yöntemler ve Ekonometrik Modeller Kullanılarak Hanehalkı Enerji Tüketimi Alanında Yapılan Çalışmalar

Regresyon analizi ile yerel enerji kullanım kalıpları üzerine bir araştırma yapan çalışmada; konut özelliklerinden, hanehalkı özelliklerinden ve cihaz sahipliğinden kaynaklı elektrik enerjisi tüketim düzeyi hakkında tanımlayıcı bilgiler sunulmuştur (Tso, 2003). Çok terimli bir logit modeli kullanılan çalışmada; hem kırsal hem de kentsel alanlarda aylık hane halkı harcamaları ve eğitim, yakıt tercihlerinin belirlenmesinde önemli bir rol oynamıştır. Araştırma, hanehalkı üyelerinin eğitim düzeyi arttıkça modern yakıtları tercih ettiklerini, kırsal hanelerin oduna bağımlı kaldığını göstermiştir. Bulgular ayrıca, hem kentsel hem de kırsal alanlarda hane reisinin kadın olduğu hanelerin modern yakıtları tercih etmesiyle birlikte, cinsiyetin yakıt seçiminde bir fark yarattığını da ortaya koymuştur (Narasimha Rao ve Reddy, 2007). Nicel bir görüşme anketi ve kontrollü deneyler olmak üzere iki farklı yöntem kullanılan çalışmada; kadınların erkeklere göre oda sıcaklığından daha az memnun oldukları ve daha yüksek oda sıcaklıklarını tercih ettikleri vurgulanmıştır (Karjalainen, 2007).

Sıradan en küçük kareler (ordinary least squares - OLS) regresyon modellerini kullanan çalışmada; cihaz sahipliğine bağlı olarak konut enerji talebini etkileyen faktörler araştırılmıştır. Sonuç olarak, bir ev tarafından kullanılan alan ve su ısıtma yöntemlerinin, evsel enerji kullanımını açıklamak için elektrikli cihazlardan daha önemli olduğu belirtilmiştir (Leahy ve Lyons, 2010).

Kantil regresyon analizi kullanılan çalışmada; konut enerji tüketimi yelpazesi anlaşılmaya çalışılmıştır. Sonuç olarak, konut boyutunun alan düzenlemesi için önemli olmasına rağmen, konut tipinin daha incelikli bir etkiye sahip olduğu gösterilmiştir (Kaza,

2010). OLS regresyon tahmini kullanan çalışmada; kişi başına hane elektrik tüketimi ile hane özellikleri arasındaki ilişki iki ölçek ile incelenmiştir. Hanehalkı ve konut özelliklerinin konut elektrik tüketimi üzerinde önemli bir etkiye sahip olduğu iki ölçek için tutarlı sonuçlarla gösterilmiştir (Wiesmann, Lima Azevedo, Ferrão ve Fernández, 2011).

Regresyon analizi uygulanan çalışmada; binanın yaşı, ısınma için temel bir değişken olarak bulunmuştur (Brounen, Kok ve Quigley, 2012). Çok terimli logit model kullanılan çalışmada; Türkiye'de hanelerin enerji tercihlerini etkileyen faktörler araştırılmıştır. Araştırmada hanelerin aylık gelirinin ve hanedeki kişilerin yaşının enerji tercihleri üzerinde önemli bir etkiye sahip olduğu tespit edilmiştir (Özcan, Gülay ve Üçdoğruk, 2013).

Hanehalkı enerji tüketiminin fiziksel ve sosyo-ekonomik itici güçlerini konut düzeyinde araştıran bir çalışmada; tanımlayıcı sonuçlar Tablolar halinde sunulmuştur. Konutların yapısı, hanehalkının ekonomik özellikleri ve enerji verimliliği önlemlerinin enerji tüketimi üzerindeki etkisi karşılaştırmalı bir şekilde incelemiştir. Ayrıca enerji verimliliğindeki değişiklikleri izlemek, tüketicilere sunulan bilgileri iyileştirmek, programlar geliştirmek ve değerlendirmek için bir veri çerçevesi geliştirmiştir (Wyatt, 2013).

Aydınlatma ve cihaz kullanımının hanelerde elektrik tüketimine etkisini belirlemeyi amaçlayan çalışmada, üç regresyon modeli kullanılmıştır. Hanenin büyüklüğü ve tipi, kurutucu kullanımı, yıkama döngüleri ve duş sayısının elektrik tüketimi için önemli değişkenler olduğu belirlenmiştir (Bedir, Hasselaar ve Itard, 2013). Multinomial logit model kullanılan çalışmada; aydınlatma, pişirme ve ısınma için yakıt seçiminde gelir ve eğitimin kritik değişkenler olduğu vurgulanmıştır (Rahut, Das, de Groote ve Behera, 2014).

Kantil regresyon kullanılan çalışmada; evli çiftlerin olduğu hanelerin evli olmayan çiftlerin bulunduğu hanelere göre metrekare başına daha az enerji tükettiği, yetişkin sayısı arttıkça metrekare başına düşen enerjinin arttığı tespit edilmiştir (Valenzuela vd., 2014). Çok değişkenli probit, tobit ve heckman iki aşamalı seçim modelleri kullanılan çalışmada; hanehalkının daha temiz enerji kaynakları seçiminin gelir, yaş, cinsiyet ve hane reisinin eğitimi ile önemli ölçüde ilişkili olduğu bulunmuştur (Rahut, Behera ve Ali, 2016).

Regresyon analizi ile hanelerin sosyal sınıfının ve yaşam tarzının belirlendiği çalışmada; hanelerin enerji tüketim tercihlerinde gelir, eğitim, cinsiyet, yaş, meslek gibi sosyoekonomik ve demografik özelliklerin önemli bir rol oynadığı belirtilmiştir (Salari ve Javid, 2017).

Pearson korelasyon analizi ve çoklu regresyon analizi kullanılan çalışmada; erkeklerin ve düşük gelirli ailelerin yüksek gelirli ailelere göre anlamlı düzeyde daha yüksek enerji tasarrufuna sahip olduğu gösterilmiştir. Ayrıca, hanenin büyüklüğünün ve hanenin mülkiyet durumunun enerji tüketimi üzerinde önemli bir etkisi olduğu vurgulanmıştır (Ding vd., 2017). Üç aşamalı bir ampirik analiz yapılan çalışmada; ev enerji tüketimi için on tüketici tipolojisi oluşturulmuştur. En düşük enerji tüketicilerinin yalnız yaşayanlar veya 40 yaşın altındaki çocuksuz çiftler (%60) olduğu, en yüksek tüketicilerin (%42) ise 40 yaşın üzerindeki aileler (%55) veya daha yüksek yaş grubundaki bekar insanlar ve çocuksuz çiftler (%42) olduğu keşfedilmiştir (Lévy ve Belaïd, 2018).

Çift engel modeli kullanılan çalışmada; hanehalkı enerji harcamasını etkileyen faktörler incelenmiştir. Yakıt tercihlerine bağlı olarak kırsal ve kentsel farklılıklar sunulmuştur (Adusah-Poku ve Takeuchi, 2019). Tobit model tahmini kullanılan çalışmada; kırsal kesimdeki hanelerin enerji tüketiminin özelliklerinin ayrıntılı bir analizi sunulmuştur (Zou ve Luo, 2019). Yakıt seçimlerinin itici güçlerini tahmin etmek için çok değişkenli bir probit modeli ve hanehalkının enerji tüketimine bağımlılığını araştırmak için görünüşte ilişkisiz regresyon (Seemingly Unrelated Regression - SUR) tahmin modeli kullanılan çalışmada; hanehalkının temiz enerji seçimleri araştırılmıştır. Sonuç olarak, nispeten daha eğitilmiş reisler tarafından yönetilen ve ekonomik açıdan varlıklı hanelerin, temiz enerjiyi seçme ve harcama olasılığının daha yüksek olduğu tespit edilmiştir (Pallegedara, Mottaleb ve Rahut, 2021).

Ki-kare testleri ve çok değişkenli probit model kullanılan çalışmada; aydınlatma için kırsal hanehalkı enerji seçimlerinin gelir düzeyi, aile büyüklüğü, konum, eğitim düzeyi, teknoloji maliyeti ve pazara olan mesafe değişkenlerinden önemli ölçüde etkilendiği gösterilmiştir. Hanehalkı enerji tercihlerini ve geçişini belirlemede, gelir temel bir faktör olmaya devam ederken çeşitli bu tercihlerde gelir dışı faktörlerin de önemli bir rol oynadığı ortaya konulmuştur (Wassie, Rannestad ve Adaramola, 2021).

Kantil regresyon ve bayesyen regresyon yaklaşımları kullanılan çalışmada; ALASSO (Adaptive Lasso) değişken seçim yöntemi ile enerji harcamaları incelenmiştir. Gelir, yaş, hanehalkı büyüklüğü, konut büyüklüğü ve istihdam durumu, enerji harcaması için göze çarpan belirleyiciler olarak bulunmuştur. Ayrıca, enerji harcamaları konusunda sosyo-ekonomik özellikler çok daha büyük bir etkiye sahipken, konut özelliklerinin orta düzeyde bir etkiye sahip olduğu tespit edilmiştir (Belaid ve Rault, 2021).

3.1.2. Yapay Zeka Yöntemleri Kullanılarak Hanehalkı Enerji Tüketimi Alanında Yapılan Çalışmalar

Elektrik enerjisi tüketiminin tahminini yapan bir çalışmada; üç modelleme yöntemi karşılaştırılmış, karar ağacı ve sinir ağları makine öğrenmesi yöntemlerinin regresyon analizinden daha iyi alternatifler olduğu bulunmuştur (Tso ve Yau, 2007). Denetimli makine öğrenmesi tekniklerini kullanan bir çalışmada; bir hanenin elektrik tüketiminden hareketle hanenin belirli “özelliklerini” otomatik olarak tahmin etmek için bir sistem önerilmiştir (Beckel, Sadamori, Staake ve Santini, 2014). Karar ağacı ve yapay sinir ağları (Artificial Neural Networks – ANN) gibi yapay zeka yöntemlerini kullanarak bina elektrik enerjisini tahmin eden çalışmada; tek tahmin yönteminin yanı sıra, iki tahmin yönteminin birleşiminin daha doğru sonuçlar verdiği belirtilmiştir (A. S. Ahmad vd., 2014).

Denetimsiz kümeleme yöntemleri ile hanehalkı elektrik kullanımının incelendiği çalışmada; k-means, k-medoid ve SOM (Self Organising Maps) olmak üzere üç yöntem kullanılmıştır. Haneleri gün içindeki elektrik kullanım modellerine göre kümelere ayırmak için en iyi performans gösteren tekniğin SOM olduğu bulunmuştur (McLoughlin, Duffy ve Conlon, 2015).

Konutlarda elektrik tüketimini tahmin etmek için hibrit bir model yaklaşımı öneren çalışmada; beş farklı makine öğrenme algoritması kullanılarak gerçek zamanlı ölçülen veriler ile hibrit model karşılaştırılmıştır. Önerilen hibrit modelin tahmin doğruluğunu %6–10 oranında iyileştirdiği vurgulanmıştır (Dong, Li, Rahman ve Vega, 2016).

Hanelerde enerji tüketimini kontrol altına almayı veya karbon yoğunluğunu azaltmayı amaçlayan politikalar tasarlamak için bir kümeleme yöntemi kullanılan çalışmada; farklı enerji tüketim düzeylerinin konut özelliklerinin yanında, sosyo-

ekonomik ve bölgesel özelliklerle ilişkili olduğu bulunmuştur. Ayrıca hanelerin enerji tüketiminin temel bileşenleri belirlenmiş ve hanelerin enerji tüketim tipolojisi oluşturulmuştur (Hache, Leboullenger ve Mignon, 2017). Enerji tüketimini tahmin etmek için dokuz makine öğrenmesi yöntemi kullanılan çalışmada; iki veya daha fazla tahmin tekniğini birleştiren hibrit modelin çeşitli kombinasyonlarının, enerji tahmininde daha etkili olduğu bulunmuştur (Deb, Zhang, Yang, Lee ve Shah, 2017).

Bina enerji tüketiminin tahmini için random forest ve yapay sinir ağları modellerinin karşılaştırıldığı bir çalışmada; yapay sinir ağlarının, random forest modelinden marjinal olarak daha iyi performans gösterdiği sunulmuştur. Ayrıca iki modelin de karşılaştırılabilir tahmin gücüne sahip olduğu ve bina enerji uygulamalarında neredeyse eşit derecede uygulanabilir olduğu belirtilmiştir (M. W. Ahmad, Mourshed ve Rezgui, 2017). Ticari binalarda enerji tüketimini tahmin etmek için makine öğrenmesi yöntemleri kullanılan çalışmada; gradyan artırmalı regresyon modellerinin diğer yöntemlerden daha iyi performans gösterdiği bulunmuştur (Robinson vd., 2017).

Makine öğrenme modelleri kullanılarak hanehalklarının kullanım kalıplarına bağlı enerji tüketimlerini tahmin eden çalışmada; hibrit modelin, tekli ve toplu modellerden daha doğru sonuçlar verdiği vurgulanmıştır (Chou ve Tran, 2018). Elektrik enerjisi tüketim tahmini için hibrit bir yapay sinir ağı kullanılan çalışmada; önerilen modelin tahmin doğruluğu açısından daha önceki yöntemlerden daha üstün performansa ulaştığı sunulmuştur (Li vd., 2018). Yük tahmini için farklı yaklaşımların doğruluğunu araştıran çalışmada; en uygun tahmin modelini aramak için ilk adımda, iki çeşit otoregresif hareketli ortalama modeli olan ARIMA ve SARIMA kullanılmıştır. İkinci adımda makine öğrenmesi ve derin öğrenme yöntemleri kullanılmış ve XGBOOST modeli en iyi performansı göstermiştir (Hadri vd., 2019).

Ev aletlerinin enerji tüketimini tahmin eden bir çalışmada; destek vektör makinesi (Support Vector Machine - SVM), k- en yakın komşu (K-Nearest Neighbor - KNN), aşırı rastgele orman (Extreme Random Forest - ERF), ve uzun-kısa süreli bellek ağı (Long Short-Term Memory Network - LSTM) olmak üzere dört tahmin modeli kullanılmıştır. Derin öğrenme yöntemi olan LSTM en iyi performansı göstermiştir (Wang vd., 2020). Enerji tüketimini tahmin etmek için bina enerji tüketiminin beş yıllık veri kümesi

kullanılan çalışmada; önerilen rastgele orman tabanlı bir tahmin modeli, iyi bir tahmin doğruluğu ortaya koymuştur (Pham, Ngo, Ha Truong, Huynh ve Truong, 2020).

Çoklu doğrusal regresyon, logaritmik çoklu doğrusal regresyon yöntemleri ve dışsal girdili doğrusal olmayan otoregresif yöntem olmak üzere üç farklı makine öğrenme yöntemini kullanan çalışmada; konut ve ticari sektörlerde enerji tüketimi model tahmini yapılmıştır. Dışsal girdili doğrusal olmayan otoregresif (Non-Linear Autoregressive with Exogenous Input - NARX) modelin sadece çoklu doğrusal regresyon ve logaritmik çoklu doğrusal regresyon yöntemlerinden daha iyi olmadığı, aynı zamanda literatürdeki mevcut yöntemlerin geri kalanıyla karşılaştırıldığında en iyi model olduğu gösterilmiştir (Nabavi vd., 2020).

Hem yeni bir regresyon tabanlı topluluk tahmin modeli hem de kümeleme yöntemi kullanılan çalışmada; akıllı evler için enerji kullanımı tahmin modelleri sunulmuştur. Elektrik talebini tahmin etmek için otomatik optimizasyona sahip yeni bir regresyon tabanlı topluluk tahmin modeli önerilmiştir (Hoque, Jamil, Saharudin ve Amin, 2020). Binalarda enerji tasarrufu sağlamak için uyarlanabilecek faktörleri araştıran bir çalışmada kümeleme, karar ağacı ve birliktelik kurallarının kullanıldığı bir yaklaşım önerilmiştir (Nazeriye vd., 2021).

Akıllı sayaç verilerini kullanarak hanehalkı profillerini belirlemek için yarı denetimli öğrenme yaklaşımı öneren çalışmada; önerilen yaklaşımın yalnızca sınırlı etiketli veri mevcut olduğu zaman denetimli öğrenme yöntemlerinden daha iyi performans gösterdiği belirlenmiştir (Wang vd., 2022). Veri boyutunun modelin performansı üzerindeki etkisinin araştırıldığı çalışmada; yıllık enerji tüketimini tahmin etmek için derin öğrenme, topluluk öğrenme ve makine öğrenme modelleri karşılaştırılmıştır. Dokuz modelin geliştirildiği çalışmada, derin öğrenme modelinin diğer modellerden daha verimli olduğu açıklanmıştır (Olu-Ajayi vd., 2022).

Literatürdeki çalışmalar incelendiğinde hanehalkı enerji profillerini inceleyen modellerin daha çok istatistiksel modeller olduğu görülmüştür. Makine öğrenmesi ile modellenen enerji çalışmalarının ise enerji tasarrufu, bina kullanımı ve akıllı sayaç sistemleri v.b ile ilgili olduğu tespit edilmiştir. Makine öğrenmesi yöntemlerini kullanan birçok performans tahmin modeli yayınlanmış olmasına rağmen, hanehalkı enerji

harcamalarını, makine öğrenmesi alanındaki son gelişmeler ışığında kapsamlı bir şekilde inceleyen çalışmalar yetersizdir.

Literatürde enerji ile ilgili yapılan çalışmalar incelendiğinde; hanehalkı enerji profillerini inceleyen çalışmalarda daha çok istatistiksel yaklaşımlar ve ekonometrik modeller kullanıldığı görülmüştür. Konu olarak; hanelerin toplam enerji harcamalarından ziyade, ‘mikro ve makro düzeyde enerji talep tahmini’, ‘bina enerji kullanımı’, ‘enerji verimliliğini artırma yolları’, ‘daha fazla enerjiyi nereden temin edebiliriz’ başlıkları üzerine yoğunlaşıldığı gözlemlenmektedir. Makine öğrenmesi ile modellenen enerji çalışmalarının ise enerji tasarrufu, bina kullanımı ve akıllı sayaç sistemleri v.b ile ilgili olduğu tespit edilmiştir. Makine öğrenmesi yöntemlerini kullanan birçok çalışma yayınlanmış olmasına rağmen, hanehalkı enerji harcamalarını, yapay zeka alanında hızla ilerleyen gelişmeler ışığında, yeni yaklaşımlar ve modellerle inceleyen çalışmalar hâlâ yetersizdir.

Çalışmamız hanehalkı enerji harcamalarını bir bütün olarak ele almıştır. Bu bağlamda, “ısınma ve mutfak harcama”, “aydınlatma harcama”, “yakıt harcama” değerlerinin toplanarak hanelerin “toplam enerji harcama” değerlerinin elde edilmesi, alana yeni bir bakış açısı kazandırmıştır. Hanehalkı “toplam” enerji harcamalarının topluluk öğrenmesi yöntemleri ile analizi literatüre yeni bir yaklaşım sunmuştur.

Bu çalışmada kullanılan değişkenlerin analizimizde yer alma gerekçeleri devam eden paragraflarda genel olarak açıklanmıştır. Sosyo - Demografik özellikler, konut özellikleri, enerji ile doğrudan ilişkili değişkenler ve enerji ile dolaylı olarak ilişkili değişkenler analizimizdeki hanehalkı profilini oluşturmaktadır.

Gelir, eğitim, cinsiyet, yaş, meslek gibi sosyoekonomik ve demografik özellikler hanelerin sosyal sınıfını ve yaşam tarzlarını belirlediği için hanehalkının enerji tüketim seçimlerinde önemli rollere sahip oldukları bilinmektedir (Salari ve Javid, 2017). Hanehalkı enerji harcamaları, kişi başına düşen gelir, hane reisinin eğitim durumu, hanehalkı üyelerinin mesleğinden etkilenmektedir (Narasimha Rao ve Reddy, 2007). Ayrıca, erkekler ve düşük gelirli aileler, yüksek gelirli ailelere göre önemli ölçüde daha yüksek enerji tasarrufu düzeyine sahiptir (Ding vd, 2017).

Evli haneler evli olmayan hanelere göre metrekare başına daha az enerji tüketmekte, yetişkin sayısı arttıkça metrekare başına düşen enerjinin de arttığını görülmektedir (Valenzuela vd., 2014). Bu nedenlerde, hanehalkı geliri, hanehalkı reisinin yaşı, medeni durumu, mesleği, eğitim durumu, hanedeki erkek ve kadın sayısı hanehalkı enerji harcamaları ile doğrudan ilgili olduğu için analizimize dahil edilen değişkenler arasında yer almıştır.

Konut özellikleri dikkate alındığında, hane büyüklüğü ve konutun mülkiyet durumu enerji tüketim davranışları üzerinde önemli bir etkiye sahiptir (Ding vd., 2017). Büyük konutlardaki haneler metrekare başına daha küçük konut birimlerine göre daha az enerji tüketir (Valenzuela vd., 2014). Ayrıca, konut tipi, bina yaşı ve hanenin büyüklüğü hane ısıtma enerjisinin en önemli belirleyicilerindendir (Rahut vd., 2016). Dolayısıyla, hanehalkı enerji harcamaları için hane büyüklüğü, konut tipi ve bina yaşı değişkenleri çalışmamızda yer almıştır.

Hanehalkı toplam enerji harcamaları ile doğrudan ilişkili değişkenler, ısıtma ve mutfak harcama, aydınlatma harcama ve araçlar için kullanılan yakıt harcama değişkenleridir. Burada hanelerde kullanılan doğalgaz, elektrik ve yakıt tüketimi değişkenleri detayları ile ele alınmıştır. Doğal gaz, küresel birincil enerji tüketiminde, petrolün ve kömürün ardından 3. sırada yer alırken Türkiye’de toplam nihai enerji tüketiminde doğal gaz; petrolün ardından 2. sırada yer almaktadır (TSKB, 2020).

Hanehalkı büyüklüğü ve bir hanede oturanların ortalama yaşı arttıkça ısıtma harcamaları da artmaktadır. Bu yüzden yaşlı insanlar genç olanlara kıyasla daha fazla konut enerjisi kullanırlar (Tonn ve Eisenberg, 2007). Ayrıca kadınlar erkeklere göre oda sıcaklığından daha az memnundur ve erkeklere göre kadınlar daha yüksek oda sıcaklıklarını tercih ederler (Karjalainen, 2007).

Türkiye toplam elektrik talebi 2000-2019 yılları arasında 2001, 2009 ve 2019 yılları haricinde bir düşüş göstermemiş ve artış trendini sürdürmüştür (TSKB, 2020). Bina sakinlerinin fazlalığı, binada gençlerin varlığı, konut elektrik tüketiminde önemli göstergelerdendir. Ayrıca, konut yaşının, oda sayısının, konutun alanının, evsel elektrik kullanımı üzerinde önemli bir pozitif etki gösterdiği yani evlerin yaşlandıkça ve büyüdükçe elektrik kullanımının arttığı bilinmektedir (Jones, Fuertes ve Lomas, 2015). Bunlara ek olarak, hane büyüklüğü, konut tipi, kurutucu kullanımı, yıkama döngüleri, duş

sayısı, aydınlatma ve cihaz kullanımı değişkenleri konutlardaki elektrik tüketimini önemli bir şekilde etkilemektedir (Bedir vd., 2013).

Gelişen kentleşme ve kentlerde yaşayan nüfusun artması, elektrikli araçların yaygınlaşması gibi nedenlere bağlı olarak nihai enerji tüketiminde elektriğin payının, önümüzdeki 30 yıl içinde önemli ölçüde artması beklenmektedir (TSKB, 2020). Hanelerdeki elektrik tüketimi iki ana faktör tarafından belirlenmektedir. İlki hanedeki elektrikli aletlerin tipi ve sayısı, ikincisi bu cihazların hane sakinleri tarafından kullanımındır. Hanehalkı sakinleri, hem elektrikli ev aletleri satın alarak hem de bu aletleri kullanarak bir hanenin elektrik kullanımını etkilemektedir (Firth, Lomas, Wright ve Wall, 2008). Bilgisayar, internet, buzdolabı, bulaşık makinesi, kurutma makinesi v.b.cihazların daha fazla kullanımı, konutlarda elektrik kullanımının artmasına daha fazla neden olmaktadır (Jones vd., 2015).

Şehirlerin sürekli gelişimi ve konut inşaatının büyümesiyle, konut binalarının enerji tüketimi son yıllarda artış göstermektedir. Konut sakinlerinin ekonomik düzeyindeki artış, ev aletleri alımında artışa yol açmaktadır. Ayrıca konut binalarının enerji tüketiminin %57'si ev aletleri tarafından tüketilmektedir (Wang vd., 2020). Hanehalkı toplam enerji harcama ana değişkenlerimizden biri olan elektrik tüketiminin hanedeki elektrikli alet sayısı ile doğrudan ilgili olduğu görülmüş ve bu sebeple elektrikli alet sayısı değişkeni analizimize eklenmiştir.

Petrol, Türkiye'nin hem enerji ürünleri arzı hem de nihai enerji tüketiminde önemli bir yer tutmaktadır. Türkiye enerji ürünleri arzının ortalama %30'u, nihai enerji tüketiminin ise ortalama %35'i petrol ve petrol ürünlerinden oluşmaktadır (TSKB, 2020). Hem kırsal hem de kentsel bölgelerde; aylık hanehalkı harcaması, hanehalkı büyüklüğü ve eğitim değişkenleri yakıt tercihlerinin belirlenmesinde kritik bir rol oynamaktadır. Hanehalkı üyelerinin eğitim durumu arttıkça haneler modern yakıtları tercih etmektedir. Kırsal haneler yakacak oduna bağımlı olup, buna karşın kentsel haneler modern yakıtları tercih etmektedir. Bu durumlara ek olarak, cinsiyet yakıt seçiminde bir fark oluşturmakta; kadınların reisi olduğu haneler genellikle hem kentsel hem de kırsal alanlarda modern yakıtları tercih etmektedir (Narasimha Rao ve Reddy, 2007).

Hanehalkı toplam enerji harcama ana değişkenlerimizden olan yakıt harcamalarının konutun tipi, hanehalkı eğitim durumu, hanedeki yaşlı, erkek ve kadın sayısı ile doğrudan

ilgili olduğu görülmüş ve bu sebeple bahsi geçen değişkenlere analizimizde yer verilmiştir.

TÜİK'in yayınladığı Hanehalkı Bütçe Araştırması anketinde kır kent ayrımının verilmemiş olması çalışmanın sınırlayıcılarından biridir. Bu önemli eksikliği gidermek açısından tarla, bahçe ve sera sahipliği değişkenleri birleştirilerek tek bir değişken haline getirilmiştir. Bu değişken, kır kent ayrımında olduğu gibi kesin bir bilgi sunmasada tarla, bahçe ve sera sahiplerinin büyük bir kısmının kırdaki yaşadığı varsayımı ile kır kent ayrımını belli ölçüde temsil ettiği düşünülmüştür. Ayrıca kamu ve bazı zorunlu hizmetlere ulaşım güçlüğü nedeniyle dezavantajlı konumda olan hanehalklarının ekstra enerji kullanımı ile ilgili olduğu için tarla-bahçe-sera sahipliği, günlük alışveriş hizmetlerine ulaşım zorluğu, bankacılık hizmetlerine ulaşım zorluğu, zorunlu eğitim hizmetlerine ulaşım zorluğu gibi değişkenler analizimize dahil edilmiştir.

Kısaca; hanehalkı enerji tüketim literatürü gözden geçirilerek, sosyo-demografik özellikler, konut özellikleri, enerji ile doğrudan ilişkili değişkenler ve enerji ile dolaylı olarak ilişkili olan değişkenler olmak üzere hanehalkı profilleri dört ana başlık halinde toplanmıştır. Hanehalkı profilleri Şekil 14'te gösterilmiştir.



Şekil 14. Hanehalkı Profilleri

3.2. ARAŞTIRMANIN VERİ SETİ

Çalışmada, Türkiye İstatistik Kurumu (TÜİK) tarafından yayınlanan Hanehalkı Bütçe Araştırması anketinin mikro veri seti kullanılmıştır. Örneklem yöntemi olarak tabakalı iki aşamalı küme örnekleme yöntemi kullanılmıştır. Hanehalkı Bütçe Araştırması anketinin örnekleme çerçevesi dahilinde, birinci aşama örnekleme birimi olan blokların seçiminde kullanılan temel örnekleme çerçevesi, Ulusal Adres Veri Tabanıdır. Bu çerçeve kullanılarak bloklar oluşturulmuş, kentsel ve kırsal alanlardan yerleşim yerinin büyüklüğü ile orantılı olasılıkla bloklar belirlenmiş ve her blokta hanehalkları sistematik olarak seçilmiştir (TÜİK, 2020).

3.3. ARAŞTIRMADA KULLANILAN YÖNTEMLER

Çalışmada, hanehalkı toplam enerji harcamalarını tahmin etmek için makine öğrenmesi yöntemlerine ek olarak topluluk öğrenme yöntemleri uygulanmıştır. Model tahmininde; K-En Yakın Komşu (K-Nearest Neighbor), Naive Bayes, Karar ağacı (Decision Tree), Derin Sinir Ağları (Deep Neural Networks), Random Forest, Gradyan Artırmalı Ağaç (Gradient Boosted Trees) sınıflandırıcıları uygulanmıştır. Modelde sınıflandırıcıların bireysel performanslarının yanı sıra bagging (torbalama), boosting (artırma), voting (oylama) ve stacking (istifleme) topluluk öğrenme yöntemlerinin performansları da değerlendirilmiştir. Verileri sınıflandırmak ve topluluk yöntemlerinin performansını değerlendirmek için uygulanan veri ön işleme, öznitellik seçimi, modelleme, tahmin ve performans değerlendirme aşamaları için RapidMiner (RapidMiner Studio 9.10.10) açık kaynak kodlu yazılım programı kullanılmıştır.

3.4. ARAŞTIRMADA KULLANILAN DEĞİŞKENLER

Bu çalışmada, TÜİK tarafından yayınlanan 2019 yılı Hanehalkı Bütçe Araştırması anketinde yer alan hanehalkı tüketim verileri kullanılmıştır. Tüketim verileri içinde halihazırda var olan değişkenlerin yanında, literatür çerçevesinde ve uzman görüşü rehberliğinde yeni değişkenler de oluşturulmuştur. Değişkenlerin nasıl seçildiği ve oluşturulduğu çalışmanın üçüncü bölümünde detaylı bir şekilde açıklanmıştır. Hanehalkı toplam enerji harcamalarının sınıflandırılması için kullanılan değişkenler Tablo 2'de gösterilmiştir.

Tablo 2. Değişkenlerin Listesi ve Tanımı

Değişkenler	Kategoriler
Toplam Enerji Harcama	Aydınlatma Harcama + Isınma ve Mutfak Harcama + Yakıt Harcama (Araç)
Aydınlatma Harcama	Pozitif Reel Sayı (TL)
Isınma ve Mutfak Harcama	
Yakıt Harcama (Araç)	
Yıllık Kullanılabilir Gelir	
Elektrik Harcaması	
Hanehalkı Harcama Değeri	
Konutun Tipi	1 Müstakil Konut 2 Apartman Dairesi
Mülkiyet Durumu	1 Ev Sahibi 2 Kiracı 3 Lojman 4 Ev Sahibi Değil ama Kira Ödemiyor
Binanın İnşa Tarihi	1 1961-1980 2 1981-2000 3 2001 ve sonrası 4 1960 ve öncesi
Oda Sayısı	(Salon dahil)
Konutun Alanı	m ²
İkinci Konut Durumu	1 Var 2 Yok
Doğal Gaz Sahipliği	
Tarla Bahçe Sera Sahipliği	
Cinsiyet	1 Erkek 2 Kadın
Yaş	Bitirdiği Yaş
Eğitim Durumu	1 Bir Okul Bitirmesi-İlkokul 2 Ortaokul 3 Lise 4 Yüksek Okul, Lisans- Lisansüstü
Medeni Durum	1 Hiç Evlenmedi 2 Evli 3 Boşandı-Eşi öldü
Çalışma Durumu	1 Çalışıyor 2 Çalışmıyor

Tablo 2. (Devamı)

Meslek	1 Yöneticiler 2 Profesyonel Meslek Mensupları 3 Teknisyenler, Teknikerler ve Yardımcı Profesyonel Meslek Mensupları 4 Büro Hizmetlerinde Çalışan Elemanlar 5 Hizmet ve Satış Elemanları 6 Nitelikli Tarım, Ormancılık ve Su Ürünleri Çalışanları 7 Sanatkarlar ve İlgili İşlerde Çalışanlar 8 Tesis ve Makine Operatörleri ve Montajcılar 9 Nitelik Gerektirmeyen İşlerde Çalışanlar 10 Çalışmıyor
Hanedeki Kadın Oranı (Hanedeki Kadın Sayısı / Hanehalkı Büyüklüğü)	Pozitif Reel Sayı
Harcama Oranı (Hanehalkı Harcama Değeri / Toplam Enerji Harcama)	
Hanede Okuyan Sayısı	Pozitif Tam Sayı
Hanedeki Çalışan Sayısı	
Hanedeki Çocuk Sayısı	
Hanedeki Yaşlı Sayısı	
Hanehalkı Büyüklüğü	
Otomobil Sayısı	
Hanedeki Elektrikli Alet Sayısı	
Günlük Alışveriş Hizmetlerine Ulaşım Zorluğu	1. Çok Kolay 2. Kolay 3. Orta 4. Zor 5. Çok Zor
Toplu Ulaşım Hizmetlerine Ulaşım Zorluğu	
Bankacılık Hizmetlerine Ulaşım Zorluğu	
Sağlık Hizmetlerine Ulaşım Zorluğu	
Posta Hizmetlerine Ulaşım Zorluğu	
Zorunlu Eğitim Hizmetlerine Ulaşım Zorluğu	

3.5. ARAŞTIRMA VERİLERİNE CRISP-DM METODOLOJİSİNİN UYGULANMASI

Bu çalışmada, sosyal bilimlerde yaygın olarak kullanılan CRISP-DM (Cross Industry Standard Process for Data Mining) metodolojisi kullanılmıştır. Bu metodoloji birinci bölümde bahsedildiği gibi altı aşamadan oluşmaktadır.

3.5.1. İş Süreçlerinin Anlaşılması (Business Understanding)

İlk aşamada, toplam hanehalkı enerji harcamalarının sınıflandırılması için yüksek performanslı bir topluluk öğrenme modelinin oluşturulması hedeflenmiştir.

3.5.2. Verinin Anlaşılması (Data Understanding)

İkinci aşamada, 2019 yılında TÜİK'in yayınladığı Hanehalkı Bütçe Araştırması anket verilerinden elde edilen 11.521 hanenin veri seti incelenmiştir. Veri setinde hanehalkı tüketimine ilişkin bilgileri içeren üç farklı veri dosyası olduğu görülmüştür. Bunlar; bireysel özellikler (cinsiyet, eğitim düzeyi, medeni durum vb.), konut özellikleri (konut tipi, bina inşa yılı, oda sayısı vb.), tüketim özellikleri (tüketilen mal ve hizmetlerle ilgili değerler)'dir. Bu veri seti kullanılarak değişkenler, alan bilgisi ve uzman görüşü rehberliğinde yapılan literatür taraması sonucunda çalışmanın amacına uygun olarak birleştirilmiştir. Veri setinde halihazırda var olan değişkenlere ek olarak toplam hanehalkı enerji harcama değerleri ile ilgili yeni değişkenler oluşturulmuştur.

3.5.3. Veri Ön İşleme (Preprocessing)

Üçüncü aşamada veriler ön işleme aşamalarından geçirilmiştir. Burada ikinci aşamada belirlenen verilere, filtreleme, düzenleme, yapılandırma, öznitelik seçimi v.b. işlemler yapılarak veriler modelleme işlemine hazır hale getirilmiştir. Bu anlamda ilk olarak “ısınma ve mutfak harcama”, “aydınlatma harcama”, “yakıt harcama” değerleri toplanarak “hanehalkı toplam enerji harcama” değerleri elde edilmiştir. Bazı değişkenlerin çalışmamızın amacı ve analizi üzerinde anlamlı bir etkisi olmamıştır. Bu nedenle analiz için anlamsız olan hanehalkı toplam enerji harcama değerleri, ısınma ve mutfak harcama değerleri ve aydınlatma harcama değerleri içerisinde sıfıra eşit olan haneler, sırasıyla veri setinden çıkarılmıştır. Ayrıca, otomobil sahibi olup yakıt harcama değeri sıfıra eşit olan haneler, analizimiz için anlamsız olup veri setinden çıkarılmıştır.

Devam eden süreçte aykırı (uç) değerleri bulmak için yoğunluğa dayalı bir aykırı değer hesaplama yöntemi olan lokal aykırı faktör (LOF- Local Outlier Faktör) yöntemi kullanılmıştır. LOF, bir nesnenin kendisini çevreleyen komşuluğa göre ne kadar izole olduğunun derecesi olarak sayısal bir ölçekte aykırılık sağlayarak yerel aykırı değerleri

tespit için geliştirilmiş bir yöntemdir (Breunig, Kriegel, Ng ve Sander, 2000; Lee vd., 2011). Değerlendirme sonucunda aykırı değerler veri setinden çıkarılmıştır. Yapılan uygulamalar sonucunda analizimiz için hazır halde 8.494 haneye ait temiz bir veri seti elde edilmiştir.

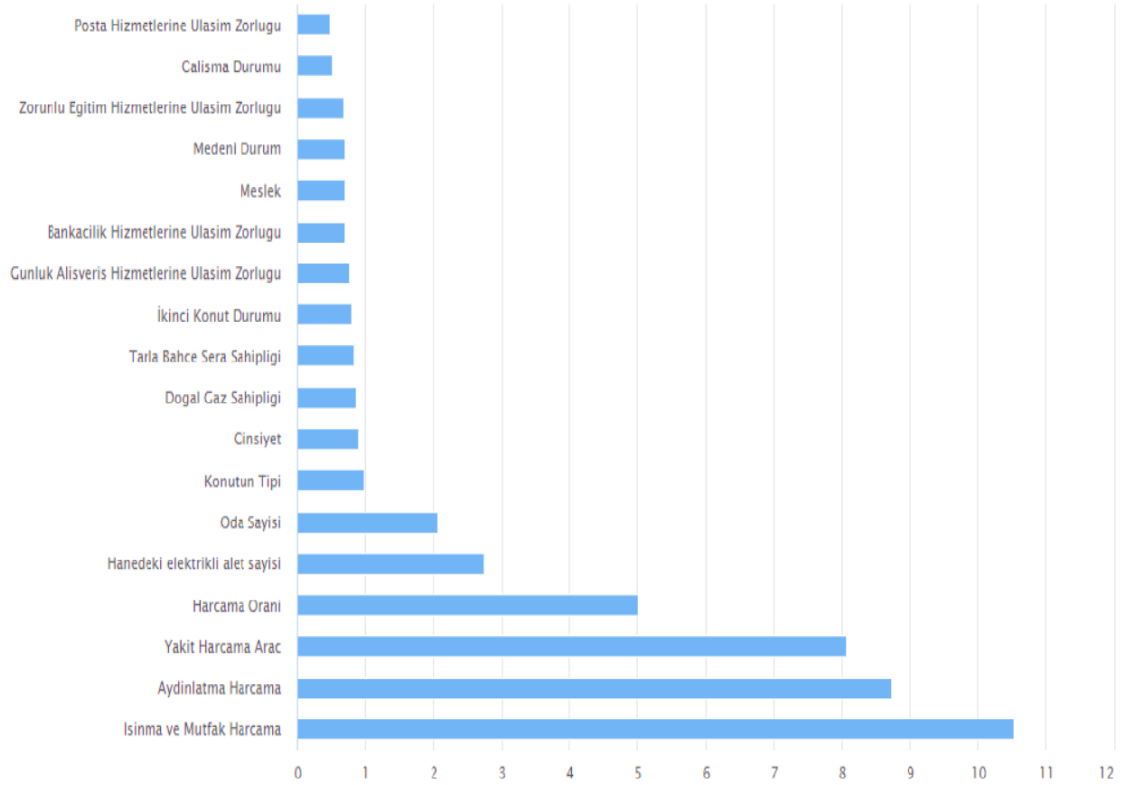
“Toplam enerji harcama” değişkeni “hedef değişken” olarak belirlenmiştir. Literatür taraması sonucunda oluşturulan Tablo 2’deki değişkenlerden; toplam enerji harcama değişkeni dışındaki değişkenler “açıklayıcı değişkenler” olarak tespit edilmiştir.

Daha sonra literatürde verileri farklı etiketlendirme şekilleri araştırılmıştır. Türkiye'nin işgücü yapısının incelendiği çalışmada (Ekmekçi, 2019), karar ağacı modelleri ile bir sınıflandırma yapılmış ve hedef değişken (iş arama süresi) iki sınıf olarak etiketlenmiştir. Kitap satış listelerinin KNN ile sınıflandırıldığı çalışmada (Denizli, 2019) hedef değişken (kitap türü) iki sınıf olarak etiketlenmiştir. Konutlarda elektrik tüketim düzeylerinin yapay sinir ağları ile tahmin edildiği çalışmada (Ercan, Irmak, Çevik ve Canbazoglu, 2020) hedef değişken (elektrik tüketim düzeyi) iki sınıf olarak etiketlenmiştir. Türkiye'deki otomobil talebinin tahmin edildiği çalışmada (Alper, 2000) hedef değişken (araba modelleri) çeyreklere (kartillere) ayrılarak etiketlenmiştir. Bu çalışmalardan yola çıkarak çalışmamızdaki hedef değişken verileri, çeyreklerine (kartillerine) göre üst, orta üst, orta alt ve alt olmak üzere dört sınıfa ayrılmıştır. Sonuçta, hedef değişkenimiz olan toplam enerji harcaması 4 sınıflı değişken olarak etiketlenmiştir. Veri ön işleme aşamaları neticesinde, bir hedef değişken ve 35 açıklayıcı değişken elde edilmiştir.

Belirlediğimiz 35 değişken, modelleme sürecinde sınıflandırma performansına katkılarına göre öznitelik seçme yöntemleri kullanılarak elenmiştir. Bu çalışmada, İleri Seçim (Forward Selection), Geriye Eleme (Backward Elimination), Optimize Edilmiş Seçim (Optimize Selection), Optimize Edilmiş Seçim –Evrimsel (Optimize Selection-Evolutionary) öznitelik seçim yöntemleri denenmiştir. Model performans sonucuna en fazla katkıyı sağlayan Optimize Edilmiş Seçim (Evrimsel) yöntemi kullanılarak analize devam edilmiştir. Bu evrimsel yöntemde bir genetik algoritma kullanılmaktadır. Genetik algoritma, doğal evrim sürecini taklit eden bir buluşsal arama algoritmasıdır (Panthong ve Srivihok, 2015).

Çalışmamızda, öznitelik seçiminden sonra 35 açıklayıcı değişken içerisinde sınıflandırma performansına daha fazla katkı sağlayan 18 değişken ortaya çıkmıştır. Sonuçlar; hanahalkı toplam enerji harcamalarını etkileyen ilk üç faktörünün, ısınma ve mutfak harcama, aydınlatma harcama, yakıt harcama (araç) değişkenleri olduğunu göstermiştir. Bu anlamda, bahsi geçen üç değişkenin çalışmamızın özgün değeri olan “hanehalkı toplam enerji harcamalarını” oluşturması neticesinde yapılan analizin tutarlı olduğu keşfedilmiştir. Ayrıca sonraki önemli değişkenlerin, harcama oranı, hanedeki elektrikli alet sayısı ve oda sayısı olduğu tespit edilmiştir.

Konutun tipi, cinsiyet, doğal gaz sahipliği, tarla-bahçe-sera sahipliği, ikinci konut durumu, günlük alışveriş hizmetlerine ulaşım zorluğu, bankacılık hizmetlerine ulaşım zorluğu, meslek, medeni durum, zorunlu eğitim hizmetlerine ulaşım zorluğu, çalışma durumu değişkenlerin hemen hemen eşit derecede hanehalkı toplam enerji harcamalarına katkıda bulunduğu görülmüştür. En az etkinin ise, çalışma durumu ve posta hizmetlerine ulaşım zorluğu değişkenleri olduğu belirlenmiştir. Öznitelik seçimi sonrasında değişkenlerin RapidMiner çıktısı Şekil 15’te gösterilmiştir.



Şekil 15. Öznitelik Seçimi

3.5.4. Modelleme (Modelling)

Modelleme aşamasında, hanehalkının toplam enerji harcamasını tahmin etmek için makine öğrenmesi yöntemlerine ek olarak topluluk öğrenme yöntemleri de uygulanmıştır. Modeldeki verileri sınıflandırmak için K-En Yakın Komşu (K-Nearest Neighbor - KNN), Naive Bayes (NB), Karar ağacı (Decision Tree - DT), Derin Sinir Ağları (Deep Neural Networks), Random Forest (RF), Gradyan Artırmalı Ağaç (Gradient Boosted Trees - GBT) sınıflandırıcıları uygulanmıştır. Sonrasında bagging (torbalama), boosting (artırma), voting (oylama) ve stacking (istifleme) topluluk öğrenme yöntemleri kullanılarak birleştirilen modellerden en yüksek performansa sahip model belirlenmiştir.

Modeldeki verilerin eğitim ve test verisi olarak ayrılması işleminde 10-katlı çapraz geçerlilik yöntemi uygulanmıştır. 10-katlı çapraz geçerlilik yönteminde veri seti 10 parçaya ayrılır. Bu parçalardan her biri sırayla, test veri seti ve geri kalan dokuz parça eğitim veri seti olarak kullanılır. Bu işlem, her bir parça test veri seti olana kadar 10 kez tekrar edilir. Doğruluk değerlerinin ortalaması alınarak modelin doğruluk oranı belirlenir. 10-katlı çapraz geçerlilik yöntemi Şekil 16'da ifade edilmiştir.

	1. Parça	2. Parça	3. Parça	4. Parça	5. Parça	6. Parça	7. Parça	8. Parça	9. Parça	10. Parça
1. Adım	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
2. Adım	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
3. Adım	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
4. Adım	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
5. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
6. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim
7. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim
8. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim
9. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim
10. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test

Şekil 16. 10-Katlı Çapraz Geçerlilik Yöntemi

3.5.5. Değerlendirme (Evaluation)

Beşinci aşamada, farklı makine öğrenmesi yöntemleri kullanılarak analiz edilen ve topluluk öğrenme modelleri ile birleştirilen modellerin sınıflandırma performans sonuçları elde edilmiştir. Sınıflandırıcıların ve topluluk modellerinin performansını değerlendirmek için kullanılan metrikler için karışıklık matrisi (confusion matrix) Tablo 3'te verilmiştir. Sınıflandırıcıların ve topluluk yöntemlerinin performansları doğruluk, kappa katsayısı, duyarlılık ağırlıklı ortalama, kesinlik ağırlıklı ortalama metrikleri ile değerlendirilmiştir.

Tablo 3. Karışıklık Matrisi

		Gerçek Değerler		
		Pozitif Doğru Pozitif (DP)	Negatif Yanlış Pozitif (YP)	Toplam DP+YN
Tahmin	Pozitif			
Değerleri	Negatif	Yanlış Negatif (YN)	Doğru Negatif (DN)	YP+DN
	Toplam	DP+YN	YP+DN	DP+YP+YN+DN

$$\text{Doğruluk} = \frac{DP + DN}{DP + YP + YN + DN}$$

$$\text{Kappa} = \frac{N * (DP + DN) - X}{N^2 - X}$$

$X = (DP + YN) * (DP + YP) * (DN + YN)$, N, kullanılan toplam örnek sayısıdır.

$$\text{Duyarlılık} = \frac{DP}{DP + YP}$$

$$\text{Kesinlik} = \frac{DP}{DP + YN}$$

Doğru Pozitifler (DP): Veri noktalarının gerçek sınıfı “Doğru” olduğunda, tahmin edilen sınıfın da “Doğru” olması durumudur. Eğer bir kişi gerçekte sağlıklı ise; kişinin incelemeler sonucunda da sağlıklı çıkması durumudur.

Doğru Negatifler (DN): Veri noktalarının gerçek sınıfı “Yanlış” olduğunda, tahmin edilen sınıfın da “Yanlış” olması durumudur. Eğer bir kişi gerçekte hasta ise; kişinin incelemeler sonucunda da hasta çıkması durumudur.

Yanlış Pozitifler (YP): Veri noktalarının gerçek sınıfı “Doğru” olduğunda, tahmin edilen sınıfın “Yanlış” olması durumudur. Eğer bir kişi gerçekte sağlıklı ise; kişinin incelemeler sonucunda hasta çıkması durumudur.

Yanlış Negatifler (YN): Veri noktalarının gerçek sınıfı “Yanlış” olduğunda, tahmin edilen sınıfın “Doğru” olması durumudur. Eğer bir kişi gerçekte hasta ise; kişinin incelemeler sonucunda sağlıklı çıkması durumudur (Gürsaka, 2018).

3.5.5.1. Performans Değerlendirme Metrikleri

3.5.5.1.1. Doğruluk (Accuracy)

Doğruluk değeri, ölçümün gerçek değere yakınlığıdır. Ayrıca, doğru sınıflandırılmış birimlerin yüzdesini gösterir. Doğruluk değeri, bir modelin performansını ölçmek için kullanılan en yaygın ölçü birimidir.

3.5.5.1.2. Kappa Katsayısı

Kappa katsayısı, sınıflandırma modelinin sınıflandırma başarısı hakkında bilgi verir. Kappa değeri -1 ile +1 arasında bir değer alır. Kappa değer aralıkları aşağıdaki gibi yorumlanmaktadır:

- <0 : Hiç uyum yok
- 0 - 0,20: Önemsiz uyum
- 0,21 - 0,40: Orta derecede uyum
- 0,41 - 0,60: Makul uyum
- 0,61 - 0,80: Önemli uyum
- 0,81 - 1: Neredeyse mükemmel uyum (Özdemir, 2020)

3.5.5.1.3. Duyarlılık (Sensitivity, recall)

Duyarlılık, sınıflandırma modeli tarafından doğru tahmin edilen pozitif örneklerin oranını ifade etmektedir. Gerçekte pozitif bir sonuç elde edildiğinde, test sonucunun da pozitif olması olasılığı anlamına gelir (Özdemir, 2020).

3.5.5.1.4. Kesinlik (Precision)

Kesinlik gerçek pozitif değerlere ilişkin, sınıflandırma modeli tarafından tahmin edilmiş pozitif değerlerin yüzdesini göstermektedir (Özdemir, 2020).

3.5.5.1.5. Duyarlılık Ağırlıklı Ortalama (Weighted_mean_recall)

Duyarlılık ağırlıklı ortalama, duyarlılık ölçümlerinin ağırlıklı ortalamasını ifade eder. İkiiden çok sınıflı tahminler için her bir sınıfın duyarlılık değerinin sınıf sayısına bölünmesi yoluyla hesaplanır.

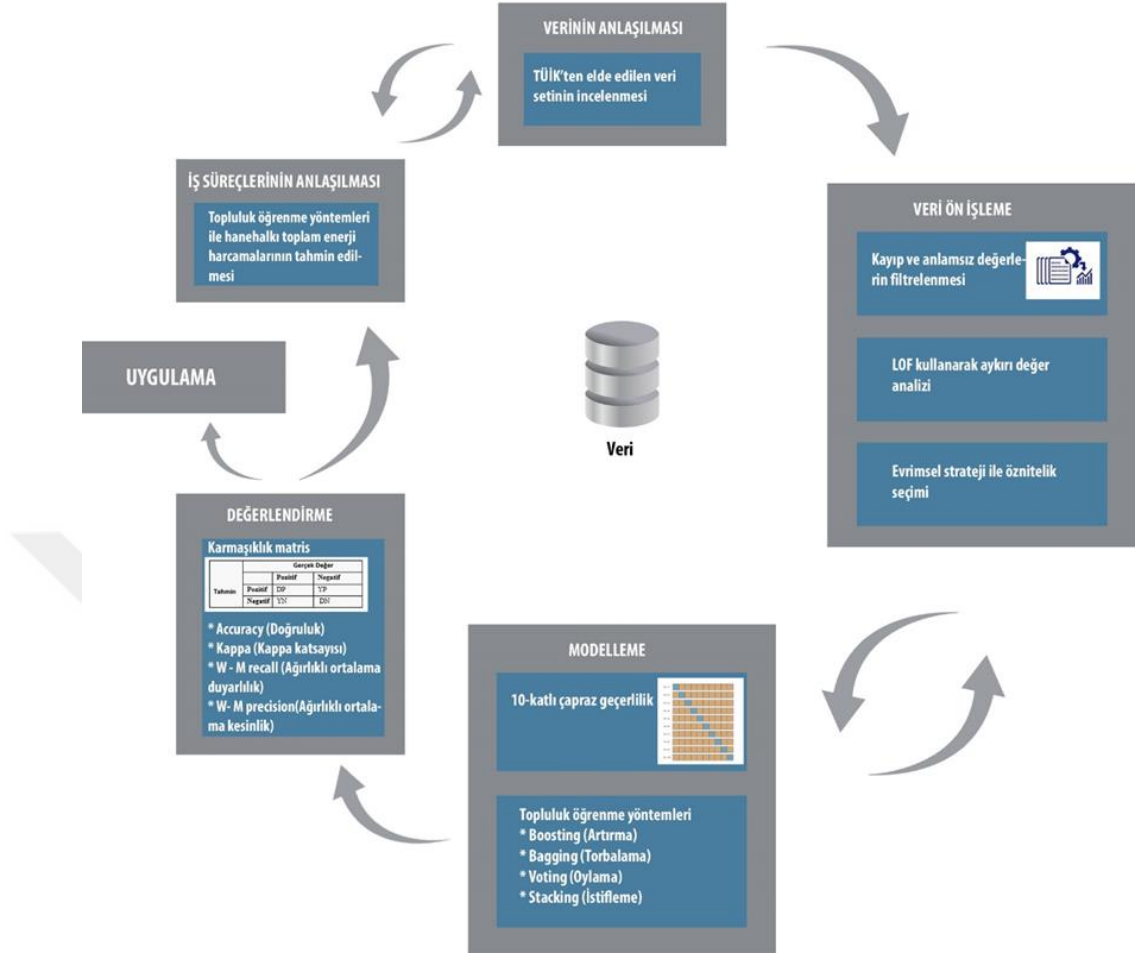
3.5.5.1.6. Kesinlik Ağırlıklı Ortalama (Weighted_mean_precision)

Kesinlik ağırlıklı ortalama, kesinlik ölçümlerinin ağırlıklı ortalamasını ifade eder. İkiiden çok sınıflı tahminler için her bir sınıfın kesinlik değerinin toplanıp sınıf sayısına bölünmesi yoluyla hesaplanır.

3.5.6. Uygulama (Deployment)

Uygulama aşamasında, bir önceki aşamada performans metrikleri ile değerlendirilen topluluk öğrenme model sonuçlarının yüksek performansa sahip olduğu keşfedilmiştir. Keşfedilen bilgi ile birlikte CRISP –DM döngüsü tamamlanmıştır. Bu çalışmada, topluluk öğrenme yöntemleri ile hanehalkı toplam enerji harcamalarının sınıflandırılmasının uygunluğu araştırılmış ve istikrarlı bir sonuç bulmuştur. Bu çalışma başlangıçtaki hedefine ulaşmıştır.

Bu çalışmanın CRISP- DM sürecinin tüm aşamaları Şekil 17'de özet olarak sunulmuştur.



Şekil 17. CRISP-DM Uygulama Süreci

3.6. BULGULAR

Bu çalışmada, hane halkı toplam enerji harcaması, makine öğrenmesi ve topluluk öğrenme yöntemleri kullanılarak sınıflandırılmıştır. TÜİK tarafından yayınlanan Hane halkı Bütçe Araştırması anketi 2019 veri setinden elde edilen 11.521 hanenin tüketim verileri kullanılmıştır. Bu çalışmanın amacı, en iyi sınıflandırma performansına sahip topluluk öğrenme yöntemini bulmaktır. Bu amaçla veriler ön işleme aşamalarından geçtikten sonra modellemeye hazır hale getirilmiştir.

Modelleme aşamasında, öncelikle topluluk öğrenme yöntemleri kullanılmadan KNN, Naive Bayes, Karar ağacı, Derin Sinir Ağları, Random Forest, Gradyan Artırmalı Ağaç modelleri bireysel olarak çalıştırılmıştır. Daha sonra KNN, Naive Bayes, Karar ağacı, Derin Sinir Ağları, Random Forest, Gradyan Artırmalı Ağaç sınıflandırıcıları, tüm

topluluk öğrenme yöntemleri ile tek tek çalıştırılmıştır. Ayrıca KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest, Gradyan Artırmalı Ağaç sınıflandırıcılarının birleşmesinden oluşan ALL modeli, voting ve stacking topluluk öğrenme yöntemleri ile çalıştırılmıştır.

Yöntemlerin performansları doğruluk, kappa katsayısı, duyarlılık ağırlıklı ortalama, kesinlik ağırlıklı ortalama metrikleri açısından karşılaştırılmıştır. Makine öğrenmesi ve topluluk öğrenme yöntemlerinin hanehalkı toplam enerji harcaması sınıflandırma performansları Tablolar ile gösterilmiştir.

3.6.1. Makine Öğrenmesi Sınıflandırıcılarının Performanslarının Değerlendirilmesi

Hanehalkı toplam enerji harcamalarının tahmin edilmesi için öncelikle sınıflandırıcıların bireysel makine öğrenmesi model performansları sunulmuştur. KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç olmak üzere en fazla performans gösteren altı farklı sınıflandırıcı, modele dahil edilmiştir.

Tablo 4'te; KNN sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. KNN sınıflayıcısının doğruluk değeri 0.977, kappa katsayısı 0.969, duyarlılık ağırlıklı ortalama değeri 0.977 ve kesinlik ağırlıklı ortalama değeri 0.977 olarak elde edilmiştir.

Tablo 4. Knn Sınıflandırıcısının Performans Değerlendirmesi

Performans Ölçütü	Değer
Doğruluk	0.977
Kappa katsayısı	0.969
Duyarlılık ağırlıklı ortalama	0.977
Kesinlik ağırlıklı ortalama	0.977

Tablo 5'te; Naive Bayes sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Naive Bayes sınıflayıcısının doğruluk değeri 0.785, kappa katsayısı 0.713, duyarlılık ağırlıklı ortalama değeri 0.784 ve kesinlik ağırlıklı ortalama değeri 0.791 olarak elde edilmiştir.

Tablo 5. Naive Bayes Sınıflandırıcısının Performans Değerlendirmesi

Performans Ölçütü	Değer
Doğruluk	0.785
Kappa katsayısı	0.713
Duyarlılık ağırlıklı ortalama	0.784
Kesinlik ağırlıklı ortalama	0.791

Tablo 6’da; Karar Ağacı sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Karar ağacı sınıflayıcısının doğruluk değeri 0.775, kappa katsayısı 0.700, duyarlılık ağırlıklı ortalama değeri 0.774 ve kesinlik ağırlıklı ortalama değeri 0.787 olarak elde edilmiştir.

Tablo 6. Karar Ağacı Sınıflandırıcısının Performans Değerlendirmesi

Performans Ölçütü	Değer
Doğruluk	0.775
Kappa katsayısı	0.700
Duyarlılık ağırlıklı ortalama	0.774
Kesinlik ağırlıklı ortalama	0.787

Tablo 7’de; Derin Sinir Ağları sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Derin Sinir Ağları sınıflayıcısının doğruluk değeri 0.973, kappa katsayısı 0.964, duyarlılık ağırlıklı ortalama değeri 0.973 ve kesinlik ağırlıklı ortalama değeri 0.973 olarak elde edilmiştir.

Tablo 7. Derin Sinir Ağları Sınıflandırıcısının Performans Değerlendirmesi

Performans Ölçütü	Değer
Doğruluk	0.973
Kappa katsayısı	0.964
Duyarlılık ağırlıklı ortalama	0.973
Kesinlik ağırlıklı ortalama	0.973

Tablo 8’de; Random Forest sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Random Forest sınıflayıcısının doğruluk değeri 0.802, kappa katsayısı 0.736, duyarlılık ağırlıklı ortalama değeri 0.802 ve kesinlik ağırlıklı ortalama değeri 0.805 olarak elde edilmiştir.

Tablo 8. Random Forest Sınıflandırıcısının Performans Değerlendirmesi

Performans Ölçütü	Değer
Doğruluk	0.802
Kappa katsayısı	0.736
Duyarlılık ağırlıklı ortalama	0.802
Kesinlik ağırlıklı ortalama	0.805

Tablo 9’da; Gradyan Artırmalı Ağaç sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Gradyan Artırmalı Ağaç sınıflayıcısının doğruluk değeri 0.917, kappa katsayısı 0.893, duyarlılık ağırlıklı ortalama değeri 0.919 ve kesinlik ağırlıklı ortalama değeri 0.920 olarak elde edilmiştir.

Tablo 9. Gradyan Artırmalı Ağaç Sınıflandırıcısının Performans Değerlendirmesi

Performans Ölçütü	Değer
Doğruluk	0.917
Kappa katsayısı	0.893
Duyarlılık ağırlıklı ortalama	0.919
Kesinlik ağırlıklı ortalama	0.920

Çalışmanın bulgularına göre, bireysel makine öğrenmesi yönteminin sınıflandırıcıları arasında en yüksek performansı KNN sınıflandırıcısı elde etmiştir. Naive Bayes ve Karar Ağacı sınıflandırıcısı, topluluk yöntemine tabi tutulmayan sınıflandırıcılar arasında diğer modellere göre daha düşük performans elde etmiştir. Makine öğrenmesi yöntemlerinin sınıflandırma performanslarının karşılaştırılması Tablo 10’da gösterilmiştir.

Tablo 10. Makine Öğrenmesi Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması

		Performans Değerlendirmesi			
		Doğruluk	Kappa katsayısı	Duyarlılık ağırlıklı ortalama	Kesinlik ağırlıklı ortalama
Makine Öğrenmesi Yöntemleri	KNN	0.977	0.969	0.977	0.977
	Naive Bayes	0.785	0.713	0.784	0.791
	Karar Ağaçları	0.775	0.700	0.774	0.787
	Derin Sinir Ağları	0.973	0.964	0.973	0.973
	Random Forest	0.802	0.736	0.802	0.805
	Gradyan Artırmalı Ağaç	0.917	0.893	0.919	0.920

3.6.2. Boosting Topluluk Öğrenme Yöntemi ile Sınıflandırıcıların Performanslarının Değerlendirilmesi

Hanehalkı toplam enerji harcamalarının tahmin edilmesi için boosting topluluk öğrenme yönteminin sınıflandırma performansları sunulmuştur. KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç olmak üzere en fazla performans gösteren altı farklı sınıflandırıcı, modele tek tek dahil edilmiştir.

Tablo 11’de; boosting topluluk öğrenme yöntemi ile KNN sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Boosting topluluk öğrenme yönteminde KNN sınıflayıcısının doğruluk değeri 0.977, kappa katsayısı 0.969, duyarlılık ağırlıklı ortalama değeri 0.977 ve kesinlik ağırlıklı ortalama değeri 0.977 olarak elde edilmiştir.

Tablo 11. Boosting Topluluk Öğrenme Yöntemi İle KNN Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.977
Kappa katsayısı	0.969
Duyarlılık ağırlıklı ortalama	0.977
Kesinlik ağırlıklı ortalama	0.977

Tablo 12’de; boosting topluluk öğrenme yöntemi ile Naive Bayes sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Boosting topluluk öğrenme yönteminde, Naive Bayes sınıflayıcısının doğruluk değeri 0.831, kappa katsayısı 0.775, duyarlılık ağırlıklı ortalama değeri 0.832 ve kesinlik ağırlıklı ortalama değeri 0.834 olarak elde edilmiştir.

Tablo 12. Boosting Topluluk Öğrenme Yöntemi İle Naive Bayes Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.831
Kappa katsayısı	0.775
Duyarlılık ağırlıklı ortalama	0.832
Kesinlik ağırlıklı ortalama	0.834

Tablo 13’te; boosting topluluk öğrenme yöntemi ile Karar Ağacı sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Boosting topluluk öğrenme yönteminde, Karar Ağacı sınıflandırıcısının doğruluk değeri 0.904, kappa katsayısı 0.873, duyarlılık ağırlıklı ortalama değeri 0.905 ve kesinlik ağırlıklı ortalama değeri 0.910 olarak elde edilmiştir.

Tablo 13. Boosting Topluluk Öğrenme Yöntemi İle Karar Ağacı Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.904
Kappa katsayısı	0.873
Duyarlılık ağırlıklı ortalama	0.905
Kesinlik ağırlıklı ortalama	0.910

Tablo 14’te; boosting topluluk öğrenme yöntemi ile Derin Sinir Ağları sınıflandırıcısının hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Boosting topluluk öğrenme yönteminde Derin Sinir Ağları sınıflandırıcısının doğruluk değeri 0.980, kappa katsayısı 0.974, duyarlılık ağırlıklı ortalama değeri 0.980 ve kesinlik ağırlıklı ortalama değeri 0.980 olarak elde edilmiştir.

Tablo 14. Boosting Topluluk Öğrenme Yöntemi İle Derin Sinir Ağları Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.980
Kappa katsayısı	0.974
Duyarlılık ağırlıklı ortalama	0.980
Kesinlik ağırlıklı ortalama	0.980

Tablo 15’te; boosting topluluk öğrenme yöntemi ile Random Forest sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Boosting topluluk öğrenme yönteminde, Random Forest sınıflandırıcısının doğruluk değeri 0.644, kappa katsayısı 0.526, duyarlılık ağırlıklı ortalama değeri 0.645 ve kesinlik ağırlıklı ortalama değeri 0.596 olarak elde edilmiştir.

Tablo 15. Boosting Topluluk Öğrenme Yöntemi İle Random Forest Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.644
Kappa katsayısı	0.526
Duyarlılık ağırlıklı ortalama	0.645
Kesinlik ağırlıklı ortalama	0.596

Tablo 16’da; boosting topluluk öğrenme yöntemi ile Gradyan Artırmalı Ağaç sınıflandırıcısının hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Boosting topluluk öğrenme yönteminde, Gradyan Artırmalı Ağaç sınıflandırıcısının doğruluk değeri 0.952, kappa katsayısı 0.936, duyarlılık ağırlıklı Ortalama Değeri 0.952 Ve Kesinlik Ağırlıklı Ortalama Değeri 0.952 Olarak Elde Edilmiştir.

Tablo 16. Boosting Topluluk Öğrenme Yöntemi İle Gradyan Artırmalı Ağaç Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.952
Kappa katsayısı	0.936
Duyarlılık ağırlıklı ortalama	0.952
Kesinlik ağırlıklı ortalama	0.952

Çalışmanın bulgularına göre, boosting topluluk öğrenme yönteminin sınıflandırıcıları arasında en yüksek performansı Derin Sinir Ağları sınıflandırıcısı elde etmiştir. Naive Bayes ve Karar Ağacı sınıflandırıcıları bireysel makine öğrenmesi sınıflandırma performansında diğer modellere göre düşük performans göstermiştir. Boosting genel olarak bu bireysel modelleri iyileştirerek daha iyi bir performans sağlamalarına neden olmuştur. Boosting topluluk öğrenme yöntemi; diğer sınıflandırıcıların aksine, Random Forest sınıflandırma performansının artmasını sağlamamıştır. Boosting topluluk öğrenme yöntemlerinin sınıflandırma performanslarının karşılaştırılması Tablo 17’de gösterilmiştir.

Tablo 17. Boosting Topluluk Öğrenme Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması

		Performans Değerlendirmesi			
		Doğruluk	Kappa katsayısı	Duyarlılık ağırlıklı ortalama	Kesinlik ağırlıklı ortalama
Boosting Topluluk Öğrenme Yöntemi	KNN	0.977	0.969	0.977	0.977
	Naive Bayes	0.831	0.775	0.832	0.834
	Karar Ağaçları	0.904	0.873	0.905	0.910
	Derin Sinir Ağları	0.980	0.974	0.980	0.980
	Random Forest	0.644	0.526	0.645	0.596
		Gradyan Artırmalı Ağaç	0.952	0.936	0.952

3.6.3. Bagging Topluluk Öğrenme Yöntemi ile Sınıflandırıcıların Performanslarının Değerlendirilmesi

Hanehalkı toplam enerji harcamalarının tahmin edilmesi için bagging topluluk öğrenme yönteminin sınıflandırma performansları sunulmuştur. KNN, Naive Bayes,

Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç olmak üzere en fazla performans gösteren altı farklı sınıflandırıcı, modele tek tek dahil edilmiştir.

Tablo 18’de; bagging topluluk öğrenme yöntemi ile KNN sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Bagging topluluk öğrenme yönteminde, KNN sınıflandırıcısının doğruluk değeri 0.977, kappa katsayısı 0.970, duyarlılık ağırlıklı ortalama değeri 0.977 ve kesinlik ağırlıklı ortalama değeri 0.977 olarak elde edilmiştir.

Tablo 18. Bagging Topluluk Öğrenme Yöntemi İle KNN Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.977
Kappa katsayısı	0.970
Duyarlılık ağırlıklı ortalama	0.977
Kesinlik ağırlıklı ortalama	0.977

Tablo 19’da; Bagging topluluk öğrenme yöntemi ile Naive Bayes sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Bagging topluluk öğrenme yönteminde Naive Bayes sınıflandırıcısının, doğruluk değeri 0.786, kappa katsayısı 0.714, duyarlılık ağırlıklı ortalama değeri 0.785 ve kesinlik ağırlıklı ortalama değeri 0.792 olarak elde edilmiştir.

Tablo 19. Bagging Topluluk Öğrenme Yöntemi İle Naive Bayes Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.786
Kappa katsayısı	0.714
Duyarlılık ağırlıklı ortalama	0.785
Kesinlik ağırlıklı ortalama	0.792

Tablo 20’de; Bagging topluluk öğrenme yöntemi ile Karar Ağacı sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Bagging topluluk öğrenme yönteminde, Karar Ağacı sınıflandırıcısının doğruluk değeri

0.816, kappa katsayısı 0.754, duyarlılık ağırlıklı ortalama değeri 0.816 ve kesinlik ağırlıklı ortalama değeri 0.823 olarak elde edilmiştir.

Tablo 20. Bagging Topluluk Öğrenme Yöntemi İle Karar Ağacı Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.816
Kappa katsayısı	0.754
Duyarlılık ağırlıklı ortalama	0.816
Kesinlik ağırlıklı ortalama	0.823

Tablo 21’de; Bagging topluluk öğrenme yöntemi ile Derin Sinir Ağları sınıflandırıcısının, hane halkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Bagging topluluk öğrenme yönteminde Derin Sinir Ağları sınıflandırıcısının doğruluk değeri 0.984, kappa katsayısı 0.979, duyarlılık ağırlıklı ortalama değeri 0.984 ve kesinlik ağırlıklı ortalama değeri 0.984 olarak elde edilmiştir.

Tablo 21. Bagging Topluluk Öğrenme Yöntemi İle Derin Sinir Ağları Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.984
Kappa katsayısı	0.979
Duyarlılık ağırlıklı ortalama	0.984
Kesinlik ağırlıklı ortalama	0.984

Tablo 22’de; bagging topluluk öğrenme yöntemi ile Random Forest sınıflandırıcısının, hane halkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Bagging topluluk öğrenme yönteminde Random Forest sınıflandırıcısının doğruluk değeri 0.808, kappa katsayısı 0.744, duyarlılık ağırlıklı ortalama değeri 0.808 ve kesinlik ağırlıklı ortalama değeri 0.812 olarak elde edilmiştir.

Tablo 22. Bagging Topluluk Öğrenme Yöntemi İle Random Forest Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.808
Kappa katsayısı	0.744
Duyarlılık ağırlıklı ortalama	0.808
Kesinlik ağırlıklı ortalama	0.812

Tablo 23’de; bagging topluluk öğrenme yöntemi ile Gradyan Artırmalı Ağaç sınıflandırıcısının hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Bagging topluluk öğrenme yönteminde Gradyan Artırmalı Ağaç sınıflandırıcısının doğruluk değeri 0.925, kappa katsayısı 0.897, duyarlılık ağırlıklı ortalama değeri 0.922 ve kesinlik ağırlıklı ortalama değeri 0.923 olarak elde edilmiştir.

Tablo 23. Bagging Topluluk Öğrenme Yöntemi İle Gradyan Artırmalı Ağaç Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.925
Kappa katsayısı	0.897
Duyarlılık ağırlıklı ortalama	0.922
Kesinlik ağırlıklı ortalama	0.923

Çalışmanın bulgularına göre, bagging topluluk öğrenme yönteminin sınıflandırıcıları arasında en yüksek performansı Derin Sinir Ağları sınıflandırıcısı elde etmiştir. Bagging topluluk öğrenme yöntemi genel olarak bireysel modelleri iyileştirerek daha iyi bir performans sağlamalarına neden olmuştur. Bagging topluluk öğrenme yöntemi yalnızca Naive Bayes sınıflandırıcısının performansında artışa neden olmamıştır. Bagging topluluk öğrenme yöntemi sınıflandırıcılarının performans karşılaştırması Tablo 24’te gösterilmiştir.

Tablo 24. Bagging Topluluk Öğrenme Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması

		Performans Değerlendirmesi			
		Doğruluk	Kappa katsayısı	Duyarlılık ağırlıklı ortalama	Kesinlik ağırlıklı ortalama
Bagging Topluluk Öğrenme Yöntemi	KNN	0.977	0.970	0.977	0.977
	Naive Bayes	0.786	0.714	0.785	0.792
	Karar Ağaçları	0.816	0.754	0.816	0.823
	Derin Sinir Ağları	0.984	0.979	0.984	0.984
	Random Forest	0.808	0.744	0.808	0.812
	Gradyan Artırmalı Ağaç	0.925	0.897	0.922	0.923

3.6.4. Voting Topluluk Öğrenme Yöntemi ile Sınıflandırıcıların Performanslarının Değerlendirilmesi

Hanehalkı toplam enerji harcamalarının tahmin edilmesi için voting topluluk öğrenme yöntemlerinin sınıflandırma performansları sunulmuştur. KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç olmak üzere en fazla performans gösteren altı farklı sınıflandırıcı, modele tek tek dahil edilmiştir. Ayrıca, KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç sınıflandırıcılarının birleştirilmesi ile elde edilen ALL model sınıflandırma modellerine eklenmiştir.

Tablo 25'te; voting topluluk öğrenme yöntemi ile KNN sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Voting topluluk öğrenme yönteminde, KNN sınıflandırıcısının doğruluk değeri 0.976, kappa katsayısı 0.968, duyarlılık ağırlıklı ortalama değeri 0.976 ve kesinlik ağırlıklı ortalama değeri 0.976 olarak elde edilmiştir.

Tablo 25. Voting Topluluk Öğrenme Yöntemi İle KNN Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.976
Kappa katsayısı	0.968
Duyarlılık ağırlıklı ortalama	0.976
Kesinlik ağırlıklı ortalama	0.976

Tablo 26’da; voting topluluk öğrenme yöntemi ile Naive Bayes sınıflandırıcısının hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Voting topluluk öğrenme yönteminde, Naive Bayes sınıflandırıcısının doğruluk değeri 0.784, kappa katsayısı 0.711, duyarlılık ağırlıklı ortalama değeri 0.782 ve kesinlik ağırlıklı ortalama değeri 0.791 olarak elde edilmiştir.

Tablo 26. Voting Topluluk Öğrenme Yöntemi İle Naive Bayes Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.784
Kappa katsayısı	0.711
Duyarlılık ağırlıklı ortalama	0.782
Kesinlik ağırlıklı ortalama	0.791

Tablo 27’de; voting topluluk öğrenme yöntemi ile Karar Ağacı sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Voting topluluk öğrenme yönteminde, Karar Ağacı sınıflandırıcısının doğruluk değeri 0.761, kappa katsayısı 0.680, duyarlılık ağırlıklı ortalama değeri 0.759 ve kesinlik ağırlıklı ortalama değeri 0.777 olarak elde edilmiştir.

Tablo 27. Voting Topluluk Öğrenme Yöntemi İle Karar Ağacı Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.761
Kappa katsayısı	0.680
Duyarlılık ağırlıklı ortalama	0.759
Kesinlik ağırlıklı ortalama	0.777

Tablo 28’de; voting topluluk öğrenme yöntemi ile Derin Sinir Ağları sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Voting topluluk öğrenme yönteminde, Derin Sinir Ağları sınıflandırıcısının doğruluk değeri 0.978, kappa katsayısı 0.970, duyarlılık ağırlıklı ortalama değeri 0.978 ve kesinlik ağırlıklı ortalama değeri 0.978 olarak elde edilmiştir.

Tablo 28. Voting Topluluk Öğrenme Yöntemi İle Derin Sinir Ağları Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.978
Kappa katsayısı	0.970
Duyarlılık ağırlıklı ortalama	0.978
Kesinlik ağırlıklı ortalama	0.978

Tablo 29’da; voting topluluk öğrenme yöntemi ile Random Forest sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Voting topluluk öğrenme yönteminde, Random Forest sınıflandırıcısının doğruluk değeri 0.806, kappa katsayısı 0.740, duyarlılık ağırlıklı ortalama değeri 0.806 ve kesinlik ağırlıklı ortalama değeri 0.809 olarak elde edilmiştir.

Tablo 29. Voting Topluluk Öğrenme Yöntemi İle Random Forest Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.806
Kappa katsayısı	0.740
Duyarlılık ağırlıklı ortalama	0.806
Kesinlik ağırlıklı ortalama	0.809

Tablo 30’da; voting topluluk öğrenme yöntemi ile Gradyan Artırmalı Ağaç sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Voting topluluk öğrenme yönteminde, Gradyan Artırmalı Ağaç sınıflandırıcısının doğruluk değeri 0.916, kappa katsayısı 0.888, duyarlılık ağırlıklı ortalama değeri 0.916 ve kesinlik ağırlıklı ortalama değeri 0.917 olarak elde edilmiştir.

Tablo 30. Voting Topluluk Öğrenme Yöntemi İle Gradyan Artırmalı Ağaç Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.916
Kappa katsayısı	0.888
Duyarlılık ağırlıklı ortalama	0.916
Kesinlik ağırlıklı ortalama	0.917

Tablo 31’de; voting topluluk öğrenme yöntemi ile ALL modelin, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. ALL modeli; KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç sınıflandırıcılarının birlikte çalıştırılması ile elde edilen bir modeldir. Voting topluluk öğrenme yönteminde, ALL modeli sınıflandırmasının doğruluk değeri 0.925, kappa katsayısı 0.900, duyarlılık ağırlıklı ortalama değeri 0.925 ve kesinlik ağırlıklı ortalama değeri 0.929 olarak elde edilmiştir.

Tablo 31. Voting Topluluk Öğrenme Yöntemi İle ALL Modelin Sınıflandırma Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.925
Kappa katsayısı	0.900
Duyarlılık ağırlıklı ortalama	0.925
Kesinlik ağırlıklı ortalama	0.929

Çalışmanın bulgularına göre, voting topluluk öğrenme yönteminin sınıflandırıcıları arasında en yüksek performansı Derin Sinir Ağları sınıflandırıcısı elde etmiştir. Naive Bayes ve Karar Ağacı sınıflandırıcıları diğer modellere göre daha düşük sınıflandırma performansı göstermiştir. Voting genel olarak diğer modelleri iyileştirerek daha iyi bir performans sağlamalarına neden olmuştur. Voting topluluk öğrenme yöntemi sınıflandırıcılarının performans değerlendirmesi Tablo 32’de gösterilmiştir.

Tablo 32. Voting Topluluk Öğrenme Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması

		Performans Değerlendirmesi			
		Doğruluk	Kappa katsayısı	Duyarlılık ağırlıklı ortalama	Kesinlik ağırlıklı ortalama
Voting Topluluk Öğrenme Yöntemleri	KNN	0.976	0.968	0.976	0.976
	Naive Bayes	0.784	0.711	0.782	0.791
	Karar Ağaçları	0.761	0.680	0.759	0.777
	Derin Sinir Ağları	0.978	0.970	0.978	0.978
	Random Forest	0.806	0.740	0.806	0.809
	Gradyan Artırmalı Ağaç	0.916	0.888	0.916	0.917
*ALL		0.925	0.900	0.925	0.929

*ALL = KNN + Naive Bayes + Karar Ağaçları + Derin Sinir Ağları + Random Forest + Gradyan Artırmalı Ağaç

3.6.5. Stacking Topluluk Öğrenme Yöntemi ile Sınıflandırıcıların Performanslarının Değerlendirilmesi

Hanehalkı toplam enerji harcamalarının tahmin edilmesi için stacking topluluk öğrenme sınıflandırma performansları sunulmuştur. KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç olmak üzere en fazla performans gösteren altı farklı sınıflandırıcı, modele tek tek dahil edilmiştir. Ayrıca, KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç sınıflandırıcılarının birleştirilmesi ile elde edilen ALL model sınıflandırma modellerine eklenmiştir.

Tablo 33'de; stacking topluluk öğrenme yöntemi ile KNN sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Stacking topluluk öğrenme yönteminde, KNN sınıflandırıcısının doğruluk değeri 0.977, kappa katsayısı 0.969, duyarlılık ağırlıklı ortalama değeri 0.977 ve kesinlik ağırlıklı ortalama değeri 0.977 olarak elde edilmiştir.

Tablo 33. Stacking Topluluk Öğrenme Yöntemi İle KNN Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.977
Kappa katsayısı	0.969
Duyarlılık ağırlıklı ortalama	0.977
Kesinlik ağırlıklı ortalama	0.977

Tablo 34’te; stacking topluluk öğrenme yöntemi ile Naive Bayes sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Stacking topluluk öğrenme yönteminde, Naive Bayes sınıflandırıcısının doğruluk değeri 0.801, kappa katsayısı 0.735, duyarlılık ağırlıklı ortalama değeri 0.800 ve kesinlik ağırlıklı ortalama değeri 0.808 olarak elde edilmiştir.

Tablo 34. Stacking Topluluk Öğrenme Yöntemi İle Naive Bayes Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.801
Kappa katsayısı	0.735
Duyarlılık ağırlıklı ortalama	0.800
Kesinlik ağırlıklı ortalama	0.808

Tablo 35’de; stacking topluluk öğrenme yöntemi ile Karar Ağacı sınıflandırıcısının hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Stacking topluluk öğrenme yönteminde, Karar Ağacı sınıflandırıcısının doğruluk değeri 0.842, kappa katsayısı 0.789, duyarlılık ağırlıklı ortalama değeri 0.841 ve kesinlik ağırlıklı ortalama değeri 0.847 olarak elde edilmiştir.

Tablo 35. Stacking Topluluk Öğrenme Yöntemi İle Karar Ağacı Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.842
Kappa katsayısı	0.789
Duyarlılık ağırlıklı ortalama	0.841
Kesinlik ağırlıklı ortalama	0.847

Tablo 36’da; stacking topluluk öğrenme yöntemi ile Derin Sinir Ağları sınıflandırıcısının, hanehalkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Stacking topluluk öğrenme yönteminde Derin Sinir Ağları sınıflandırıcısının doğruluk değeri 0.974, kappa katsayısı 0.965, duyarlılık ağırlıklı ortalama değeri 0.974 ve kesinlik ağırlıklı ortalama değeri 0.974 olarak elde edilmiştir.

Tablo 36. Stacking Topluluk Öğrenme Yöntemi İle Derin Sinir Ağları Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.974
Kappa katsayısı	0.965
Duyarlılık ağırlıklı ortalama	0.974
Kesinlik ağırlıklı ortalama	0.974

Tablo 37’de; stacking topluluk öğrenme yöntemi ile Random Forest sınıflandırıcısının, hane halkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Stacking topluluk öğrenme yönteminde, Random Forest sınıflandırıcısının doğruluk değeri 0.818, kappa katsayısı 0.757, duyarlılık ağırlıklı ortalama değeri 0.818 ve kesinlik ağırlıklı ortalama değeri 0.822 olarak elde edilmiştir.

Tablo 37. Stacking Topluluk Öğrenme Yöntemi İle Random Forest Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.818
Kappa katsayısı	0.757
Duyarlılık ağırlıklı ortalama	0.818
Kesinlik ağırlıklı ortalama	0.822

Tablo 38’de; stacking topluluk öğrenme yöntemi ile Gradyan artırmalı Ağaç sınıflandırıcısının, hane halkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. Stacking topluluk öğrenme yönteminde, Gradyan artırmalı Ağaç sınıflandırıcısının doğruluk değeri 0.920, kappa katsayısı 0.893, duyarlılık ağırlıklı ortalama değeri 0.920 ve kesinlik ağırlıklı ortalama değeri 0.921 olarak elde edilmiştir.

Tablo 38. Stacking Topluluk Öğrenme Yöntemi İle Gradyan Artırmalı Ağaç Sınıflandırıcısının Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.920
Kappa katsayısı	0.893
Duyarlılık ağırlıklı ortalama	0.920
Kesinlik ağırlıklı ortalama	0.921

Tablo 39’da; stacking topluluk öğrenme yöntemi ile ALL modelin, hane halkı toplam enerji harcamaları sınıflandırma performans sonuçları gösterilmiştir. ALL model; KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç sınıflandırıcılarının birlikte çalıştırılması ile elde edilen bir modeldir. Stacking topluluk öğrenme yönteminde, ALL modeli sınıflandırmasının doğruluk değeri 0.984, kappa katsayısı 0.979, duyarlılık ağırlıklı ortalama değeri 0.984 ve kesinlik ağırlıklı ortalama değeri 0.984 olarak elde edilmiştir.

Tablo 39. Stacking Topluluk Öğrenme Yöntemi İle ALL Modelin Sınıflandırma Performanslarının Değerlendirilmesi

Performans Ölçütü	Değer
Doğruluk	0.984
Kappa katsayısı	0.979
Duyarlılık ağırlıklı ortalama	0.984
Kesinlik ağırlıklı ortalama	0.984

Çalışmanın bulgularına göre, stacking topluluk öğrenme yönteminin sınıflandırıcıları arasında en yüksek performans ALL model sınıflandırmasında elde edilmiştir. ALL model; KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç sınıflandırıcılarının birlikte çalıştırılması ile elde edilen bir modeldir. Stacking topluluk öğrenme modeli genel olarak diğer sınıflandırıcıları iyileştirerek daha iyi bir performans sağlamalarına neden olmuştur. Stacking topluluk öğrenme yöntemi sınıflandırıcılarının performans değerlendirmesi Tablo 40’da gösterilmiştir.

Tablo 40. Stacking Topluluk Öğrenme Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması

		Performans Değerlendirmesi			
		Doğruluk	Kappa katsayısı	Duyarlılık ağırlıklı ortalama	Kesinlik ağırlıklı ortalama
Stacking Topluluk Öğrenme Yöntemleri	KNN	0.977	0.969	0.977	0.977
	Naive Bayes	0.801	0.735	0.800	0.808
	Karar Ağaçları	0.842	0.789	0.841	0.847
	Derin Sinir Ağları	0.974	0.965	0.974	0.974
	Random Forest	0.818	0.757	0.818	0.822
	Gradyan Artırmalı Ağaç	0.920	0.893	0.920	0.921
*ALL		0.984	0.979	0.984	0.984

*ALL = KNN + Naive Bayes + Karar Ağaçları + Derin Sinir Ağları + Random Forest + Gradyan Artırmalı Ağaç

Çalışmadaki bulguları en genel hali ile değerlendirecek olursak, ALL model sınıflandırmasında stacking topluluk öğrenme yöntemi ve Derin Sinir Ağları sınıflandırmasında bagging topluluk öğrenme yöntemi, topluluk öğrenme yöntemleri arasında en yüksek performansı elde etmiştir. Bahsi geçen modeller için bagging ve stacking topluluk öğrenme yöntemlerinin her ikisinde doğruluk değeri 0.984, kappa katsayısı değeri 0.979, duyarlılık ağırlıklı ortalama değeri 0.984 ve kesinlik ağırlıklı ortalama değeri 0.984 olarak elde edilmiştir.

Sonuç olarak; bireysel makine öğrenmesi yöntemleri ve topluluk öğrenme yöntemlerinin sınıflandırma performans değerlendirmeleri Tablo 41’de bir bütün olarak gösterilmiştir.

Tablo 41. Topluluk Öğrenme Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması

		Performans		Değerlendirmesi	
		Doğruluk	Kappa katsayısı	Duyarlılık ağırlıklı ortalama	Kesinlik ağırlıklı ortalama
Makine Öğrenmesi Yöntemleri	KNN	0.977	0.969	0.977	0.977
	Naive Bayes	0.785	0.713	0.784	0.791
	Karar Ağaçları	0.775	0.700	0.774	0.787
	Derin Sinir Ağları	0.973	0.964	0.973	0.973
	Random Forest	0.802	0.736	0.802	0.805
	Gradyan Artırmalı Ağaç	0.917	0.893	0.919	0.920
Boosting Topluluk Öğrenme Yöntemleri	KNN	0.977	0.969	0.977	0.977
	Naive Bayes	0.831	0.775	0.832	0.834
	Karar Ağaçları	0.904	0.873	0.905	0.910
	Derin Sinir Ağları	0.980	0.974	0.980	0.980
	Random Forest	0.644	0.526	0.645	0.596
	Gradyan Artırmalı Ağaç	0.952	0.936	0.952	0.952
Bagging Topluluk Öğrenme Yöntemleri	KNN	0.977	0.970	0.977	0.977
	Naive Bayes	0.786	0.714	0.785	0.792
	Karar Ağaçları	0.816	0.754	0.816	0.823
	Derin Sinir Ağları	0.984	0.979	0.984	0.984
	Random Forest	0.808	0.744	0.808	0.812
	Gradyan Artırmalı Ağaç	0.925	0.897	0.922	0.923
Voting Topluluk Öğrenme Yöntemleri	KNN	0.976	0.968	0.976	0.976
	Naive Bayes	0.784	0.711	0.782	0.791
	Karar Ağaçları	0.761	0.680	0.759	0.777
	Derin Sinir Ağları	0.978	0.970	0.978	0.978
	Random Forest	0.806	0.740	0.806	0.809
	Gradyan Artırmalı Ağaç	0.916	0.888	0.916	0.917
*ALL		0.925	0.900	0.925	0.929
Stacking Topluluk Öğrenme Yöntemleri	KNN	0.977	0.969	0.977	0.977
	Naive Bayes	0.801	0.735	0.800	0.808
	Karar Ağaçları	0.842	0.789	0.841	0.847
	Derin Sinir Ağları	0.974	0.965	0.974	0.974
	Random Forest	0.818	0.757	0.818	0.822
	Gradyan Artırmalı Ağaç	0.920	0.893	0.920	0.921
*ALL		0.984	0.979	0.984	0.984

*ALL = KNN + Naive Bayes + Karar Ağaçları + Derin Sinir Ağları + Random Forest + Gradyan Artırmalı Ağaç

SONUÇ ve TARTIŞMA

Enerji, ülkelerin ekonomik ve sosyal kalkınmasında stratejik bir role sahiptir. Son yıllarda, enerji talebi tüm dünyada katlanarak artmaktadır ve enerji talebini tahmin etmek birçok ülkede ana endişelerden biri haline gelmektedir (Nabavi vd., 2020). Türkiye'de hanehalkı enerji tüketiminin sektörlere göre dağılımına bakıldığında, toplam enerji tüketiminin %35'lik payını oluşturan bir orana sahip olduğu görülmektedir. Bu nedenle, hanelerin enerji tüketimi, Türkiye'de toplam enerji tüketiminin önemli bir bölümünü oluşturmaktadır. Hanehalklarının toplam enerji harcamalarının keşfedilmesi, özellikle politika yapıcılar açısından dikkatle takip edilmesi gereken bir konudur. Ayrıca Türkiye gibi enerjiye bağımlı ülkelerde enerji harcama profillerinin farklı yaklaşımlarla ele alınması, hâlâ araştırılması gereken bir alandır.

Birçok ülkede, hanehalkları önemli miktarda enerji tüketir. Hanehalkı enerji tüketimini anlamak uzun zamandır önemli bir araştırma konusudur. Tüketimi etkileyen çok sayıda davranışsal, teknolojik faktör ve politika faktörü konuyu karmaşık hale getirmektedir. Ekonomik gelişmelerle birlikte tüketim alışkanlıkları yelpazesinin kontrolsüz bir şekilde büyümesi sonucunda; enerji kullanımı, dünyamızın varlığını birçok yönden tehdit eden daha büyük bir bileşen haline gelmiştir (TSKB; 2020). Tüketiciler için etkili bir enerji politikası tasarlamak, bazı hanelerin neden diğerlerinden daha fazla enerjiye ihtiyaç duyduğunu anlamayı gerektirmektedir (Vringer vd., 2007). Bu açıdan haneler, gelecekteki enerji tasarrufu için büyük bir potansiyeli temsil etmektedir. Sonuç olarak, politika yapıcıların en önemli temel amaçları; hanehalkı enerji kullanımının birincil faktörlerini anlamak, hanehalkı enerji tüketimini azaltmak için etkili önlemleri keşfetmek ve artan enerji ihtiyaçlarını sürdürülebilir ve güvenli bir şekilde karşılamak olmalıdır (Karaaslan, Adar ve Delice, 2021; Salari ve Javid, 2017).

Son yıllarda, hanehalkı enerji talebini tahmin etmek için çeşitli yöntemler kullanılmaktadır. 1980'lerden bu yana, hanehalkı enerji tüketim talebini tahmin etmek için kullanılan istatistiksel yöntemler hakkında geniş bir literatür ortaya konulmuştur. Bununla birlikte, hanehalkı enerji tüketimi alanında makine öğrenmesi yöntemlerini kullanan literatür büyümeye devam etmektedir. Yapay zeka alanında hızla ilerleyen gelişmeler ışığında, yeni yaklaşımlar ve modeller ile yapılan çalışmalar hâlâ yetersizdir. İstikrarlı tahminlerin yapıldığı başarılı sınıflandırma yaklaşımları karar vericiler için

kritik bir öneme sahiptir. Bu bağlamda, bu çalışmada, topluluk öğrenme yöntemleri kullanılarak yeni bir sınıflandırma yaklaşımı sunulmuştur.

Topluluk öğrenme yöntemlerini günlük hayatımızda karar verme mekanizmamızdan ortaya çıkan bir alan olarak tanımlayabiliriz. Günlük hayatta, bir konu hakkında nihai bir karara varmadan önce çeşitli fikirleri tartıp birleştirerek doğru kararı veririz. Her karar vericinin muhakkak bir kusuru vardır. Bu nedenle birden fazla karar vericiye danışılarak alınan kararlarda hata payı en aza indirilmektedir (Polikar, 2012). Topluluk öğrenme yaklaşımları, daha iyi bir sınıflandırma tahmini yapmak için bireysel makine öğrenmesi modellerinin yanı sıra bireysel modelleri birleştirmeyi amaçlamaktadır. Topluluk öğrenme yöntemleri, çeşitli makine öğrenmesi problemlerine, farklı birleştirme teknikleri kullanarak sezgisel, basit, zarif ve güçlü çözümler sunmaktadır. Bu yaklaşım, makine öğrenmesi sonuçlarını iyileştirmektedir ve tek bir modelin tahmininden daha iyi tahminler üretmektedir.

Bu çalışmada Türkiye İstatistik Kurumu (TÜİK)'ten elde edilen Hanehalkı Bütçe Araştırması anketi 2019 veri seti kullanılmıştır. 11.521 hanenin hanehalkı tüketim verileri incelenmiştir. Literatür taraması sonucunda oluşturulan değişkenler makine öğrenmesi ve topluluk öğrenmesi yöntemleri kullanılarak sınıflandırılmıştır. Topluluk öğrenme yöntemleri ve makine öğrenme yöntemlerinin sınıflandırma performansı, hanehalkı toplam enerji harcama tahmini için değerlendirilmiştir. KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç bireysel sınıflandırıcılarının tahmin performansının yanı sıra bagging, boosting, voting ve stacking topluluk öğrenme yöntemleri ile birleştirilen performanslar da değerlendirilmiştir.

KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç sınıflandırıcıları; makine öğrenmesi ve topluluk öğrenme yöntemleri ile tek tek sınıflandırılmıştır. Ayrıca, KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç sınıflandırıcılarının birleştirilmesi ile elde edilen ALL model, sınıflandırma modellerine eklenmiştir. Çalışmanın bulgularına göre, bireysel makine öğrenmesi yöntemine tabi tutulan sınıflandırıcılar arasında en yüksek performansı KNN sınıflandırıcısı elde etmiştir. KNN sınıflandırıcısının doğruluk değeri 0.977, kappa katsayısı 0.969, duyarlılık ağırlıklı ortalama değeri 0.977 ve kesinlik

ağırlıklı ortalama değeri 0.977 olarak elde edilmiştir. Genel olarak, KNN'nin tüm topluluk öğrenme yöntemlerinde de yüksek performansa sahip olduğu bulunmuştur. Bu sonuç, literatürdeki çalışmalar ile uyum sağlamaktadır (Hassanat, Abbadi, Altarawneh ve Alhasanat, 2014). Bir başka açıdan KNN topluluk öğrenme yöntemleri ile performansını çok fazla artıran bir yöntem olmamıştır. KNN, tek başına yani topluluk öğrenme yöntemleri ile birleştirilmediğinde de yüksek tahmin performansı sunmuştur. Bireysel makine öğrenmesi sınıflandırma performansında Naive Bayes ve Karar Ağacı sınıflandırıcıları diğer modellere göre düşük tahmin performansı göstermiştir.

Çalışmanın bulgularına göre, boosting topluluk öğrenme yöntemi sınıflandırıcıları arasında en yüksek performansı Derin Sinir Ağları sınıflandırıcısı elde etmiştir. Boosting topluluk öğrenme yönteminde Derin Sinir Ağları sınıflandırıcısının doğruluk değeri 0.980, kappa katsayısı 0.974, duyarlılık ağırlıklı ortalama değeri 0.980 ve kesinlik ağırlıklı ortalama değeri 0.980 olarak elde edilmiştir. Boosting genel olarak bireysel makine öğrenmesi modellerini iyileştirerek daha iyi bir performans sağlamalarına neden olmuştur. Boosting topluluk öğrenme yöntemi; diğer makine öğrenmesi yöntemlerinin aksine, Random Forest sınıflandırıcısının performansının artmasına katkıda bulunmamıştır. Tüm topluluk öğrenme model sonuçları incelendiğinde, Random Forest tahmin edici olduğunda yanıltıcı sonuçlar elde edilebileceği tespit edilmiştir. Bu sonuç, literatürdeki çalışma (Archer ve Kimes, 2008) ile uyum sağlamaktadır. Literatürde; farklı tahmin modelleri ile karşılaştırıldığında, Random Forest modelinin en iyi tahmin modeli olduğunu gösteren çalışma (Wang vd., 2018) da bulunmaktadır.

Çalışmanın bulgularına göre, voting topluluk öğrenme yönteminin sınıflandırıcıları arasında en yüksek performansı Derin Sinir Ağları sınıflandırıcısı elde etmiştir. Naive Bayes ve Karar Ağacı sınıflandırıcıları, voting topluluk öğrenme yönteminde, diğer modellere göre daha düşük sınıflandırma performansı göstermiştir. Voting, Naive Bayes ve Karar Ağacı dışındaki modelleri genel olarak iyileştirerek daha iyi bir performans sağlamalarına neden olmuştur. Voting topluluk öğrenme yönteminde Derin Sinir Ağları sınıflandırıcısının doğruluk değeri 0.978, kappa katsayısı 0.970, duyarlılık ağırlıklı ortalama değeri 0.978 ve kesinlik ağırlıklı ortalama değeri 0.978 olarak elde edilmiştir.

Çalışmada, voting ve stacking topluluk öğrenme yöntemlerinde ALL model kullanılmıştır. ALL model; KNN, Naive Bayes, Karar Ağacı, Derin Sinir Ağları, Random Forest ve Gradyan Artırmalı Ağaç sınıflandırıcılarının birlikte çalıştırılması ile elde edilen bir modeldir. ALL modelinde stacking topluluk öğrenme yöntemi ve Derin Sinir Ağları modelinde bagging topluluk öğrenme yöntemi, topluluk öğrenme yöntemleri arasında en yüksek performansı elde etmiştir. Bagging ve stacking topluluk öğrenme yöntemlerinin her ikisinin de doğruluk değeri 0.984, kappa katsayısı değeri 0.979, duyarlılık ağırlıklı ortalama değeri 0.984 ve kesinlik ağırlıklı ortalama değeri 0.984 olarak elde edilmiştir. Bu çalışmada, bagging ve stacking topluluk öğrenme modellerinin daha iyi sınıflandırma performansına sahip olmaları literatürle uyumludur (Dietterich, 2000; Doğaner vd., 2021; Erdoğan ve Namlı, 2019; Onan, Korukoğlu ve Bulut, 2016; Prusa, Khoshgoftaar ve Dittman, 2015; Seker ve Ocak, 2019; Wang vd., 2011).

Yöntemin sonuçları en genel şekilde incelendiğinde, hanehalkı toplam enerji harcaması sınıflandırmasında en yüksek performansı stacking ve bagging topluluk öğrenme yöntemi sağlamıştır. Boosting ve voting topluluk öğrenme yöntemleri ise bazı sınıflandırıcıların performansına katkı sağlarken bazılarında önemli bir etki oluşturamadıkları gözlemlenmiştir.

Sonuç olarak bu çalışmada, hanehalkı toplam enerji harcamalarının sınıflandırma performansları, topluluk öğrenme yöntemleri ile değerlendirilmiştir. Topluluk öğrenme yöntemleri ile hanehalkı enerji harcamalarının sınıflandırılmasının uygunluğunu araştıran çok fazla çalışma bulunmamaktadır. Bu boşluğu gidermek için, bu çalışmada yeni bir topluluk öğrenme yaklaşımı sunulmuştur. Araştırmamızın amacı, topluluk öğrenme yöntemlerinin hanehalkının toplam enerji harcamalarını tahmin edip edemeyeceği ve kullanabilecekleri temel makine öğrenmesi sınıflandırıcılarına kıyasla sınıflandırmalarını iyileştirip iyileştiremeyeceğidir. Önerilen topluluk öğrenme yaklaşımının sonuçları ile hanehalkı toplam enerji harcamaları çok yüksek doğrulukla sınıflandırılmıştır. Çalışmada; “ısınma ve mutfak harcama”, “aydınlatma harcama”, “yakıt harcama” değerlerinin toplanarak hanelerin “toplam enerji harcama” değerlerinin elde edilmesi, bu alana yeni bir bakış açısı sunmaktadır. Bu bakış açısı ile hanehalkı “toplam” enerji harcamalarını bir bütün olarak ele alan bu çalışma, büyüyen literatüre yeni bir yaklaşım eklemiştir. Bu yönü ile çalışma; hanehalkı toplam enerji harcamalarını sınıflandırmada kullanılan topluluk öğrenme yöntemleri hakkında, gelişen literatüre katkıda bulunmuştur.

Çalışma, topluluk öğrenme yöntemleri ile hanehalkı” toplam” enerji harcaması üzerine yapılan ilk çalışma olup, diğer yöntemlere göre daha kararlı sonuçlar elde edilmiştir. Bu bağlamda, çalışmada topluluk öğrenme yöntemlerinin önemi vurgulanmıştır. Elde edilen sonuçlar, politika yapıcıları enerji projeleri geliştirmeden önce bilinçli kararlar almaya sevkedecek ve enerji verimliliğini yönetmek için önerilen yaklaşımı kullanmaya motive edecektir. Ayrıca, hem tahminin doğruluğu hem de modellerin kullanıma uygunluğu enerji yönetiminin planlanmasında kullanıcıları destekleyecektir.

Gelecekte yapılacak olan çalışmalar için; topluluk öğrenme yöntemlerinde modele daha fazla değişken dahil etmenin, modelin sınıflandırma performansını artıracığı söylenebilir. Ayrıca, topluluk öğrenme yöntemlerinde sınıflandırma performansını artırmak için farklı türde öznitelik seçimi ve çapraz doğrulama yöntemleri kullanılabilir. Kanıtlanmış yöntemlerle istikrarlı bir şekilde sınıflara ayrılan hanehalkları için enerji tüketim tipolojisi oluşturulabilir. Oluşturulan tipoloji sayesinde, teknolojik ve kurumsal bir bağlamda hanelerin enerji tüketim kararları öngörülebilir. Başarılı bir tahmin modeli kullanılarak sınıflara ayrılan hanehalkları için etkili ekonomik, sosyal ve teknolojik faktörlerin tanımlanmasında politika yapıcıların daha istikrarlı kararlar alması sağlanabilir.

KAYNAKÇA

- Adusah-Poku, F., & Takeuchi, K. (2019). "Household Energy Expenditure in Ghana: A Double-Hurdle Model Approach" [Gana'da Hane Halkı Enerji Harcaması: Çift Engel Modeli Yaklaşımı]. *World Development*, 117, 266-277.
- Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H., & Saidur, R. (2014). "A Review on Applications of ANN And SVM For Building Electrical Energy Consumption Forecasting" [Elektrik Enerjisi Tüketimi Tahmini Oluşturmak İçin YSA ve SVM Uygulamaları Üzerine Bir İnceleme. Yenilenebilir ve Sürdürülebilir Enerji İncelemeleri]. *Renewable and Sustainable Energy Reviews*, 33, 102-109.
- Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). "Trees Vs Neurons: Comparison Between Random Forest And ANN For High-Resolution Prediction of Building Energy Consumption" [Ağaçlar ve Nöronlar: Bina Enerji Tüketiminin Yüksek Çözünürlüklü Tahmini İçin Rastgele Orman ve YSA Karşılaştırması]. *Energy and buildings*, 147, 77-89.
- Aizenberg, I. N., Aizenberg, N. N., & Vandewalle, J. (2000). Multiple-Valued Threshold Logic and Multi-Valued Neurons. In *Multi-valued and Universal Binary Neurons* (pp. 25-80). Springer, Boston, MA.
- Akar, Ö., & Güngör, O. (2012). "Classification of Multispectral Images Using Random Forest Algorithm" [Random Forest Algoritması Kullanılarak Multispektral Görüntülerin Sınıflandırılması]. *Journal of Geodesy and Geoinformation*, 1(2), 105-112.
- Albayrak, A. S., & Yilmaz, S. K. (2009). "Veri Madenciliği: Karar Ağacı Algoritmaları Ve İmkb Verileri Üzerine Bir Uygulama". *Suleyman Demirel University Journal of Faculty of Economics & Administrative Sciences*, 14(1). 31-52.
- Alobaidi, M. H., Chebana, F., & Meguid, M. A. (2018). "Robust Ensemble Learning Framework for Day-Ahead Forecasting of Household Based Energy Consumption" [Hane Bazlı Enerji Tüketiminin Gün Öncesi Tahmini için Sağlam Topluluk Öğrenme Çerçevesi]. *Applied energy*, 212, 997-1012.

- Alper, C. E., & Mumcu-Serdar, A. (2000). Türkiye’de Otomobil Talebinin Tahmini. *Ekonomi ve Ekonometri Merkezi, Boğaziçi Üniversitesi*.
- Archer, K. J., & Kimes, R. V. (2008). “Empirical Characterization of Random Forest Variable Importance Measures” [Random Forest Değişken Önem Ölçülerinin Ampirik Karakterizasyonu] . *Computational Statistics & Data Analysis*, 52(4), 2249-2260.
- Arunadevi, J., & Nithya, M. J. (2016). “Comparison Of Feature Selection Strategies For Classification Using RapidMiner” [RapidMiner Kullanarak Sınıflandırma İçin Öznitelik Seçme Stratejilerinin Karşılaştırılması]. *International Journal of Innovative Research in Computer and Communication Engineering*, 4(7), 13556-13563.
- Atems, B., & Hotaling, C. (2018). “The Effect of Renewable and Nonrenewable Electricity Generation on Economic Growth” [Yenilenebilir Ve Yenilenemez Elektrik Üretiminin Ekonomik Büyümeye Etkisi]. *Energy Policy*, 112, 111-118.
- Ayık, Y. Z., Özdemir, A., & Yavuz, U. (2007). “Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkisinin Veri Madenciliği Tekniği İle Analizi”. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2), 441-454.
- Azevedo, A., & Santos, M. F. (2008). “KDD, SEMMA and CRISP-DM: A Parallel Overview” [KDD, SEMMA ve CRISP-DM: Paralel Bir Genel Bakış]. *IADS-DM*.
- Bagui, S. C. (2005). “Combining Pattern Classifiers: Methods and Algorithms” [Örüntü Sınıflandırıcıları Birleştirme: Yöntemler ve Algoritmalar]. *Technometrics*, 47(4), 517-518.
- Beckel, C., Sadamori, L., Staake, T., & Santini, S. (2014). “Revealing Household Characteristics From Smart Meter Data” [Akıllı Sayaç Verilerinden Hanehalkı Karakteristiklerini Ortaya Çıkarmak]. *Energy*, 78, 397-410.
- Bedir, M., Hasselaar, E., & Itard, L. (2013). “Determinants Of Electricity Consumption In Dutch Dwellings” [Hollanda Konutlarında Elektrik Tüketiminin Belirleyicileri]. *Energy and buildings*, 58, 194-207.
- Belaid, F., & Rault, C. (2021).” Energy Expenditure in Egypt: Empirical Evidence Based on a Quantile Regression Approach” [Mısır'da Enerji Harcaması: Niceliksel

- Regresyon Yaklaşımına Dayalı Ampirik Kanıtlar]. *Environmental Modeling & Assessment*, 26(4), 511-528.
- Blum, A. L., & Langley, P. (1997). "Selection of Relevant Features and Examples in Machine Learning" [Makine Öğreniminde İlgili Özelliklerin Ve Örneklerin Seçimi]. *Artificial intelligence*, 97(1-2), 245-271.
- Breiman, L. (1996). "Bagging Predictors" [Bagging Tahmin Edicileri]. *Machine learning*, 24(2), 123-140.
- Brounen, D., Kok, N., & Quigley, J. M. (2012). "Residential Energy Use and Conservation: Economics and Demographics" [Konutlarda Enerji Kullanımı ve Korunması: Ekonomi ve Demografi]. *European Economic Review*, 56(5), 931-945.
- Burnett, J. W., & Kiesling, L. L. (2022). "How Do Machines Predict Energy Use? Comparing Machine Learning Approaches for Modeling Household Energy Demand İn The United States" [Makineler Enerji Kullanımını Nasıl Tahmin Ediyor? Amerika Birleşik Devletleri'nde Hanehalkı Enerji Talebini Modellemek için Makine Öğrenimi Yaklaşımlarını Karşılaştırma]. *Energy Research & Social Science*, 91, 102715.
- Chandrashekar, G., & Sahin, F. (2014). "A Survey on Feature Selection Methods" [Öznitelik Seçme Yöntemleri Üzerine Bir Anket]. *Computers & Electrical Engineering*, 40(1), 16-28.
- Chou, J. S., & Tran, D. S. (2018). "Forecasting Energy Consumption Time Series Using Machine Learning Techniques Based on Usage Patterns of Residential Householders" [Konut Sakinlerinin Kullanım Kalıplarına Dayalı Makine Öğrenimi Tekniklerini Kullanarak Enerji Tüketimi Zaman Serilerini Tahmin Etme]. *Energy*, 165, 709-726.
- Cios, K. J., Swiniarski, R. W., Pedrycz, W., & Kurgan, L. A. (2007). "The Knowledge Discovery Process" [Bilgi Keşif Süreci]. In *Data Mining* (pp. 9-24). Springer, Boston, MA.
- Cover, T., & Hart, P. (1967). "Nearest Neighbor Pattern Classification" [En Yakın Komşu Örüntü Sınıflandırması]. *IEEE Transactions on Information Theory*, 13(1), 21-27.

- Çoban T. (2018). *Sinemada Yapay Zeka* (Yüksek lisans tezi). Ordu: Ordu Üniversitesi
- Denizli, Z. (2019). *Veri Madenciliği Modelleri Ve Örnek Bir Uygulama*. (Yüksek Lisans Tezi). İzmir: Dokuz Eylül Üniversitesi.
- Dåderman, A., & Rosander, S. (2018). Evaluating Frameworks For Implementing Machine Learning In Signal Processing: A Comparative Study of CRISP-DM, semma and KDD. KTH, School of Electrical Engineering and Computer Science (EECS).
- Doğaner, A. (2020). *Topluluk Öğrenme Yöntemleri İle Renal Hücreli Karsinom'un Tahmin Edilmesi* (Doktora Tezi). Malatya: İnönü Üniversitesi.
- Dasarathy, B. V., & Sheela, B. V. (1979). "A Composite Classifier System Design: Concepts And Methodology" [Bir Bileşik Sınıflandırıcı Sistem Tasarımı: Kavramlar ve Metodoloji]. *Proceedings of the IEEE*, 67(5), 708-713.
- Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). "A Review on Time Series Forecasting Techniques for Building Energy Consumption" [Bina Enerji Tüketimi için Zaman Serisi Tahmin Teknikleri Üzerine Bir İnceleme]. *Renewable and Sustainable Energy Reviews*, 74, 902-924.
- Deng, L. (2014). Dong Yu and others. "Deep Learning: Methods and Applications" [Derin Öğrenme: Yöntemler ve Uygulamalar]. *Foundations and Trends R in Signal Processing*, 7(3-4), 197-387.
- De Ville, B. (2013). "Decision trees" [Karar ağaçları]. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(6), 448-455.
- Dhar, V. (2013). "Data Science And Prediction" [Veri Bilimi Ve Tahmin]. *Communications of the ACM*, 56(12), 64-73.
- Dietterich, T. G. (2000). "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization" [Karar Ağaçları Topluluğu Oluşturmak için Üç Yöntemin DeneySEL Karşılaştırması: Bagging, Boosting Ve Randomization]. *Machine learning*, 40(2), 139-157.

- Ding, Z., Wang, G., Liu, Z., & Long, R. (2017). “Research on Differences in The Factors Influencing The Energy-Saving Behavior of Urban And Rural Residents in China—A Case Study of Jiangsu Province” [Çin’de Kentsel ve Kırsal Alanda Yaşayanların Enerji Tasarrufu Davranışlarını Etkileyen Faktörlerdeki Farklılıklar Üzerine Araştırma—Jiangsu Eyaleti Vaka Çalışması]. *Energy policy*, 100, 252-259.
- Doğaner, A., Colak, C., Küçükdurmaz, F., & Ölmez, C. (2021). “Prediction of Renal Cell Carcinoma Based on Ensemble Learning Methods” [Toplu Öğrenme Yöntemlerine Dayalı Renal Hücreli Karsinom Tahmini]. *Middle Black Sea Journal of Health Science*, 7(1), 104-114.
- Dong, B., Li, Z., Rahman, S. M., & Vega, R. (2016). “A Hybrid Model Approach for Forecasting Future Residential Electricity Consumption” [Gelecekteki Konut Elektrik Tüketimini Tahmin Etmek için Hibrit Bir Model Yaklaşımı]. *Energy and Buildings*, 117, 341-351.
- Ekmekçi H.İ. (2019), Türkiye’nin İşgücü Yapısının Özelliklerinin İncelenmesi: Lojistik Regresyon ve Karar Ağacı Uygulaması (Yüksek Lisans Tezi). Ankara: Hacettepe Üniversitesi.
- Enerji Görünümü; TSKB; 2020. <https://www.tskb.com.tr>.
- Eraslan, D. (2019). Experimenting With Some Data Mining Techniques To Establish Pediatric Reference Intervals for Clinical Laboratory Tests / Tıbbi Laboratuvar Testlerinin Pediatrik Referans Aralıklarının Belirlenmesi için Bazı Veri Madenciliği Teknikleri ile Deneyler (Yüksek Lisans Tezi). İzmir: Dokuz Eylül Üniversitesi.
- Ercan, U., Irmak, S., Çevik, K. K., & Canbazoglu, E. (2020). “Yapay Sinir Ağları Kullanılarak Konutlarda Elektrik Tüketimi Düzeylerinin Tahmin Edilmesi”. *Sosyoekonomi*, 28(46), 173-186.
- Erdoğan, Z. (2019). *Türkiye’de Lisans Düzeyindeki İstatistik Ders İçeriklerinin Veri Madenciliği Yöntemleri İle Analizi*. (Yüksek Lisans Tezi). Adana: Çukurova Üniversitesi.

- Erdoğan, Z., & Namlı, E. (2019). “A Living Environment Prediction Model Using Ensemble Machine Learning Techniques Based on Quality of Life Index” [Yaşam Kalitesi Endeksine Dayalı Topluluk Makine Öğrenimi Tekniklerini Kullanan Bir Yaşam Ortamı Tahmin Modeli]. *Journal of Ambient Intelligence and Humanized Computing*, 1-17.
- Enerji ve Tabii Kaynaklar Bakanlığı, <https://enerji.gov.tr/enerji-isleri-genel-mudurlugu-denge-Tabloları>, 2021.
- Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996, August). “Knowledge Discovery and Data Mining: Towards a Unifying Framework” [Bilgi Keşfi ve Veri Madenciliği: Birleştirici Bir Çerçeveye Doğru]. In *KDD* (Vol. 96, pp. 82-88).
- Firth, S., Lomas, K., Wright, A., & Wall, R. (2008). “Identifying Trends In The Use Of Domestic Appliances From Household Electricity Consumption Measurements” [Ev Elektrik Tüketimi Ölçümlerinden Ev Aletleri Kullanımındaki Eğilimlerin Belirlenmesi]. *Energy and Buildings*, 40(5), 926-936.
- Flores, V., & Keith, B. (2019). “Gradient Boosted Trees Predictive Models for Surface Roughness In High-Speed Milling in The Steel and Aluminum Metalworking Industry” [Çelik Ve Alüminyum Metal İşleme Endüstrisinde Yüksek Hızlı Frezelemede Yüzey Pürüzlülüğü için Gradyan Artırmalı Ağaçlar Öngörücü Modeller]. *Complexity*, 2019.
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). “Knowledge Discovery In Databases: An Overview” [Veritabanlarında Bilgi Keşfi: Genel Bir Bakış]. *AI magazine*, 13(3), 57-57.
- Freund, Y., & Schapire, R. E. (1996, July). “Experiments With A New Boosting Algorithm” [Yeni Bir Boosting Algoritması ile Deneyler]. In *icml* (Vol. 96, pp. 148-156).
- Freund, Y. (1997). Yoav Freund and Robert E Schapire: “A Decision-Theoretic Generalization of On-Line Learning to Boosting” [Çevrimiçi Öğrenmenin Boosting’e Yönelik Bir Karar-Teorik Genellemesi]. *Journal of Computer and System Science*, 55, 119-139.

- Friedman, J. H. (2001). "Greedy Function Approximation: A Gradient Boosting Machine" [Greedy Fonksiyon Yaklaşımı: Bir Gradyan Artırma Makinesi]. *Annals of statistics*, 1189-1232.
- Getahun, A. (2020). *Telecom Voice Traffic Termination Fraud Detection Using Ensemble Learning: The Case of Ethio Telecom/ Topluluk Öğrenme Kullanarak Telekom Ses Trafikçi Sonlandırma Dolandırıcılık Tespiti: Ethio Telekom Örneği* (Doktora Tezi), Ethiopia: St. Mary's University.
- Ghasemi, F., Neysiani, B. S., & Nematbakhsh, N. (2020, April). "Feature Selection in Pre-Diagnosis Heart Coronary Artery Disease Detection: A Heuristic Approach For Feature Selection Based on Information Gain Ratio and Gini Index" [Ön Tanı Kalp Koroner Arter Hastalığı Tespitinde Özellik Seçimi: Bilgi Kazanım Oranı ve Gini İndeksine Dayalı Özellik Seçimi için Sezgisel Bir Yaklaşım]. In *2020 6th International Conference on Web Research (ICWR)* (pp. 27-32). IEEE.
- Gilbert, C., Browell, J., & McMillan, D. (2019). "Leveraging Turbine-Level Data for Improved Probabilistic Wind Power Forecasting" [Gelişmiş Olasılıklı Rüzgar Enerjisi Tahmini için Türbin Düzeyinde Verilerden Yararlanma]. *IEEE Transactions on Sustainable Energy*, 11(3), 1152-1160.
- Goldstein, B. A., Navar, A. M., & Carter, R. E. (2017). "Moving Beyond Regression Techniques in Cardiovascular Risk Prediction: Applying Machine Learning to Address Analytic Challenges" [Kardiyovasküler Risk Tahmininde Regresyon Tekniklerinin Ötesine Geçmek: Analitik Zorlukların Üstesinden Gelmek için Makine Öğrenimini Uygulamak]. *European Heart Journal*, 38(23), 1805-1814.
- Gullo, F. (2015). "From Patterns in Data to Knowledge Discovery: What Data Mining Can Do" [Verideki Modellerden Bilgi Keşfine: Veri Madenciliği Ne Yapabilir?]. *Physics Procedia*, 62, 18-22.
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). "Development And Validation of A Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs" [Retina Fundus Fotoğraflarında Diyabetik Retinopati Tespiti için Bir Derin Öğrenme Algoritmasının Geliştirilmesi ve Doğrulanması]. *Jama*, 316(22), 2402-2410.

- Guyon, I., & Elisseeff, A. (2003). “An Introduction to Variable and Feature Selection” [Değişken ve Özellik Seçimine Giriş]. *Journal of Machine Learning Research*, 3(Mar), 1157-1182.
- Coşkun, F., & Gülleroğlu, H. D. (2021). “Geçmişten Günümüze Yapay Zekanın Gelişimi ve Eğitim Alanında Kullanılması”. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 1, 20.
- Gürsakar, N. (2018). Makine Öğrenmesi, Bursa: Dora, 1.
- Hache, E., Leboullenger, D., & Mignon, V. (2017). “Beyond Average Energy Consumption In The French Residential Housing Market: A Household Classification Approach” [Fransız Konut Piyasasında Ortalama Enerji Tüketiminin Ötesinde: Bir Hane Sınıflandırması Yaklaşımı]. *Energy Policy*, 107, 82-95.
- Hadri, S., Naitmalek, Y., Najib, M., Bakhouya, M., Fakhri, Y., & Elaroussi, M. (2019). “A Comparative Study of Predictive Approaches for Load Forecasting in Smart Buildings” [Akıllı Binalarda Yük Tahmini için Tahmine Dayalı Yaklaşımların Karşılaştırmalı Bir Çalışması]. *Procedia Computer Science*, 160, 173-180.
- Håkansson, A. (2013). “Portal of Research Methods And Methodologies for Research Projects And Degree Projects” [Araştırma Projeleri Ve Derece Projeleri için Araştırma Yöntemleri ve Metodolojileri Portalı]. In *The 2013 World Congress in Computer Science, Computer Engineering, and Applied Computing WORLDCOMP 2013; Las Vegas, Nevada, USA, 22-25 July* (pp. 67-73). CSREA Press USA.
- Hanehalkı Bütçe Araştırması Tüketim Harcamaları Birleşik Mikro Veri Seti, 2016-2017-2018; TÜİK; 2020. <https://www.tuik.gov.tr>.
- Hansen, L. K. ve Salamon, P. (1990). “Neural Network Ensembles” [Sinir Ağı Toplulukları]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10), 993-1001.
- Hassanat, A. B., Abbadi, M. A., Altarawneh, G. A., & Alhasanat, A. A. (2014). “Solving The Problem of The K Parameter in The KNN Classifier Using An Ensemble Learning Approach” [Bir Topluluk Öğrenme Yaklaşımı Kullanarak KNN

Sınıflandırıcısındaki K Parametresi Problemini Çözme]. *arXiv preprint arXiv:1409.0919*.

Hoque, M. S., Jamil, N., Saharudin, S. A., & Amin, N. (2020, August). “Energy Usage Prediction for Smart Home With Regression Based Ensemble Model” [Regresyon Tabanlı Topluluk Modeli ile Akıllı Ev için Enerji Kullanımı Tahmini]. In *2020 8th International Conference on Information Technology and Multimedia (ICIMU)* (pp. 378-383). IEEE.

Ij, H. (2018). “Statistics Versus Machine Learning” [İstatistiğe Karşı Makine Öğrenmesi]. *Nat Methods*, 15(4), 233.

Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). “Artificial Neural Networks: A Tutorial” [Yapay Sinir Ağları: Bir Öğretici]. *Computer*, 29(3), 31-44.

Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). “Supervised Machine Learning Algorithms: Classification And Comparison” [Denetimli Makine Öğrenimi Algoritmaları: Sınıflandırma ve Karşılaştırma]. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), 128-138.

Jones, R. V., Fuertes, A., & Lomas, K. J. (2015). “The Socio-Economic, Dwelling and Appliance Related Factors Affecting Electricity Consumption in Domestic Buildings” [Konut Binalarında Elektrik Tüketimini Etkileyen Sosyo-Ekonomik, Konut ve Cihazla İlgili Faktörler]. *Renewable and Sustainable Energy Reviews*, 43, 901-917.

Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons.

Karaaslan, A., Adar, T., & Delice, E. K. (2022). “Regional Evaluation of Renewable Energy Sources in Turkey by New İntegrated AHP-MARCOS Methodology: A Real Application” [Yeni Entegre Edilen AHP-MARCOS Metodolojisi İle Türkiye'deki Yenilenebilir Enerji Kaynaklarının Bölgesel Değerlendirmesi: Gerçek Bir Uygulama]. *International Journal of Sustainable Energy*, 41(2), 103-125.

- Karjalainen, S. (2007). "Gender Differences In Thermal Comfort and Use of Thermostats in Everyday Thermal Environments" [Termal Konforda Cinsiyet Farklılıkları ve Günlük Termal Ortamlarda Termostat Kullanımı]. *Building and environment*, 42(4), 1594-1603.
- Kaza, N. (2010). "Understanding The Spectrum of Residential Energy Consumption: A Quantile Regression Approach" [Konut Enerji Tüketimi Yelpazesini Anlamak: Bir Nicel Regresyon Yaklaşımı]. *Energy policy*, 38(11), 6574-6585.
- Kuncheva, L. I. (2014). *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley & Sons.
- Leahy, E., & Lyons, S. (2010). "Energy Use and Appliance Ownership in Ireland" [İrlanda'da Enerji Kullanımı ve Cihaz Sahipliği]. *Energy Policy*, 38(8), 4265-4279.
- Schowe, B. (2011, June). "Feature Selection for High-Dimensional Data With Rapidminer" [Rapidminer ile Yüksek Boyutlu Veriler için Özellik Seçimi]. In *Proceedings of the 2nd RapidMiner Community Meeting And Conference (RCOMM 2011)*, Aachen.
- Lévy, J. P., & Belaïd, F. (2018). "The Determinants of Domestic Energy Consumption In France: Energy Modes, Habitat, Households and Life Cycles" [Fransa'da Ev İçi Enerji Tüketiminin Belirleyicileri: Enerji Modları, Habitat, Haneler Ve Yaşam Döngüleri]. *Renewable and Sustainable Energy Reviews*, 81, 2104-2114.
- Li, K., Xie, X., Xue, W., Dai, X., Chen, X., & Yang, X. (2018). "A Hybrid Teaching-Learning Artificial Neural Network for Building Electrical Energy Consumption Prediction" [Elektrik Enerjisi Tüketimi Tahmini Oluşturmak için Hibrit Bir Öğretme-Öğrenme Yapay Sinir Ağı]. *Energy and Buildings*, 174, 323-334.
- Liu, Z., Wu, D., Liu, Y., Han, Z., Lun, L., Gao, J., ... & Cao, G. (2019). "Accuracy Analyses and Model Comparison Of Machine Learning Adopted in Building Energy Consumption Prediction" [Bina Enerji Tüketimi Tahmininde Benimsenen Makine Öğreniminin Doğruluk Analizleri ve Model Karşılaştırması]. *Energy Exploration & Exploitation*, 37(4), 1426-1451.

- Mahesh, B. (2020). "Machine Learning Algorithms-A Review" [Makine Öğrenimi Algoritmaları-Bir İnceleme]. *International Journal of Science and Research (IJSR)*. [Internet], 9, 381-386.
- Markus, M. (2000). "LOF: Identifying Density-Based Local Outliers" [LOF: Yoğunluk Tabanlı Yerel Aykırı Değerleri Tanımlama]. In *ACM SIGCOM 2000 Int. Conf. On Management of Data, Dallas, TX*. (ss. 93-104). New York: ACM Press.
- McCarthy, J. (2004). What Is Artificial Intelligence. URL: <http://www-formal.stanford.edu/jmc/whatisai.html>. (Erişim Tarihi: 20.05.2022).
- McLoughlin, F., Duffy, A., & Conlon, M. (2015). "A Clustering Approach to Domestic Electricity Load Profile Characterisation Using Smart Metering Data" [Akıllı Ölçüm Verilerini Kullanarak Yerel Elektrik Yüğü Profili Karakterizasyonuna Yönelik Bir Kümeleme Yaklaşımı]. *Applied Energy*, 141, 190-199.
- Mitchell, T. M. (2006). *The Discipline of Machine Learning* (Vol. 9). Pittsburgh: Carnegie Mellon University, School of Computer Science, Machine Learning Department.
- Morgan, J. N. ve Sonquist, J. A. (1963). "Problems in The Analysis of Survey Data, and A Proposal" [Anket Verilerinin Analizinde Yaşanan Sorunlar ve Öneri]. *Journal of the American Statistical Association*, 58(302), 415-434.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT press.
- Nabavi, S. A., Aslani, A., Zaidan, M. A., Zandi, M., Mohammadi, S., & Hossein Motlagh, N. (2020). "Machine Learning Modeling for Energy Consumption of Residential and Commercial Sectors" [Konut ve Ticari Sektörlerin Enerji Tüketimi için Makine Öğrenmesi Modellemesi]. *Energies*, 13(19), 5171.
- Rao, M. N., & Reddy, B. S. (2007). "Variations In Energy Use By Indian Households: An Analysis of Micro Level Data" [Hintli Haneler Tarafından Enerji Kullanımındaki Farklılıklar: Mikro Düzeyde Verilerin Analizi]. *Energy*, 32(2), 143-153.
- Nazeriye, M., Haeri, A., Haghighat, F., & Panchabikesan, K. (2021). "Understanding The Influence of Building Characteristics on Enhancing Energy Efficiency in Residential Buildings: A Data Mining Based Study" [Konut Binalarında Enerji

- Verimliliğinin Artırılmasında Bina Özelliklerinin Etkisinin Anlaşılması: Veri Madenciliğine Dayalı Bir Çalışma]. *Journal of Building Engineering*, 43, 103069.
- Olson, D. L., & Delen, D. (2008). Performance Evaluation for Predictive Modeling. in *Advanced Data Mining Techniques* (pp. 137-147). Springer, Berlin, Heidelberg.
- Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., & Ajayi, S. (2022). “Building Energy Consumption Prediction for Residential Buildings Using Deep Learning And Other Machine Learning Techniques” [Derin Öğrenme ve Diğer Makine Öğrenimi Tekniklerini Kullanarak Konut Binaları için Enerji Tüketimi Tahmini Oluşturma]. *Journal of Building Engineering*, 45, 103406.
- Onan, A. Korukoğlu, S., & Bulut, H. (2017). “A Hybrid Ensemble Pruning Approach Based on Consensus Clustering and Multi-Objective Evolutionary Algorithm for Sentiment Classification” [Duygu Sınıflandırması için Fikir Birliği Kümeleme ve Çok Amaçlı Evrimsel Algoritmaya Dayalı Hibrit Bir Topluluk Budama Yaklaşımı]. *Information Processing & Management*, 53(4), 814-833.
- Özcan, K. M., Gülay, E., & Üçdoğruk, Ş. (2013). “Economic and Demographic Determinants of Household Energy Use In Turkey” [Türkiye’de Hanehalkı Enerji Kullanımının Ekonomik ve Demografik Belirleyicileri]. *Energy Policy*, 60, 550-557.
- Özdemir, O. (2020). *Gelir Ve Yaşam Koşulları Verisinin Veri Madenciliği Yaklaşımı ile İncelenmesi ve Model Performanslarının Değerlendirilmesi* (Doktora Tezi). İstanbul: Yıldız Teknik Üniversitesi.
- Pal, A. (2020). <https://medium.com/>
- Pallegedara, A., & Mottaleb, K. A. (2021). “Exploring Choice and Expenditure on Energy For Domestic Works By The Sri Lankan Households: Implications For Policy” [Sri Lankalı Hanelerin Ev İşleri İçin Seçim ve Enerji Harcamalarını Araştırmak: Politika için Çıkarımlar]. *Energy*, 222, 119899.
- Panthong, R., & Srivihok, A. (2015). “Wrapper Feature Subset Selection for Dimension Reduction Based on Ensemble Learning Algorithm” [Topluluk Öğrenme

Algoritmasına Dayalı Boyut Küçültme için Sarmalayıcı Özellik Alt Kümesi Seçimi]. *Procedia Computer Science*, 72, 162-169.

Pham, A. D., Ngo, N. T., Truong, T. T. H., Huynh, N. T., & Truong, N. S. (2020). “Predicting Energy Consumption in Multiple Buildings Using Machine Learning for Improving Energy Efficiency And Sustainability” [Enerji Verimliliğini ve Sürdürülebilirliği İyileştirmek için Makine Öğrenimini Kullanarak Birden Fazla Binada Enerji Tüketimini Tahmin Etme]. *Journal of Cleaner Production*, 260, 121082.

Polikar, R., & Polikar, R. (2006). “Ensemble Based Systems in Decision Making” [Karar Vermede Topluluk Tabanlı Sistemler]. *IEEE Circuit Syst. Mag.* 6, 21–45 (2006).

Polikar, R. (2012). *Ensemble Learning*. Springer, Boston, MA.

Poloczek, J., Treiber, N. A., & Kramer, O. (2014). “KNN Regression As Geo-Imputation Method For Spatio-Temporal Wind Data” [Mekansal-Zamansal Rüzgar Verileri İçin Coğrafi Değerlendirme Yöntemi Olarak KNN Regresyonu]. In *International Joint Conference SOCO'14-CISIS'14-ICEUTE'14* (pp. 185-193). Springer, Cham.

Popenici, S. A. D., & Kerr, S. (2017). “Exploring The Impact of Artificial Intelligence On Teaching And Learning in Higher Education” [Yapay Zekanın Yüksek Öğretimde Öğretme ve Öğrenme Üzerindeki Etkisini Araştırmak]. *Res Pract Technol Enhanc Learn* 12 (1): 22.

Diamantini, C., Genga, L., Marozzo, F., Potena, D., & Trunfio, P. (2017, August). “Discovering Mobility Patterns of Instagram Users Through Process Mining Techniques” [Süreç Madenciliği Teknikleri ile Instagram Kullanıcılarının Hareketlilik Modellerini Keşfetme]. In *2017 IEEE International Conference on Information Reuse and Integration (IRI)* (pp. 485-492). IEEE.

Behera, B., & Ali, A. (2016). “Household Energy Choice and Consumption İntensity: Empirical Evidence from Bhutan” [Hanehalkı Enerji Seçimi ve Tüketim Yoğunluğu: Bhutan'dan Ampirik Kanıtlar]. *Renewable and Sustainable Energy Reviews*, 53, 993-1009.

- Das, S., De Groote, H., & Behera, B. (2014). “Determinants of Household Energy Use in Bhutan” [Bhutan'da Ev Enerji Kullanımının Belirleyicileri]. *Energy*, 69, 661-672.
- Rapidminer, <https://rapidminer.com/>, erişim tarihi; mayıs 2022.
- Robinson, C., Dilkina, B., Hubbs, J., Zhang, W., Guhathakurta, S., Brown, M. A. ve Pendyala, R. M. (2017). “Machine Learning Approaches for Estimating Commercial Building Energy Consumption” [Ticari Bina Enerji Tüketimini Tahmin Etmek için Makine Öğrenimi Yaklaşımları]. *Applied Energy*, 208, 889-904.
- Rokach, L. (2010). *Pattern Classification Using Ensemble Methods* (Vol. 75). World Scientific.
- Kabari, L. G., & Onwuka, U. C. (2019). “Comparison of Bagging and Voting Ensemble Machine Learning Algorithm as a Classifier” [Bir Sınıflandırıcı Olarak Torbalama ve Oylama Topluluğu Makine Öğrenimi Algoritmasının Karşılaştırılması]. *International Journals of Advanced Research in Computer Science and Software Engineering*, 9(3), 19-23.
- Salari, M. ve Javid, R. J. (2017). “Modeling Household Energy Expenditure in The United States” [Amerika Birleşik Devletleri'nde Hanehalkı Enerji Harcamalarının Modellenmesi]. *Renewable and Sustainable Energy Reviews*, 69, 822-832.
- Schapire, R. E. (1990). “The Strength of Weak Learnability” [Zayıf Öğrenilebilirliğin Gücü]. *Machine learning*, 5(2), 197-227.
- Seker, S. E., & Ocak, I. (2019). “Performance Prediction of Roadheaders Using Ensemble Machine Learning Techniques” [Toplu Makine Öğrenimi Tekniklerini Kullanarak Yol Başlıklarının Performans Tahmini]. *Neural Computing and Applications*, 31(4), 1103-1116.
- Shafique, U., & Qaiser, H. (2014). “A Comparative Study of Data Mining Process Models (KDD, CRISP-DM And SEMMA)” [Veri Madenciliği Süreç Modellerinin (KDD, CRISP-DM Ve SEMMA) Karşılaştırmalı Bir Çalışması]. *International Journal of Innovation and Scientific Research*, 12(1), 217-222.

- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.
- Şeker, S. E. (2008). <https://bilgisayarkavramlari.com/>
- Şeker, Ş. E. (2018). CRISP-DM: Endüstriler Arası Standart İşleme–Veri Madenciliği için (Cross Industry Standard Processing–Data Mining). *YBS Ansiklopedi*, 5(2).
- Taşkın, G. G. E. V. Ç. (2005). Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6(2), 221-239.
- Temel, G. O., Erdoğan, S., & Ankaralı, H. (2012). Sınıflama Modelinin Performansını Değerlendirmede Yeniden Örnekleme Yöntemlerinin Kullanımı. *Bilişim Teknolojileri Dergisi*, 5(3), 1-8.
- Tonn, B., & Eisenberg, J. (2007). “The Aging US Population and Residential Energy Demand” [Yaşlanan ABD Nüfusu ve Konut Enerji Talebi]. *Energy policy*, 35(1), 743-745.
- Tso, G. K., & Yau, K. K. (2003). “A Study Of Domestic Energy Usage Patterns in Hong Kong” [Hong Kong’da Yerel Enerji Kullanım Kalıpları Üzerine Bir Araştırma]. *Energy*, 28(15), 1671-1682.
- Tso, G. K., & Yau, K. K. (2007). “Predicting Electricity Energy Consumption: A Comparison of Regression Analysis, Decision Tree and Neural Networks” [Elektrik Enerjisi Tüketimini Tahmin Etme: Regresyon Analizi, Karar Ağacı ve Sinir Ağlarının Karşılaştırılması]. *Energy*, 32(9), 1761-1768.
- Uçar, M. K., Nour, M., Sindi, H., & Polat, K. (2020). “The Effect of Training and Testing Process on Machine Learning in Biomedical Datasets” [Biyomedikal Veri Setlerinde Eğitim ve Test Sürecinin Makine Öğrenmesine Etkisi]. *Mathematical Problems in Engineering*, 2020.
- Valenzuela, C., Valencia, A., White, S., Jordan, J. A., Cano, S., Keating, J., ... & Potter, L. B. (2014). “An Analysis of Monthly Household Energy Consumption Among Single-Family Residences in Texas, 2010” [Teksas'ta Tek Ailelik Konutlar Arasında Aylık Hane Enerji Tüketiminin Bir Analizi, 2010]. *Energy Policy*, 69, 263-272.

- Vringer, K., Aalbers, T., & Blok, K. (2007). "Household Energy Requirement and Value Patterns" [Hanehalkı Enerji İhtiyacı ve Değer Kalıpları]. *Energy Policy*, 35(1), 553-566.
- Wang, F., Lu, X., Chang, X., Cao, X., Yan, S., Li, K., ... & Catalão, J. P. (2022). "Household Profile Identification for Behavioral Demand Response: A Semi-Supervised Learning Approach Using Smart Meter Data" [Davranışsal Talep Yanıtı için Hane Profili Tanımlaması: Akıllı Sayaç Verilerini Kullanan Yarı Denetimli Bir Öğrenme Yaklaşımı]. *Energy*, 238, 121728.
- Wang, H., & Ma, J. (2011). Wang G., Hao J., Ma J., Jiang H. "A Comparative Assessment Of Ensemble Learning for Credit Scoring" [Kredi Puanlaması için Topluluk Öğreniminin Karşılaştırmalı Bir Değerlendirmesi]. *Expert Systems with Applications*, 38(1), 223-230.
- Xiang, L., Xie, T., & Xie, W. (2020). "Prediction Model of Household Appliance Energy Consumption Based On Machine Learning" [Makine Öğrenimine Dayalı Ev Aletleri Enerji Tüketimi Tahmin Modeli]. In *Journal of Physics: Conference Series* (Vol. 1453, No. 1, p. 012064). IOP Publishing.
- Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., & Ahrentzen, S. (2018). "Random Forest Based Hourly Building Energy Prediction" [Random Forest Bazlı Saatlik Bina Enerji Tahmini]. *Energy and Buildings*, 171, 11-25.
- Wassie, Y. T., Rannestad, M. M., & Adaramola, M. S. (2021). "Determinants of Household Energy Choices in Rural Sub-Saharan Africa: An Example From Southern Ethiopia" [Kırsal Sahra Altı Afrika'da Hane Halkı Enerji Tercihlerinin Belirleyicileri: Güney Etiyopya'dan Bir Örnek]. *Energy*, 221, 119785.
- Wiesmann, D., Azevedo, I. L., Ferrão, P., & Fernández, J. E. (2011). "Residential Electricity Consumption in Portugal: Findings from Top-Down And Bottom-Up Models" [Portekiz'de Konut Elektrik Tüketimi: Yukarıdan Aşağıya ve Aşağıdan Yukarıya Modellerden Elde Edilen Bulgular]. *Energy Policy*, 39(5), 2772-2779.
- Windeatt, T., & Ardeshir, G. (2004). "Decision Tree Simplification for Classifier Ensembles" [Sınıflandırıcı Toplulukları için Karar Ağacı

- Basitleştirmesi]. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(05), 749-776.
- Wirth, R., & Hipp, J. (2000, April). "CRISP-DM: Towards A Standard Process Model for Data Mining" [CRISP-DM: Veri Madenciliği için Standart Bir Süreç Modeline Doğru]. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29-39).
- Wolpert, D. H. (1992). "Stacked Generalization" [Stacking Genelleştirme]. *Neural networks*, 5(2), 241-259.
- Wyatt, P. (2013). "A Dwelling-Level Investigation into The Physical and Socio-Economic Drivers of Domestic Energy Consumption in England" [İngiltere'de Ev İçi Enerji Tüketiminin Fiziksel Ve Sosyo-Ekonomik İtici Güçlerine İlişkin Konut Düzeyinde Bir Araştırma]. *Energy Policy*, 60, 540-549.
- Xu, Y., & Goodacre, R. (2018). "On Splitting Training And Validation Set: A Comparative Study of Cross-Validation, Bootstrap And Systematic Sampling for Estimating The Generalization Performance of Supervised Learning" [Eğitim Ve Doğrulama Setini Bölme Hakkında: Denetimli Öğrenmenin Genelleme Performansını Tahmin Etmek için Çapraz Doğrulama, Önyükleme ve Sistematik Örneklemenin Karşılaştırmalı Bir Çalışması] . *Journal of Analysis and Testing*, 2(3), 249-262.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). "Systematic Review Of Research On Artificial Intelligence Applications In Higher Education—Where Are The Educators?" [Yükseköğretimde Yapay Zeka Uygulamalarına İlişkin Araştırmaların Sistemik Olarak Gözden Geçirilmesi – Eğitimciler Nerede?]. *International Journal of Educational Technology in Higher Education*, 16(1), 1-27.
- Zhang, C., & Ma, Y. (Eds.). (2012). *Ensemble Machine Learning: Methods and Applications*. Springer Science & Business Media.
- Zhang, J., Li, X., Jiang, W., Wang, Y., Li, C., Wang, Q., & Rao, S. (2005, August). "A Novel Ensemble Decision Tree Approach for Mining Genes Coding İon Channels

for Cardiopathy Subtype” [Kardiyopati Alt Tipi için İyon Kanallarını Kodlayan Madencilik Genleri için Yeni Bir Topluluk Karar Ağacı Yaklaşımı]. In *International Conference on Fuzzy Systems and Knowledge Discovery* (pp. 852-860). Springer, Berlin, Heidelberg.

Zhao, C. M., & Luan, J. (2006). “Data Mining: Going beyond Traditional Statistics” [Veri Madenciliği: Geleneksel İstatistiklerin Ötesine Geçmek]. *New Directions for Institutional Research*, 131, 7-16.

Zhou, Z. H. (2012). *Ensemble Methods: Foundations and Algorithms*. CRC press.

Zhou, Z. H. (2009, June). “When Semi-Supervised Learning Meets Ensemble Learning” [Yarı Denetimli Öğrenme Toplu Öğrenmeyle Buluştuğunda]. In *International Workshop on Multiple Classifier Systems* (pp. 529-538). Springer, Berlin, Heidelberg.

Zins, C. (2007). “Conceptual Approaches for Defining Data, Information, and Knowledge” [Veri, Bilgi ve Özbilgiyi Tanımlamaya Yönelik Kavramsal Yaklaşımlar]. *Journal of the American Society for Information Science and Technology*, 58(4), 479-493.

Zou, B., & Luo, B. (2019). “Rural Household Energy Consumption Characteristics and Determinants in China” [Çin’de Kırsal Hane Enerji Tüketimi Özellikleri ve Belirleyicileri]. *Energy*, 182, 814-823.

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı ve Soyadı	Esma KESRİKLİOĞLU
Doğum Yeri ve Yılı	
Çalıştığı Kurum	
Ünvan	
Eğitim Durumu	
Lisans Öğrenimi	
Tezsiz Yüksek Lisans	
Yüksek lisans	
Doktora Öğrenimi	
İletişim	
Email	