

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**KAYIP DOĞAL GAZ TÜKETİMİNİN MAKİNE ÖĞRENMESİ İLE
TESPİTİ: SAKARYA ÖRNEĞİ**

YÜKSEK LİSANS TEZİ

Kürşad ARSLAN

Bilgisayar Mühendisliği Anabilim Dalı

NİSAN 2023

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**KAYIP DOĞAL GAZ TÜKETİMİNİN MAKİNE ÖĞRENMESİ İLE
TESPİTİ: SAKARYA ÖRNEĞİ**

YÜKSEK LİSANS TEZİ

Kürşad ARSLAN

Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı: Dr. Öğr. Üyesi Muhammed Fatih ADAK

Ortak Danışman: Dr. Öğr. Üyesi Mustafa AKPINAR

NİSAN 2023

Kürşad ARSLAN tarafından hazırlanan “Kayıp Doğal Gaz Tüketiminin Makine Öğrenmesi ile Tespiti: Sakarya Örneği” adlı tez çalışması 13.04.2023 tarihinde aşağıdaki jüri tarafından oy birliği ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı : **Dr. Öğr. Üyesi**
Sakarya Üniversitesi

Ortak Danışman : **Dr. Öğr. Üyesi**
BAE/Higher Colleges of Technology

Jüri Üyesi : **Dr. Öğr. Üyesi**
Sakarya Üniversitesi

Jüri Üyesi : **Dr. Öğr. Üyesi**
Sakarya Uygulamalı Bilimler Üniversitesi

Jüri Üyesi : **Dr. Öğr. Üyesi**
Sakarya Uygulamalı Bilimler Üniversitesi



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “Kayıp Doğal Gaz Tüketiminin Makine Öğrenmesi ile Tespiti: Sakarya Örneği” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

13/04/2023

Kürşad ARSLAN





Bilim camiasına ve meslektaşlarıma



TEŞEKKÜR

Yüksek lisans öğrenimim boyunca bilgi ve tecrübeleri ile bana destek olarak yol gösteren danışman hocalarım Sayın Dr. Mustafa AKPINAR ve Sayın Dr. Öğr. Üyesi Muhammed Fatih ADAK'a;

Tez çalışmasında kullanılan verilerin temin edilmesi noktasında katkılarından dolayı danışman hocam Sayın Dr. Mustafa AKPINAR'a, Adapazarı Gaz Dağıtım A.Ş.'ye ve Türkiye Cumhuriyeti Çevre, Şehircilik ve İklim Değişikliği Bakanlığı Meteoroloji Genel Müdürlüğü'ne;

Öğrenim hayatım boyunca her daim yanımda olan saygıdeğer aileme;

Teşekkürlerimi ve saygılarımı sunarım.

Kürşad ARSLAN



İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	ix
İÇİNDEKİLER	xi
KISALTMALAR	xiii
SİMGELER	xv
TABLO LİSTESİ	xvii
ŞEKİL LİSTESİ	xix
ÖZET	xxi
SUMMARY	xxiii
1. GİRİŞ	1
1.1. Çalışmanın Ana Hatları	2
2. KAYNAK ARAŞTIRMASI	5
2.1. Literatür Çalışması	5
2.2. Doğal Gaz ve Tüketimi	9
2.3. Doğal Gaz Tüketimini Okuma Yöntemleri	10
2.3.1. Körüklü doğal gaz sayaçları	12
2.3.2. Rotary doğal gaz sayaçları	13
2.3.3. Türbin doğal gaz sayaçları	13
2.4. Kayıp ve hesaplanamayan doğal gaz	13
3. TÜKETİM TAHMİNİNDE KULLANILAN VERİLER VE YÖNTEMLER 17	17
3.1. Veriseti	17
3.1.1. Doğal gaz tüketim verileri.....	17
3.1.2. Abone ve sayaç işlemleri verileri.....	19
3.1.3. Meteorolojik veriler	19
3.1.4. Takvimsel olaylar veri seti	20
3.1.5. Veri setlerinin birleştirilmesi	21
3.2. Kullanılan Yöntemler	22
3.2.1. Yapay sinir ağları	24
3.2.1.1. Eğitici öğrenme	25
3.2.1.2. Eğitici öğrenme.....	29
3.2.1.3. Takviyeli öğrenme	30
3.2.1.4. İleri beslemeli hesaplama.....	31
3.2.1.5. Geri yayılım algoritması	36
3.2.2. Performans ölçüm metrikleri	39
3.2.2.1. Ortalama hata (ME).....	39
3.2.2.2. Ortalama yüzdesel hata (MPE)	40
3.2.2.3. Ortalama mutlak hata (MAE).....	40
3.2.2.4. Ortalama kare hata (MSE)	40
3.2.2.5. Kök ortalama kare hata (RMSE).....	40
3.2.2.6. Ortalama mutlak yüzdesel hata (MAPE)	41
4. TÜKETİM TAHMİNİ VE SONUÇLAR.....	43

4.1. Modelleme.....	43
4.1.1. Abone gruplarının ve veri sayılarının belirlenmesi.....	43
4.1.2. Modellerin oluşturulması ve kullanılan yaklaşım	44
4.1.3. Modellerin eğitilmesi ve tüketim tahminleri.....	46
4.2. Tahmin Sonuçlarının Elde Edilmesi.....	49
5. SONUÇ VE ÖNERİLER.....	61
5.1. Kayıp ve Hesaplanamayan Tüketime Sahip Abonelerin Belirlenmesi	61
KAYNAKLAR.....	67
ÖZGEÇMİŞ.....	73



KISALTMALAR

AGDAŞ	: Adapazarı Gaz Dağıtım Anonim Şirketi
ARIMA	: Bütünleşik otoregresif hareketli ortalama
ARIMAX	: Dışsal değişkenli bütünleşik otoregresif hareketli ortalama
ARX	: Otoregresif dışsal
EPDK	: Türkiye Cumhuriyeti Enerji Piyasası Düzenleme Kurumu
LSTM	: Uzun – kısa süreli bellek
MAPE	: Ortalama mutlak yüzdesel hata
ME	: Ortalama hata
MPE	: Ortalama yüzdesel hata
MSE	: Ortalama kare hata
MSSQL	: Microsoft SQL Server
NLME	: Doğrusal olmayan karma etkiler modeli
RMS-A	: Gaz basıncı düşürme ve ölçüm istasyonu
RMSE	: Kök ortalama kare hata
SVM	: Destek vektör makinesi
UFG	: Kayıp ve hesaplanamayan doğal gaz
YSA	: Yapay sinir ağı



SİMGELER

m^3	: Metreküp
kWh	: Kilovatsaat
$kcal$	: Kilokalori [joule]
Q	: Doğal gaz debisi [m^3/h]
Q_{min}	: Minimum doğal gaz debisi [m^3/h]
Q_{max}	: Maksimum doğal gaz debisi [m^3/h]
TG	: Elde edilen ısı enerjisi [kcal]
H_u	: Doğal gazın alt ısı değeri [kcal/ m^3]
P_{mut}	: Mutlak gaz basıncı
η	: Doğal gazı yakan cihazın verimi
X	: Model girdilerinin kümesi
Y	: Model çıktıların kümesi
T	: Beklenen çıktıların kümesi
x_i	: Girdi verisi
y_i	: Çıktı verisi
t_i	: Beklenen çıktı verisi
N	: Veri seti eleman sayısı
ϵ	: Hata değeri
θ	: Eşik değeri
w_i	: Ağırlık değeri
NET	: Nöronun çıktı değeri
b_i	: Bias değeri
OUT	: Transfer fonksiyonunun çıktı değeri
e	: Euler sabiti
φ	: Transfer fonksiyonu
ϵ_{Toplam}	: Toplam hata miktarı
λ	: Öğrenme katsayısı
α	: Momentum katsayısı
∂	: Kısmi türev işlemi
T_{ort}	: Abonenin ortalama tüketim miktarı [m^3]

T_i	: Abonenin bir aydaki tüketim miktarı [m ³]
$T_{Eğitim}$	: Eğitim veri kümesi
TSI	: Abonenin toplam sayaç işlem sayısı
$Ti_{[s]}$	: Abonenin sayaç işlem vektörü
SI	: Abonenin sayaç indisi vektörü



TABLO LİSTESİ

Sayfa

Tablo 2.1. Türkiye’de Nisan 2021-Nisan 2022 dönemindeki doğal gaz tüketiminin sektör dağılımı [27].	10
Tablo 3.1. Doğal gaz tüketim veri seti örnek veriler.	18
Tablo 3.2. Abone ve sayaç işlemleri veri seti örnek veriler.	19
Tablo 3.3. Meteoroloji veri seti örnek veriler.	20
Tablo 3.4. Takvimsel olaylar veri seti örnek veriler.	21
Tablo 3.5. Nihai veri seti değişkenler.	22
Tablo 4.1. Abonelerin fatura sayıları.	43
Tablo 4.2. Abone grupları.	44
Tablo 4.3. Modellerin çalışma süreleri.	50
Tablo 4.4. Eğitim ve test veri setlerinin gruplar bazında ortalama MAPE ve düzeltilmiş R^2 değerleri.	52
Tablo 4.5. Bağlı modelin daha başarılı olduğu abone örnekleri.	53
Tablo 4.6. Noktasal modelin daha başarılı olduğu abone örnekleri.	55
Tablo 5.1. Tüm ve azaltılmış verinin regresyon katsayıları.	63
Tablo 5.2. Tüm ve azaltılmış verinin tanımlayıcı istatistikleri.	63



ŞEKİL LİSTESİ

Sayfa

Şekil 2.1. Türkiye'nin 2011 – 2021 yılları arası doğal gaz tüketimi [26].	10
Şekil 2.2. Körüklü doğal gaz sayacı şematik gösterimi [28].	12
Şekil 3.1. Biyolojik sinir hücresinin sembolik gösterimi.	23
Şekil 3.2. Örnek bir yapay sinir ağı mimarisi.	25
Şekil 3.3. Eğitici öğrenme mimarisi.	26
Şekil 3.4. Sınıflandırma işlemi.	27
Şekil 3.5. Regresyon işlemi.	28
Şekil 3.6. Eğitici öğrenme mimarisi.	29
Şekil 3.7. Kümeleme yaklaşımı görseli.	30
Şekil 3.8. Takviyeli öğrenme mimarisi.	30
Şekil 3.9. Basamak fonksiyonu grafiği.	32
Şekil 3.10. Doğrusal fonksiyon grafiği.	33
Şekil 3.11. Log-sigmoid fonksiyon grafiği.	33
Şekil 3.12. Tan-sigmoid fonksiyon grafiği.	34
Şekil 3.13. Yapay sinir hücresi.	35
Şekil 3.14. Geri yayılım algoritması mimarisi.	37
Şekil 3.15. Hata eğrisi üzerindeki minimum noktalar.	38
Şekil 4.1. Çalışmada kullanılan YSA mimarisi.	47
Şekil 4.2. Sayaç işlemlerinin bağıl zamansal gösterimi.	48
Şekil 4.3. Sayaç işlemlerinin noktasal gösterimi.	49
Şekil 4.4. 9 numaralı gruba ait örnek bir abone için bağıl ve noktasal modelin örnek çıktı grafikleri.	54
Şekil 4.5. 2 numaralı gruba ait örnek bir abone için bağıl ve noktasal modelin örnek çıktı grafikleri.	55
Şekil 4.6. 9 numaralı gruba ait örnek bir abone için bağıl ve noktasal modelin örnek çıktı grafikleri.	56
Şekil 4.7. 4 numaralı gruba ait örnek bir abone için bağıl ve noktasal modelin örnek çıktı grafikleri.	56
Şekil 4.8. A durumuna sahip aboneler için bağıl ve noktasal modelde MAPE değişimleri (a) Değişim miktarı en yüksek ilk 50 abone (b) Değişim miktarı en düşük 50 abone.	58
Şekil 4.9. B durumuna sahip aboneler için bağıl ve noktasal modelde MAPE değişimleri (a) Değişim miktarı en yüksek ilk 50 abone (b) Değişim miktarı en düşük ilk 50 abone.	59
Şekil 5.1. Modeller ile gerçekleştirilen işlemlerin adımları.	62
Şekil 5.2. Azaltılmış verinin MAPE farklarının dağılımı.	64
Şekil 5.3. Abonelerin belirlenen bölgeler üzerinde dağılımı.	65



KAYIP DOĞAL GAZ TÜKETİMİNİN MAKİNE ÖĞRENMESİ İLE TESPİTİ: SAKARYA ÖRNEĞİ

ÖZET

Enerji, dünya üzerindeki ülkelerin en fazla harcama yaptıkları alanlar arasında bulunmaktadır. Enerjiye ulaşabilmek kadar tedarik edilen veya üretilen enerji kaynaklarının yönetimi ve korunumu da oldukça önem arz etmektedir. Bu tez çalışması kapsamında, dünyadaki en önemli enerji kaynaklarının başında gelen doğal gazın hesapsız bir biçimde tüketimi tespit edilmeye çalışılmıştır. Bu durum, bir dağıtım şebekesinden tüketicilere gönderilen gaz miktarı ile tüketicilerin faturalarına yansıyan tüketim miktarları arasındaki fark olarak tanımlanmaktadır ve hem doğal gaz tedariki sağlayan firmaları hem de ülkeleri ekonomik olarak ciddi zararlara uğratmaktadır. İletim borularındaki sızıntılar, hava sıcaklığının değişimi, aktif olmayan tükeçiler, ölçüm hataları ve arızalı olabilecek sayaçlar bu duruma neden olarak gösterilebilmektedir. Bu çalışmada, sayaçlardan kaynaklı olabilecek kayıp doğal gaz tüketimi üzerinde yoğunlaşmıştır.

Bu çalışmada, her bir doğal gaz abonesinin tüketimlerini tahmin eden modeller oluşturulmuştur ve gerçekleştirilen tahminlerden yola çıkarak tüketimlerinde anormallik olan aboneler belirlenmeye çalışılmıştır. Çalışma sahası olarak Sakarya ilinde mevcut olan ve tüketim zamanları 2011-2021 arasında değişkenlik gösteren 307.570 adet abonenin doğal gaz tüketim verileri kullanılmıştır ve aboneler tüketim zamanı büyüklüklerine göre gruplara ayrılmıştır. Doğal gaz tüketiminde önemli ölçüde belirleyici olan meteorolojik durumları belirten veri seti ve ilgili tüketim dönemleri arasındaki hafta sonları, resmî tatil ve bayramların bilgisini içeren takvimsel olaylar veri seti kullanılarak kapsamlı bir veri ambarı oluşturulmuştur. Oluşturulan bu veri ambarı, model eğitimi ve test işlemlerinde kullanılmıştır. Abonelerin tüketimlerini tahmin etmek için çok katmanlı yapay sinir ağı modeli geliştirilmiştir. Her bir abonenin bir yıllık tüketim verileri ile ilgili abonelerin modelleri eğitilmiş ve geri kalan dönemlerin tüketimleri model tarafından tahmin edilmiştir. Çalışmada, her bir doğal gaz abonesi için iki farklı model geliştirilmiştir. Bunlar, doğal gaz tüketim verileri arasında bulunan ve her bir fatura dönemi içerisinde ilgili abonenin sayacının yenilenip yenilenmediği bilgisinin kullanılmasıyla gerçekleştirilmiştir. Bu bilgiden yararlanılarak sayacı yenilenen abonelerin sonraki tüketimlerinin yenilenmiş sayaç ile gerçekleştirildiği bilgisinin ağı verildiği model oluşturulmuştur ve “bağıl model” olarak adlandırılmıştır. Diğer model ise sayaç yenilenme işlemlerinin noktasal olarak belirtildiği, yani sayaç bilgisinin ağı verilmediği modeldir ve “noktasal model” olarak adlandırılmıştır. Her abonenin belirtilen bu iki model ile de eğitim ve test işlemleri gerçekleştirilerek doğal gaz tüketim tahminleri gerçekleştirilmiştir. Yapılan tahminlerin performansları, MAPE (ortalama mutlak yüzdesel hata) yöntemi kullanılarak değerlendirilmiştir. Geliştirilen modeller ile gerçekleştirilen tahminlerde bağıl modelde ortalama MAPE değeri 2,42 olurken noktasal modelde 2,12 olmuştur.

İki modelin tahmin performansları deęerlendirilerek kullanılan sayaların doęal gaz tüketimleri üzerine olan etkileri belirlenmiştir. Bu bağlamda tüm abonelerin iki modelden elde edilen MAPE deęerlerinin dağılımları istatistiksel yöntemler kullanılarak incelenmiş ve dağılım üzerinde normal tüketim davranışlarına sahip olan abonelerin sınırları belirlenmiştir. Ardından normal sınırlar dışında kalan aboneler iki alt gruba ayrılmıştır. Bunlar ara-geiş bölgesi ve anormal bölge olarak isimlendirilmiştir. Elde edilen sınır deęerleri kullanılarak anormal tüketim davranışlarına sahip olan aboneler belirlenmiştir ve bu bölgede bulunan abonelerin kayıp ve hesaplanamayan tüketime sahip olma riskinin yüksek olduęu belirtilmiştir.



DETECTION OF LOST NATURAL GAS CONSUMPTION WITH MACHINE LEARNING: SAKARYA CASE

SUMMARY

Energy is among the areas where countries around the world spend the most. The management and protection of supplied or produced energy resources is as important as accessing energy. Within the scope of this thesis, it was tried to determine the unaccounted consumption of natural gas, one of the world's most important energy sources. This situation is defined as the difference between the amount of gas sent to consumers from a distribution network and the amount of consumption reflected on the bills of consumers, and it causes serious economic losses to both the companies providing natural gas supply and the countries. Leaks in the transmission pipes, change in air temperature, inactive consumers, measurement errors and possibly defective meters can be shown as the cause of this situation. Intentional physical interventions and aging of the meters over time can be given as causes of gas meter failures. Interventions made by subscribers are mostly examined under "natural gas theft". This study focuses on the loss of natural gas consumption that may be caused by the meters.

In this study, models estimating the consumptions of each natural gas subscriber were created and it was tried to determine the subscribers with abnormal consumptions based on the estimates made. As the study area, natural gas consumption data of 307.570 subscribers in the province of Sakarya, whose consumption times vary between 2011 and 2021, were used and the subscribers were divided into groups according to their consumption time sizes. A comprehensive data warehouse has been created by using the data set that indicates the meteorological conditions, which are important determinants in natural gas consumption, and the calendar events dataset that includes the information on weekends, public holidays, and holidays between the relevant consumption periods. This created data warehouse was used in model training and testing processes. A multi-layer artificial neural network model has been developed to predict the consumption of subscribers. The models of the subscribers were trained on the consumption data of each subscriber for one year, and the consumption of the remaining periods was estimated by the model. In the study, two different models were developed for each natural gas subscriber. These were realized by using the information, which is among the natural gas consumption data, whether the relevant subscriber's meter is renewed or not within each billing period. Using this information, the model in which the network is informed that the next consumption of the subscribers whose meters are renewed is realized with the renewed meter was created and it is called the "relative model". The other model is the one in which the meter renewal processes are specified pointwise, that is, the meter information is not given to the network, and it is called the "point model". Natural gas consumption estimations were made by performing training and testing processes for each subscriber with these two models. The performances of the predictions were evaluated using the MAPE (mean absolute percentile error) method.

In the predictions made with the developed models, the mean MAPE value was 2.42 in the relative model and 2.12 in the point model.

The estimation performances of the two models were evaluated and the effects of the meters used on natural gas consumption were determined. In this context, the distribution of MAPE values of all subscribers obtained from the two models was examined using statistical methods, and the limits of subscribers with normal consumption behaviors on the distribution were determined. Then, the subscribers outside the normal limits were divided into two subgroups. These are named intermediate-transition zone and abnormal zone. By using the obtained limit values, subscribers with abnormal consumption behaviors were determined and it was stated that the subscribers in this region had a high risk of having lost and unaccounted gas consumption.



1. GİRİŞ

Enerji arz, talep ve güvenliği günümüzde önemli konulardan biridir. Dünya üzerinde otomotiv, makine, tekstil, inşaat, tarım, savunma sanayi ve haberleşme gibi büyük sektörler doğrudan enerji sektörü ile ilişkilidir. Ülkelerin üretim, istihdam, yatırım ve kalkınma hedeflerini gerçekleştirebilmeleri kendi enerjilerini üretebilmelerine ve daha da önemlisi üretilen-tedarik edilen enerjinin yönetimini en uygun biçimde sağlamasına bağlıdır.

Ülkemizde ve dünyada enerji kaynaklarına kolay ulaşımın yaygınlaşması özellikle evsel tüketiciler ve üretim sektöründeki enerji talebini hızla arttırmaktadır. Bu noktada enerji arz güvenliği ve enerji tedarigi konuları üzerine çalışmalar artmış, enerji planlaması gelişmiş ve gelişmekte olan ülkeler için önemli bir konu haline gelmiştir. Enerji planlamasının en önemli konu başlıklarından biri de talep tahmindir.

Enerji tedariginde en önemli kaynaklardan biri de doğal gazdır. Günümüzde konutlarda ısıtma amacıyla en fazla kullanılan, temiz ve güvenli olarak bilinen bir enerji kaynağı olan doğal gaz, 1988 yılı itibariyle ülkemizde kullanılmaya başlamıştır ve tüketimi her geçen gün artmaktadır [1]. Yapılan araştırmalara ve tüketim tahminlerine göre ülkemizin 2022 yılı için doğal gaz tüketimi bir önceki yıla göre %20 artmış ve 60.44.873.569 metreküpe yükselmiştir [2].

Bu talebe karşılık doğal gazın tedarigi ve yönetiminin yanında kayıp doğal gaz tüketiminin de önüne geçilmesi gerekmektedir. Kayıp doğal gaz veya diğer adı ile hesaplanamayan gaz, bir dağıtım sistemine enjekte edilen gaz ile müşterilerin sayaçlarında ölçülen doğal gaz tüketim miktarının arasındaki fark olarak tanımlanmaktadır [3]. Bu durumun başlıca sebepleri, doğal gaz iletim borularındaki sızıntı, kullanıcılar tarafından sayaçlara yapılan kasıtlı fiziksel müdahaleler, dağıtım şirketlerinin hataları ve doğal gaz tüketicilerinin sayaçlarında zaman içerisinde meydana gelen arızalar olarak gösterilmektedir [4].

Tüketicilerin sayaçlarında meydana gelen arızalar sebebiyle ölçülen tüketim miktarı, gerçek tüketim miktarından daha düşük çıkabilmektedir. Bu durum ise öncelikle gaz dağıtım şirketlerini ve ülkeleri zarara uğratmaktadır. Bu nedenle gaz dağıtım

řirketleri talep tahmininde öngörölen tüketim miktarının üzerinde bir miktarda gaz alımı yapmak zorundadırlar.

Doğal gaz tüketiminin ve kayıp doğal gaz tüketiminin büyük bölümünü oluşturan konut sektöründe bu kayıpları azaltmak güvenlik ve ekonomik açıdan oldukça önemlidir. Gaz dağıtım řirketleri kayıp doğal gaz tüketimini tespit etmek ve bu durumun önüne geçmek için periyodik olarak konutlardaki sayaçları kontrol etmek, kalibrasyon işlemi yapmak veya gerekiyorsa gaz ölçüm sayacını değiřtirmek için personel görevlendirmektedir. Ancak bu işlem dağıtım řirketleri için maliyetleri yükseltmektedir. Kullanıcı davranışları ve çevresel faktörler göz önünde bulundurularak iyi analiz edilmiş bir tüketim tahmini ile doğal gaz tüketimindeki anomali tespit edilebilir ve kayıp doğal gaz tüketimine sahip olan aboneler ortaya çıkarılabilir.

Bu tez çalışmasında ölçülemeyen veya hatalı ölçölen doğal gaz sebebiyle meydana gelen kayıplar yapay zeka ve istatistiksel teknikler kullanılarak tespit edilmeye çalışılmıştır. Çalışmada Sakarya ili örnek olarak ele alınmıştır. Çalışmadaki amaç sayaç değışimlerinin, tüketim ve tüketim tahminleri üzerindeki etkilerini belirleyerek, ölçüm hatası olan sayaçları belirleyebilmektir.

1.1. Çalışmanın Ana Hatları

Tez çalışması 5 bölümden oluşturulmuştur. İkinci bölümde, literatürde enerji tüketim tahminine yönelik yapılan çalışmalar incelenmiştir. Bölümün ikinci kısmında doğal gaz ile ilgili genel tanımlamalar, dünyada ve ölkemizde doğal gaz ve tüketimi, doğal gaz tüketimini okuma yöntemleri ve bu çalışmanın odaklandığı konu olan kayıp - hesaplanamayan- doğal gaz konuları incelenmiştir.

Üçüncü bölümde, tez çalışmasında kullanılan veri seti ve içerdığı değışkenler detaylıca anlatılarak veri setinin hazırlanması aşamalarına değinilmiştir. Tüketim tahmininde kullanılan yöntemler ve gerçekleşen tahminlerin performanslarının değeriendirildiğı metrikler yine bu kısımda açıklanmıştır.

Dördüncü bölümde, doğal gaz abonelerinin modellenmesinde kullanılan yaklaşım gösterilmiştir. Önerilen modelleme yaklaşımları kullanılarak her bir abone için tüketim tahmini işlemleri gerçekleştirilmiş ve sonuçları incelenmiştir.

Beşinci bölümde, dördüncü bölümde açıklanan model ve yaklaşımdan elde edilen sonuçlar incelenmiştir. Bu sonuçlara göre tezin amacına yönelik olarak ulaşılan bulgular ve sonraki benzer çalışmalar için öneriler verilmiştir.





2. KAYNAK ARAŞTIRMASI

2.1. Literatür Çalışması

Bu çalışmada, tüketimi hesaplanamayan aboneler ve hatalı okumaları gerçekleştiren sayaçlar tespit edilmeye çalışılmıştır. Bu çalışmanın paralelinde literatürde yapılan çalışmalar çok azdır. Bunlardan bazıları aşağıda verilmiştir.

A. Soltanisarvestani ve ark. tarafından yapılan bir çalışmada İran’da bulunan örnek evsel doğal gaz tüketicilerin kayıp ve hesaplanamayan doğal gaz tüketimleri tespit edilmeye çalışılmıştır. Abonelerin tüketim verileri ve dış ortam sıcaklığı kullanılarak yapay sinir ağı modelleri eğitilmiş ve eğitime işleminde parçacık sürü optimizasyon algoritması kullanılmıştır. Modellerden elde edilen çıktılar, abonelerin gerçek tüketim miktarları ile karşılaştırılarak olası kayıp ve hesaplanamayan doğal gaz tüketime sahip olan aboneler belirlenmeye çalışılmıştır [5]. Bahsedilen bu çalışmada mevcut abone verilerinden olası kayıp tüketimlerin hesaplanmaya çalışıldığı gibi tespiti daha önce gerçekleştirilen kayıp tüketim vakaları üzerinde de çalışmalar yapılmıştır. L. Botev ve P. Johnson tarafından yapılan çalışmada, Birleşik Krallık’ta mevcut olan farklı miktarlardaki iki eksik okuma vakaları ele alınmıştır. Çalışmada doğal gaz iletim sisteminden elde edilen kayıp ve hesaplanamayan doğal gaz zaman serisi verileri kullanılmıştır. Mevsimsellik etkisinin bulunduğu verilere istatistiksel süreç kontrol teknikleri uygulanmıştır ve önerilen bu yöntemin hataları tespit etmek için kullanılabileceği vurgulanmıştır [3]. A. Soltanisarvestani ve A. Safavi tarafından yapılan çalışmada, kayıp ve hesaplanamayan doğal gaz tüketimi vakalarını tespit edebilmek için doğal gaz tüketimine etki eden faktörlere ve parametrelere odaklanılmışlardır. Abonelerin tüketim davranışlarını etkileyen faktörler gruplara ayrılmıştır ve bu faktörlerin tüketimlere olan etkileri bulanık bilişsel haritalama yöntemi kullanılarak belirtilmiştir. Kategorize edilen ve etkileri belirlenen faktörlerin iyi analiz edilerek dahil edilen bir model yardımıyla doğal gaz ölçüm cihazının doğruluğu hakkında fikir sahibi olunabileceğini belirtmişlerdir. Model tahminleri ile gerçekleşen ölçüm değerleri arasında önemli ölçüde fark bulunan aboneler kayıp ve hesapsız doğal gaz tüketimine sahip olma olasılığının yüksek olduğu verilmiştir [6].

M. Shafiq ve ark. kayıp ve hesaplanamayan doğal gazı sebep olan faktörleri belirleyebilmek için yüz doğal gaz tüketicisine anket gerçekleştirmişlerdir. Belirlenen her bir faktöre hipotez testi uygulayarak faktörlerin etki ağırlığı belirlenmiştir ve kayıp tüketimleri kontrol edebilmek için matematiksel model geliştirmişlerdir. Elde edilen sonuçlar değerlendirildiğinde ve saha personelleri tarafından gerekli işlemler gerçekleştirildiğinde kayıp tüketim miktarları yeniden hesaplanmıştır. Bu hesaplama sonucunda kayıp ve hesaplanamayan doğal gazın aylık bazda %17 oranda azaldığı sonucunu belirtmişlerdir [7]. F. Arpino ve ark. yaptıkları çalışmada, iletim şebekelerinden ve tüketim dengeleme sistemlerinden kaynaklanan kayıp ve hesaplanamayan doğal gaz üzerine odaklanmışlardır. Farklı iletim şebekelerindeki kayıp doğal gaz miktarlarını yıllık ve aylık bazda istatistiksel yöntemler kullanarak analiz etmişlerdir. Elde edilen hesapsız tüketim miktarının büyük bir bölümünün mevsimsel etkilerin ve sistematik ölçüm hatalarından kaynaklandığını belirtmişlerdir [8].

Bu çalışmaların temelde doğal gaz talep tahmini gerçekleştirdiği görülmüştür. Literatürde doğal gaz talep tahmini üzerine birçok çalışma yapılmıştır.

M. Akpınar, ev tipi doğal gaz tüketicilerinin enerji tüketim tahminine yönelik yaptığı çalışmada ARIMA ve çoklu regresyon yöntemlerini kullanmıştır. Yöntemleri, konjonktürel etkiden arındırılmış ve konjonktürel etkiden arındırılmamış modellere uygulamıştır ve ilgili yöntemlerin bu modeller üzerinde verdiği sonuçları karşılaştırmıştır. Konjonktürel etkiden arındırılmış model üzerinden gerçekleştirilen doğal gaz tüketim tahmininin diğer modele göre daha başarılı sonuçlar verdiği görülmüştür [9]. M. Akpınar tarafından yapılan başka bir çalışmada, gün öncesi ve yıl öncesi olmak üzere iki farklı model üzerinden doğal gaz talep tahmini gerçekleştirilmiştir. Yapılan bu çalışmada tek değişken olarak tüketilen miktar ele alınmıştır. İstatistik ve yapay zeka tekniklerinin kullanıldığı bu çalışmada, günlük veri ile gerçekleştirilen doğal gaz talep tahmininin aylık veri ile yapılan tahmine göre daha küçük hata değerine sahip olduğu belirtilmiştir [10]. N. Wei ve ark. günlük doğal gaz tüketim tahmininde kullanılmak üzere 56 günlük tüketimi içeren üç farklı model geliştirmişlerdir. Yapılan bu çalışmada eksik verinin ve verinin karmaşıklık seviyesinin yapılan tahminler üzerinde büyük bir etkisi olduğu vurgulanmıştır [11]. A. Anđelković ve D. Bajatović çalışmalarında uyarlamalı ağ tabanlı bulanık çıkarım algoritması kullanmışlardır. Diğer çalışmalarda kullanılan doğal gaz tüketim miktarı

ve hava durumu parametrelerinin haricinde saatlik hava tahmini üretip bu değeri de modele eklemişlerdir. Model çıktısı olarak bir saat önceden doğal gaz tüketimini tahmin etmeye çalışmışlardır. Model çıktılarının performanslarını ortalama mutlak yüzde hata değerini dikkate alarak değerlendirmişlerdir. Bu değerlendirmeler sonucunda önerilen modelin etkili bir yaklaşım olduğu belirtilmiştir [12]. W. Qiao ve ark. kısa dönemli doğal gaz tüketim tahmini gerçekleştirmişlerdir. Tüketim verilerine gauss yumuşatma yöntemini uygulamışlardır ve balina optimizasyon algoritmasını kullanmışlardır. Tüketim tahminini gerçekleştirmek için ise uyarlanabilir volterra filtreleme algoritmasını önermişlerdir. Algoritmanın parametrelerini balina optimizasyon algoritması ile belirlemişlerdir. Çalışmada kullanılan bu hibrit modelin kısa dönemli doğal gaz tüketim tahmininde doğru sonuçlar verdiğini belirtmişlerdir [13]. H. Lu ve ark. doğal gaz tüketimini tahmin etmek için arps düşüş eğrisi modelini kullanmışlardır. Önerilen modelin parametrelerini optimize etmek için gri kurt optimizasyon yöntemini kullanmışlardır. Oluşturulan bu hibrit modeli farklı sektörlerde bulunan doğal gaz tüketicilerinin veri setlerine uygulamışlardır. Önerilen modelin farklı türdeki veri setleri üzerinde doğru tahminler yaptığını belirtmişlerdir [14]. O. Laib ve ark. yaptıkları bir çalışmada çok katmanlı yapay sinir ağıları modellerini kullanmışlardır. İlgili model, bir sonraki günün doğal gaz tüketim miktarını tahmin etmek için kullanılmıştır. Kullanıcılar, tüketim davranışlarına göre kümeleme yaklaşımıyla gruplandırılmıştır. Daha sonra ilgili veri kümeleri uzun kısa vadeli bellek (LSTM) modelleri ile doğal gaz tüketimini tahmin etmede kullanılmıştır. Önerilen bu yöntemin özellikle tüketim davranışlarında ani değişimler olan kullanıcıların doğal gaz tüketim miktarlarını tahmin etme konusunda etkinliğinin yüksek olduğu belirtilmiştir [15]. J. Song ve ark. günlük doğal gaz tüketiminin tahminini gerçekleştirebilmek için mevsimsel ayrıştırma ve zamansal evrişimli ağ modelini kullanarak hibrit bir model önermişlerdir. Kullanılan bu yöntemin performansını, destek vektör makineleri (SVM) ve regresyon modelleri ile karşılaştırmışlardır. Bu karşılaştırma sonucunda önerilen modelin diğer algoritmalara göre daha başarılı sonuçlar verdiğini belirtmişlerdir [16]. M. Brabec ve ark. yaptıkları çalışmada günlük doğal gaz tüketimini kullanıcı bazlı tahmin etmeye çalışmışlardır. İlgili çalışmada, diğer çalışmalardan farklı olarak eksik tüketim verileri ve tüketimde ani değişiklik gösteren verileri de modele dahil etmişlerdir. Doğrusal olmayan karma etkiler modelini (NLME) kullanarak kullanıcı bazlı günlük doğal gaz tüketim tahmini gerçekleştirmişlerdir. Elde edilen sonuçlar ARIMAX ve

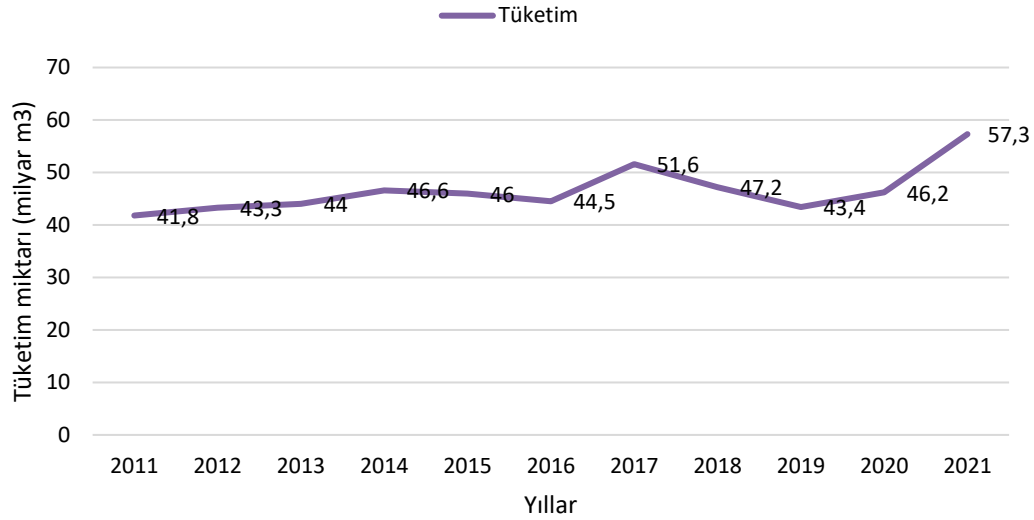
ARX yöntemleri ile elde edilen sonuçlar ile karşılaştırılmıştır [17]. M. Çelik tarafından yapılan bir çalışmada kullanıcı sayısı, birim fiyat ve sıcaklık girdileri dikkate alınarak doğal gaz tüketim tahmini gerçekleştirilmiştir. İlgili çalışmada yöntem olarak yapay sinir ağları ve çoklu regresyon yöntemleri kullanılmıştır. Oluşturulan aynı model üzerinde bu yöntemlerin etkinlikleri yorumlanmıştır [18]. S. Demirceylan, girdi parametreleri olarak sıcaklık, mevsim, ay ve birim fiyatı içeren modeli kullanarak yapay sinir ağları yöntemi ile doğal gaz tüketim miktarının tahmin edilmesine yönelik çalışma gerçekleştirmiştir. İlgili çalışmada tüketim tahmini sonuçları ve girdi parametrelerinin sonuç üzerinde ki etkileri tartışılmıştır [19]. E. Karaca, yaptığı çalışmada yapay sinir ağları ve regresyon yöntemlerini kullanarak doğal gaz tüketim tahmini gerçekleştirmiştir. Modelde mevsimler, abone sayısı, sıcaklık ve birim fiyat gibi parametreleri kullanmıştır. Model parametrelerine normalizasyon işlemi uygulamıştır ve modele uyguladığı yöntemler içerisinde yapay sinir ağlarının etkinliğinin daha yüksek olduğu verilmiştir [20]. Ö. Evcim, zaman serileri ve derin öğrenme yöntemleri kullanarak günlük doğal gaz tüketiminin tahminini gerçekleştirmiştir. Giriş değerleri olarak sıcaklık, rüzgâr ve basınç gibi hava durumunu içeren parametreleri ve tüketim miktarını içermektedir. Elde edilen sonuçlara göre yöntemlerin performansları karşılaştırılmış olup derin öğrenme modelleri ile daha doğru tahminlerde bulunduğu belirtilmiştir [1]. Y. Yılmaz, konik çok değişkenli uyarlanabilir regresyon eğrileri ve çok değişkenli uyarlanabilir regresyon eğrileri yöntemlerini kullanarak günlük doğal gaz tüketimini tahmin etme çalışması gerçekleştirmiştir. İlgili çalışmada bu iki modelin sonuçlarının karşılaştırılmasının yanı sıra, model eğitimi sırasında sıcaklık değerlerinin kullanımının da model çıktılarına etkisini mukayese etmiştir [21]. B. Akın, yapay sinir ağı modelini kullanarak bir yıllık doğal gaz tüketim miktarını tahmin etmiştir. Yapılan çalışmada on yıllık tüketim verisi ve meteorolojik veriler modelin eğitilmesi için kullanılırken, modelin test edilmesi için bir yıllık tüketim tahmini gerçekleştirilmiştir. Girdi olarak kullanılan parametrelerin tüketim üzerindeki etkisi tartışılmıştır. Yapay sinir ağları kullanılarak elde edilen sonuçlar regresyon modeli sonuçları ile karşılaştırılmıştır [22]. B. Taşkıner, doğal gaz tüketim tahmini üzerine yaptığı çalışmada model eğitimi için kullanılan giriş parametrelerinden sıcaklık değerinin model çıktısı üzerindeki etkinliğini incelemiştir. İki yıllık tüketim verilerinin eğitim için kullanıldığı bu çalışmada bir yıllık tüketimin tahmini sağlanmıştır. Model üzerinde farklı transfer fonksiyonları da kullanılarak yapay sinir

ağının performansı tartışılmıştır [23]. B. Kant tarafından yapılan çalışmada genetik algoritma, parçacık sürü optimizasyonu, yapay sinir ağları ve grup veri işleme yöntemi gibi teknikler kullanılarak doğal gaz tüketiminin tahmini yapılmıştır. Elde edilen tahmin sonuçları belirlenen performans kriterlerine göre karşılaştırılmış olup, en başarılı sonucu çok katmanlı yapay sinir ağı modelinin verdiği belirtilmiştir [24].

2.2. Doğal Gaz ve Tüketimi

Doğal gaz, dünya çapında tüketiminde en büyük artış yaşanan birincil enerji kaynağıdır. Dünya'daki bu enerji talebinin önümüzdeki dönemde artarak devam edeceği düşünülmektedir. Kullanımı birçok alanda hala yaygın olan kömüre göre kolay yanması, verimliliğinin yüksek oluşu, nakliyesinin kolay ve ayrıca çevre dostu olması doğal gazı bu yönleriyle öne çıkarmaktadır. Küresel doğal gaz tüketimi yıllık %2 - %2,5 oranında büyüme göstermektedir. Küresel doğal gaz üretimi ise yıllık %3 oranında artış göstererek diğer kaynaklar arasında en fazla artan enerji kaynağı olmaktadır [25]. Bunun nedeni olarak dünya nüfusunun artması, gelir seviyesinin yükselmesi ve sosyoekonomik durumun iyileştirilmesi olarak gösterilmektedir. Ağırlıklı olarak konutlarda ve organize sanayii bölgelerinde kullanılmaktadır. Sanayii bölgelerinde fabrikaların ihtiyaç duyduğu elektrik enerjisini jeneratörler yardımıyla üretmede kullanılırken; konutlarda odaları ısıtmak, yemek pişirmek ve sıcak su elde etmek amacıyla kullanılmaktadır.

Evsel kullanımın yanında ticari ve endüstriyel alanlardaki kullanımının yaygınlaşmasıyla Türkiye'nin doğal gaz talebi, ilk kullanılmaya başlandığından bugüne kadar sürekli artış göstermektedir. Bu talep artışının başlıca sebepleri teknolojik gelişmeler ile beraber ülkemizin sanayi bölgelerinde enerji ihtiyacının hızla artmasıdır (Şekil 2.1).



Şekil 2.1. Türkiye'nin 2011 – 2021 yılları arası doğal gaz tüketimi [26].

T.C. Enerji Piyasası Düzenleme Kurumu (EPDK)'nın Nisan 2022 numaralı raporunda belirtilen verilere göre Nisan 2021 – Nisan 2022 tarihleri arasında Türkiye'de tüketilen doğal gazın neredeyse yarısının konutlarda olduğu görülmektedir (Tablo 2.1). Bu veriler, doğal gazın Türkiye için birincil ve en fazla ihtiyaç duyduğu enerji kaynağı olduğunu göstermektedir. Türkiye bu ihtiyacının büyük bir kısmını dış ülkelere ithal ederek karşılamaktadır.

Tablo 2.1. Türkiye'de Nisan 2021-Nisan 2022 dönemindeki doğal gaz tüketiminin sektör dağılımı [27].

Sektör	Tüketimdeki payı (%)
Konut	45,40
Sanayi	22,82
Dönüşüm/Çevrim*	16,78
Hizmet	12,48
Enerji	2,06
Ulaşım	0,22

*: Elektrik üretim santralleri bu sektör içerisinde yer almaktadır.

2.3. Doğal Gaz Tüketimini Okuma Yöntemleri

Doğal gaz tüketiminin belirlenmesi için doğal gaz sayaçları kullanılır. Böylece kullanıcılara tükettikleri gaz miktarına göre bir fatura bedeli oluşturulur. Aynı şekilde gaz dağıtım şirketleri için ise aldıkları ve son kullanıcıya satışını gerçekleştirdikleri

gazın miktarını ölçmek için kullanılır. Konut sektörü ve sanayi gibi son kullanıcılar için yalnızca tükettikleri gaz miktarının ölçülmesi ve buna göre bir fatura bedelinin oluşturulması amaçlanırken; gaz dağıtım firmaları için satılan gaz, sayaçlar tarafından ölçülen gaz ve kayıp gaz miktarlarının toplamına eşittir. Bölüm 2.4'te kayıp ve hesaplanamayan doğal gaz konusuna daha detaylıca değinilecektir. Tüketimi gerçekleştirilen doğal gazın hacmi, sayaçlar tarafından m³ cinsiyle hesaplanır ve faturalandırılma aşamasında hesaplanan gaz hacmi ısı enerjisiye dönüştürülür. Hesaplanan bu kWh değeri üzerinden fatura bedeli oluşturulur. Yani gaz dağıtım şirketleri kullanıcılara belirli bir miktarda doğal gaz değil enerji satmaktadırlar.

Doğal gaz tüketimini okumak için birden çok sayaç çeşidi bulunmaktadır. Bu sayaçlar, tüketicinin türü ve dolayısıyla okunması gereken gaz hacmi göz önünde bulundurularak belirlenmektedir. Konutlarda ve düşük hacimli gazın okunmasında körüklü sayaçlar tercih edilir. Doğal gazın abone sayısı bazından en fazla tüketicisinin konut sektöründe olduğu düşünülürse, en yaygın kullanılan sayaç çeşidi körüklü sayaçlardır. Evsel kullanımda seçilecek olan körüklü gaz sayacının büyüklüğü ise bireysel tüketici ve merkezi ısıtma sistemlerinde farklılık göstermektedir. Sanayi ve doğal gazın şehirlere giriş noktaları gibi yüksek hacimde ve basınçta gaz geçişi gerçekleşen durumlarda körüklü sayaçlar yetersiz kalmaktadır. Bu durumlarda orifis, türbin ve rotary tipte sayaçlar kullanılmaktadır [28].

Sayaç büyüklüğünün saptanması işleminde ilgili sayacın kaç tane konuta doğal gaz sağladığı gibi durumlara bakılır. Bu noktada doğal gazın debisi terimi gündeme gelmektedir. Debi, bir bölgeden birim zamanda geçen miktar anlamına gelmektedir. Bir sayaçtaki gazın debisi, aşağıdaki eşitlik Denklem 2.1 kullanılarak hesaplanmaktadır [28].

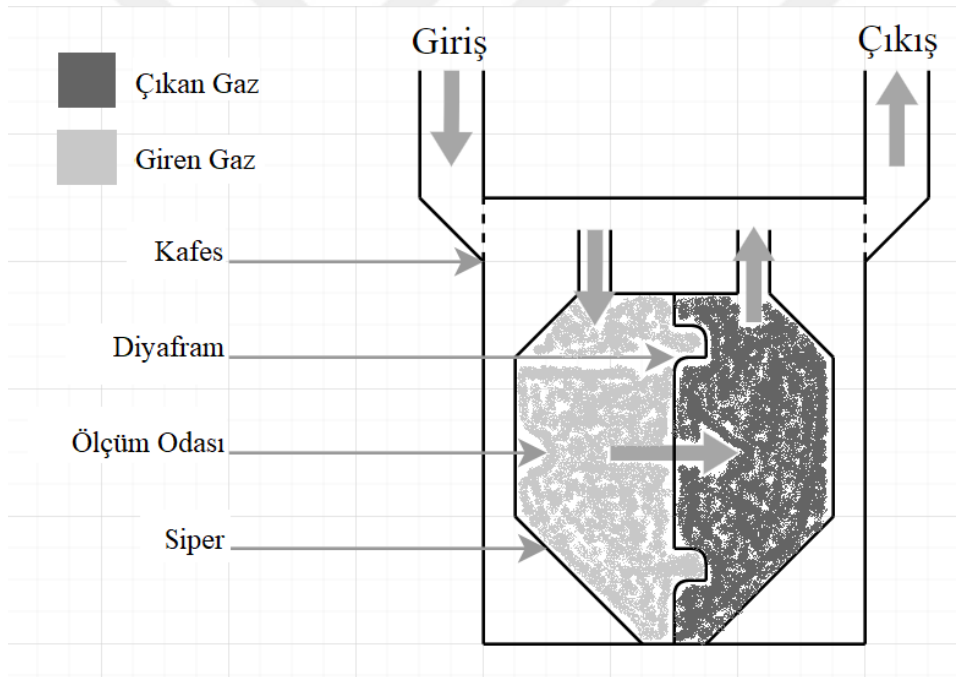
$$Q = TG/H_u \cdot P_{mut} \cdot \eta \quad (2.1)$$

Bu eşitlikte; Q sayaçta ki gazın debisi (m³/h), TG elde edilen ısı enerjisi (kcal), H_u doğal gazın alt ısı değeri (kcal/m³), P_{mut} mutlak gaz basıncı ve η ise doğal gazı yakan cihazın verimi olarak ifade edilmektedir.

2.3.1. Körüklü doğal gaz sayaçları

Körüklü doğal gaz sayaçları, çoğunlukla evsel tesisatlarda, düşük kapasiteli ölçümler yapmak amacıyla kullanılmaktadır. Ölçüm kapasiteleri ve büyüklükleri değişmekle birlikte birden fazla modelleri bulunmaktadır. Kullanım amacı ve konumuna göre uygun model seçilerek kullanılmaktadır.

Bu sayaçlar, hacimleri bilinen odacıklardan oluşur. Sayaca gelen gaz, basınç farkıyla bir odacıktan diğerine geçer. Bu gaz transfer işleminden sonra basınç farkı az öncekinin tam tersi yönünde olmaktadır. Tüm bu işlemlerde ileri ve geri sayan mekanik parçalar sayma işlemine devam etmektedir. Körük mekanizmasının hareketleri dişlileri harekete geçirir ve sayaç göstergesi çalışmaya devam eder [28]. Kaliteli ve doğrulama testleri gerçekleştirilmiş bir körüklü doğal gaz sayacı, uzun yıllar ölçüm hassasiyetini korumaktadır (Şekil 2.2).



Şekil 2.2. Körüklü doğal gaz sayacı şematik gösterimi [28].

Körüklü sayaçların belirli ölçümler için hata payları bulunmaktadır. Bunlar sırasıyla Q_{min} ve Q_{max} olarak ifade edilmektedir. Tüketilen gaz miktarı Q_{min} değerinin altında veya Q_{max} değerinin üstünde ise sayaç ya okuma yapamamakta, ya da hatalı okuma yapmaktadır. Bu durumlar, üretimi kaliteli bir biçimde gerçekleştirilmiş ve doğrulama testlerini tolere edilebilir hata payları ile tamamlamış olsa bile sayaçların uzun zaman kusursuz çalışacağı anlamına gelmemektedir. Evsel kullanımda körüklü sayaçlarda zaman içerisinde gövdenin eskimesinden dolayı kaçaklar meydana

gelebilmektedir. Bunun yanında, tüketicilerin kasıtlı olarak sayaçlara fiziksel müdahalesinden dolayı sayaçlar eksik ölçüm gerçekleştirebilir veya hiç ölçüm yapmayabilir. Bu gibi arıza durumlarında tüketiciler, olduğundan daha az veya daha yüksek fatura bedelleri ile karşılaşabilmektedir. Tez çalışmasında da temelde bu sayaçların hatalarının tespiti üzerine çalışılmıştır.

2.3.2. Rotary doğal gaz sayaçları

Rotary doğal gaz sayaçları, sanayi bölgelerinde ve endüstriyel alanlarda kullanılan sayaçlar arasındadır. İçerisinde bulunan rotorların dönme hareketiyle gaz hacminin hesaplanması prensibi ile çalışmaktadır. Rotorların hareketi mekanik dişlilere iletilir. Bu nedenle rotary sayaçlarda yağ kullanımı büyük önem taşımaktadır. Periyodik olarak yağ durumu kontrol edilmeli ve gerekiyorsa yağ ilave edilmelidir. Bu sayaçlar, rotorların hareketinden dolayı çalışırken ses ve titreşim oluşturmaktadır. Bu olay, istenmeyen bir durum olduğundan sessizliğin gerekli olduğu ortamlarda rotary sayaç kullanmak yerine büyük boyutlu körüklü sayaçlar da tercih sebebi olabilmektedir.

2.3.3. Türbin doğal gaz sayaçları

Türbin doğal gaz sayaçları da rotary sayaçlarda olduğu gibi sanayi bölgelerinde ve endüstriyel alanlarda kullanılan sayaçlardan biridir. Buna ek olarak büyük ölçekli sanayi bölgeleri, şehirlere doğal gazın dağıtıldığı bağlantı noktaları ve elektrik üretme santrallerinde de kullanılmaktadır. Bu durum, türbin sayaçlar ile yüksek miktarlarda doğal gaz akışının okunabildiğini ortaya koymaktadır. İsmi, içerisinde bulunan ve dönme hareketi yapan türbinden gelmektedir. Eğimli pervanesi sayesinde içerisinden gaz geçişi olduğunda dönmeye başlar ve dönen her bir dişli sensörler tarafından okunmaktadır. Sahip olduğu bu elektronik sistem sayesinde pervanenin hızı ve tur sayısı okunarak akışı gerçekleşen doğal gazın miktarı hesaplanmaktadır. Körüklü doğal gaz sayaçlarına göre daha sık kalibrasyon ve bakım işlemi gerektirmektedir.

2.4. Kayıp ve hesaplanamayan doğal gaz

Kayıp veya hesaplanamayan doğal gaz, bir dağıtım sistemine gönderilen gaz ile tüketicilerin kullandıkları yani sayaçlarında ölçülen gaz miktarlarının toplamı

arasındaki fark olarak tanımlanmaktadır. Gaz dağıtım şirketleri, kendi müşterilerine doğal gazı iletmek için RMS-A istasyonlarını kullanmaktadırlar. Bu istasyonlar, öncelikle bölge şebekelerinde mevcut olan doğal gazı belirli bir basınç ile doldurmaktadır. Boru gazı olarak ifade edilen bu gaz, şebekeye bağlı olarak müşteriler tarafından kullanılarak fatura edilmektedir. RMS-A istasyonuna giren gaz ile son kullanıcıların tüketimleri arasında fark bulunabilmektedir. Kayıp ve hesap dışı doğal gaz olarak ifade edilen bu durumda, yukarıdaki tanımda da belirtildiği gibi müşterilerinin sayaçlarından ölçülen doğal gazın hacimsel olarak toplamı, şirketin sisteme gönderdiği toplam doğal gaz miktarından daha düşük çıkabilmektedir. Bu nedenle bir doğal gaz dağıtım şirketi, çoğu zaman öngörülenden ve talep edilenden daha fazla doğal gaz tedarik etmek durumunda kalmaktadır. Bu durum, doğal gaz dağıtım şirketlerini maddi kayıplara uğratmaktadır.

Bir dağıtım şirketinde ölçüm hataları, bakımları geciktirmesi ve borularda meydana gelen sızıntılardan dolayı kayıp doğal gaza sebep olabilmektedir. Bununla beraber doğal gaz sayaçlarının içinden geçen gaz miktarını ölçme prensibi gereği gazın sıcaklık ve basınç parametreleri etkilidir. Yaz aylarında ve kış aylarında hava sıcaklıklarının farklı olması doğal gazın sıcaklık ve basınç değerlerini değiştirmektedir. Bu durum, gazın debisini değiştirdiğinden ölçülen doğal gaz miktarı da gerçek tüketimden farklı olabilmektedir. Doğal gaz dağıtım şirketleri bu durumu engellemek adına faturalandırma aşamasında bir düzeltme katsayısı kullanarak tüketimi dengelemeye ve gerçek değerine yaklaştırmaya çalışmaktadırlar. Yine de son kullanıcıların yani tüketicilerin bulunduğu noktalarda doğal gazın sıcaklığını ve basıncını dengeleyici bir sistem bulunmadığından ölçüm hataları meydana gelmektedir ve kayıp doğal gazın önemli bir miktarını oluşturmaktadır [29].

Aşağıda, kayıp doğal gaz tüketimine sebep olan başlıca sebepler maddeler halinde verilmiştir [4].

- a. Sızıntılar: Doğal gaz iletim borularında zaman içerisinde meydana gelen deformasyondan veya çevresel faktörlerden kaynaklanmak üzere gaz sızıntısı meydana gelebilmektedir. Bu sızıntı tespit edildikten sonra onarım sağlanana kadarki kayıp doğal gazın miktarının tespiti oldukça zordur.
- b. Sıcaklık: Tüm doğal gaz sayaçları, bir önceki paragrafta da anlatıldığı üzere hacim ölçümü gerçekleştirmektedirler. Tüketilen doğal gaz ölçülürken bir bölgeden geçen doğal gaz hacmi esas alınmaktadır. Doğal gazın debisini

etkileyen en önemli faktörlerden birisi de sıcaklık ve basınç değeridir. Hava sıcaklığının farklı olduğu zaman dilimlerinde veya coğrafi bölgelerde doğal gazın debisi de değişmektedir. Gaz dağıtım firmaları farklı mevsimlerde hava sıcaklığını esas alarak faturalandırma aşamasında düzeltme faktörü parametresini belirleyebilmektedirler. Ancak aynı mevsim içerisinde bile tüketicilerin bulunduğu bölgelere göre doğal gazın sıcaklığı farklılık gösterebilmektedir. Bu durumda tüketilen doğal gaz hacmini hesaplamak için sabit bir sıcaklık değeri parametresi kullanmak kayıp doğal gaz tüketimine yol açmaktadır.

- c. Aktif olmayan kullanıcılar: Herhangi bir tüketici, mevcut konumundan bir başka yere taşınma gerçekleştirdiğinde veya aboneliğini sonlandırdığında doğal gaz akışı kesilmediği takdirde tüketilen doğal gaz, kayıp ve hesap dışı gaz kategorisinde yer almaktadır.
- d. Ölçüm hataları: Doğal gaz tüketimini otomatik okuma cihazları yerine bazı yazılımlar kullanarak iki ayda bir tahmin yöntemiyle tespit etmeye çalışılmaktadır. Bu şekilde yapılan tüketim tahminleri, gerçek tüketimi tam anlamıyla yansıtmayabilir. Benzer şekilde doğal gaz dağıtım şirketinin sayaç okumakla görevli olan saha çalışanı tarafından sayaçların yanlış okunması da kayıp doğal gaz tüketimine neden olmaktadır.
- e. Sayaçlar: Aktif olarak kullanımda olan bir doğal gaz sayacı, zaman içerisinde aynı ölçüm performansını sürdüremeyebilmektedir. Bu durumun başlıca sebepleri olarak sayaç içerisinde bulunan mekanik aksamaların eskimesi ve sayaçların kalibrasyon bakımı ihtiyacının doğması olarak gösterilebilir. Bu gibi durumlarda sayaçlar, çoğunlukla gerçek tüketimden daha az bir tüketim değeri hesaplayıp kayıp doğal gaz tüketimine sebep olmakla birlikte gerçek tüketimden daha yüksek bir tüketim de hesaplayabilmektedir. Fazla hesaplanması durumunda tüketiciler olduğundan daha yüksek fatura bedelleri ile karşılaşarak bu durumu fark etmektedirler. Böylece bağlı bulunduğu doğal gaz dağıtım şirketine inceleme talebinde bulunabilmektedirler. Ancak sayaç tarafından hesaplanan tüketimin, gerçek tüketim miktarının altında kaldığı durumun tespiti daha zor olmaktadır. Sayaç kaynaklı meydana gelen arızalar yalnızca sayaçlardan kaynaklanmayabilir. Tüketiciler tarafından sayaçlara yapılan kasıtlı fiziksel müdahaleler de sayaçlarda okuma hatasına neden olabilmektedir. Bu gibi durumlarda sayaçların mekanik parçalarına zarar

verilir ve sayaç okumasının yavaşlatılması amaçlanmaktadır. Bu durum doğal gaz hırsızlığı başlığı altında da incelenmektedir.

Doğal gaz dağıtım şirketleri bu durumların önüne geçebilmek adına belirli periyotlarda tüketicilerin sayaçlarını kontrol etmek üzere saha elemanı görevlendirmektedir. Ancak bu durum dağıtım şirketleri için hem maliyetli hem de kayıp doğal gaz tüketimlerinin tamamını tespit etme noktasında yetersiz kalmaktadır. Bununla beraber böylesi durumlara açık kapı bırakmamak adına belirli aralıklarla doğal gaz sayaçlarında kalibrasyon işlemleri yapılır veya gerek duyulursa sayaç sökülüp yeni sayaç takılmaktadır.

Sayaçlarda meydana gelen okuma hatalarının hepsi zamansal veya kullanıcıların kasıtlı fiziksel müdahalesinden dolayı oluşmamakta, üretimden kaynaklı sorunlardan da kaynaklanabilmektedir. Bu çalışmanın odaklandığı nokta, sayaçlardan kaynaklı kayıp doğal gaz tüketiminin tespiti ile kayıp doğal gaz tüketimi gerçekleşen abonelerin belirlenmesidir.

3. TÜKETİM TAHMİNİNDE KULLANILAN VERİLER VE YÖNTEMLER

Kayıp doğal gaz tüketimini tahmin edebilmek için abonelerin doğal gaz tüketim tahminlerini gerçekleştirmek ve bu tüketim serisi içerisinde anormal tüketimlerin belirlenmesi gerekmektedir. Bu bölümde abonelerin tüketim tahminlerini gerçekleştirmek için hazırlanan veri setleri ve kullanılan yöntemler verilecektir.

3.1. Veriseti

Doğal gazın tüketimini etkileyen birçok parametre bulunmaktadır. Bunlar; hava durumu, tüketici alışkanlıkları, tüketimi gerçekleştiren paydaşların sayısı, tüketicilerin ekonomik durumu ve ilgili okuma dönemi içerisinde abonelerin evde bulunma süreleri olarak gösterilebilir. Bu çalışmada doğal gaz tüketim verileri, abone ve sayaç işlemleri verileri, meteorolojik veriler ve takvimsel durumlar adında 4 farklı veri seti hazırlanmış ve birleştirme işlemine tabii tutularak üzerinde işlem yapılacak esas veri seti elde edilmiştir. Bu bölümde, hazırlanan veri setleri açıklanacaktır.

3.1.1. Doğal gaz tüketim verileri

Bu çalışmada kullanılan ve Sakarya ilinin aylık doğal gaz tüketim bilgilerini içeren veriler, Adapazarı Gaz Dağıtım A.Ş. (AGDAŞ)’den alınmıştır. Veriler, Ocak 2011 – Aralık 2021 tarihleri arasındaki 435.440 adet abonenin tüketimlerini içermektedir. Bu veriler Microsoft SQL Server (MSSQL) veri tabanına yüklenmiş ve veriler üzerinde gerçekleştirilecek olan öznitelik seçimi ve diğer veri setleriyle birleştirme işlemleri veri tabanı üzerinden gerçekleştirilmiştir. Doğal gaz tüketimini etkileyen önemli faktörlerden biri olan tüketilen gün sayısı bilgisinin hesaplanması faturada bulunan önceki ve şimdiki okuma tarihlerinin farkı alınarak hesaplanmaktadır. Bu bağlamda okuma tarihleri arasındaki gün farkının hesaplanması ve bu parametrenin esas veri setine eklenme işlemi veri tabanına yeni bir sütun eklenerek önceki okuma ve şimdiki okuma tarihlerinin farkı alınarak veri tabanı üzerinden gerçekleştirilmiştir. Bu aşamada önceki okuma tarihi bilgisi bulunmayan kayıtların gün farkı bilgisi hesaplanamayacağından, bu parametresinde veri eksikliği bulunan aboneler esas tabloya dahil edilmemiştir. Benzer şekilde ilgili abonenin tüketimini tahmin

edebilmek için eğitilecek modelde en az 12 aylık tüketim bilgisi eğitim için kullanılacağından, 12 aydan düşük tüketimi bulunan aboneler esas tabloya dahil edilmemiştir. Bu işlemler sonucunda diğer veri setleriyle birleştirme işlemine dahil edilen abone sayısı 307.570 olmuştur. Bu abonelerin toplam tüketim verileri ise 16.369.925 satırdır. Doğal gaz tüketim veri setinde bulunan bazı önemli parametreler aşağıda listelenmiştir:

- Abone numarası
- Şimdiki okuma tarihi
- Önceki okuma tarihi
- Kaç günlük okuma yapıldığı bilgisi (Gün farkı)
- İlk okuma indeksi
- Son okuma indeksi
- İndeks farkı
- Abone sayacının ID'si
- Tüketim miktarı (m3)

Bir binaya doğal gaz hattı döşenip gaz dağıtım firması tarafından sayaç takıldıktan sonra hanede oturan aboneler değiştikçe abone numarası da değişmektedir. Burada hanede oturan abonenin değişimi; evin satışı veya yeni bir kiracının gelmesiyle olabilmektedir. Yani doğal gaz aboneliğini aktif edip evde oturmaya başlayan her abonenin kendine ait eşsiz bir abone numarası bulunmaktadır. Benzer şekilde ilgili tüketici, aboneliğini herhangi bir sebepten ötürü sonlandırdığında ilgili abone numarası adına bir fatura oluşmamaktadır ve ilgili tüketicinin tüketim serisi son bulmaktadır. Bu çalışma kapsamında her bir aboneye özgü bir tüketim modeli oluşturulacağından belirtilen bu parametreler arasından abone numarası dikkate alınmıştır ve tüketim tahmininde kullanılacak olan nihai tabloya abone numaraları eklenmiştir. Tablo 3.1'de doğal gaz tüketim veri setinin örnek verileri verilmiştir.

Tablo 3.1. Doğal gaz tüketim veri seti örnek veriler.

Abone numarası*	Şimdiki okuma tarihi	Önceki okuma tarihi	Gün farkı	İlk okuma indeksi	Son okuma indeksi	İndeks farkı	Abone sayaç ID'si*	Tüketim miktarı
9	2012-04-20	2012-03-20	31	1247	1328	81	17	84
10	2012-04-20	2012-03-21	30	7	11	4	18	4
11	2012-04-20	2012-03-20	31	2100	2122	22	19	23
12	2012-04-20	2012-03-17	34	1740	1740	0	20	0
13	2012-04-20	2012-03-21	30	2880	2938	58	21	59
14	2012-04-20	2012-03-21	30	5733	5954	221	22	226
15	2012-04-20	2012-03-21	30	891	1014	123	24	126

*: Veri gizliliği kapsamında anonimleştirilen veriye sembolik ID atamaları yapılmıştır.

3.1.2. Abone ve sayaç işlemleri verileri

Abonelerin sayaç işlemleri verilerinin bulunduğu veri setidir. Her bir aboneye sayaç takma, sayaç sökme, sayaç kalibrasyon ve bu işlemlerin gerçekleştirildiği sayaç id'sinin bilgilerini içermektedir. Gaz dağıtım firmalarının çeşitli sebeplerden ötürü Bölüm 2'de anlatıldığı gibi sayaçlarda değişim veya kalibrasyon işlemlerinin kayıtları bu veri setinde tutulmaktadır. Bu durumda ise nihai tabloda abone ID'leri dikkate alınacağı için karışıklığa neden olmaması adına bu tabloya abone ID bilgisinin eklenme işlemi veri tabanı üzerinde gerçekleştirilmiştir. Bu işlem gerçekleştirilirken abone ve sayaç işlemleri verilerinin bulunduğu veri setindeki işlem tarihini belirten değişken, doğal gaz tüketim verilerinin içerisinde bulunan şimdiki okuma ve önceki okuma tarihleri arasında kalan tarih aralığında ise ve ilgili işlem satırındaki faturaya abone sayaç işlem verisi eklenmiştir. Abone ve sayaç işlemleri veri setinde bulunan parametreler aşağıda listelenmiştir:

- Abone numarası
- İşlem tarihi
- Abone sayacının ID'si
- İşlem tipi

Tablo 3.2'de abone ve sayaç işlemleri veri seti için örnek veriler verilmiştir.

Tablo 3.2. Abone ve sayaç işlemleri veri seti örnek veriler.

Abone numarası*	İşlem tarihi	Abone sayacının ID'si*	İşlem tipi
1	2014-04-16	7	T ^a
2	2014-04-16	8	S ^b
3	2014-04-16	9	K ^c

*: Veri gizliliği kapsamında sembolik ID atamaları yapılmıştır, a: sayaç takma, b: sayaç sökme, c: sayaç kalibrasyon

3.1.3. Meteorolojik veriler

Doğal gaz tüketimini doğrudan etkileyecek olan hava olayları ve meteorolojik verilerdir. Bu çalışmada kullanılmak üzere Sakarya ili için Ocak 2003 – Aralık 2021 aralığındaki tarihsel veri Meteoroloji Genel Müdürlüğü'nden temin edilmiştir. Diğer tablolar ile birleştirme işlemi gerçekleştirildiğinde kullanılabilecek doğal gaz tüketim

verileri Ocak 2011 tarihinden itibaren başladığı için meteoroloji verilerinin de Ocak 2011 – Aralık 2021 tarih aralığındaki günlük verileri kullanılmıştır. Kullanılan meteorolojik veriler aşağıda sırasıyla verilmiştir:

- Günlük güneşlenme süresi (saat)
- Günlük maksimum sıcaklık (°C)
- Günlük minimum sıcaklık (°C)
- Günlük ortalama sıcaklık (°C)
- Günlük maksimum nem (%)
- Günlük minimum nem (%)
- Günlük fırtına kaydı (ikili değişken)
- Günlük rüzgar hızı (m/sn)

Tablo 3.3'te meteoroloji veri setinin bir örneği verilmiştir.

Tablo 3.3. Meteoroloji veri seti örnek veriler.

Tarih	Guneslenme Suresi (Saat)	Maksimum Sicaklik (C)	Minimum Sicaklik (C)	Ortalama Sicaklik (C)	Maksimum Nem	Minimum Nem	Firtina Kaydi	Ruzgar Hizi (m/sn)
2013-02-03	3,9	25,1	14,1	18,5	51	26	0	12,3
2013-02-04	5,3	21,8	14,9	19,4	46	31	0	11,8
2013-02-05	1,5	14,9	8,6	10,5	98	45	0	6,7

3.1.4. Takvimsel olaylar veri seti

Doğal gaz tüketimini etkileyen önemli unsurlardan biri de ilgili tüketim dönemi aralığında abonelerin takvimsel olaylar ile ilişkileridir. Çalışmaya dahil edilmek üzere takvimsel olayları içeren bir veri seti oluşturulmuştur. Bu veri seti kullanıcı davranışlarını özelleştirmek ve modelin başarısını arttırmak amacıyla oluşturulmuştur. Ocak 2011 – Aralık 2021 tarihleri arasındaki hergün için tarih, Cumartesi, Pazar, Kurban Bayramı, Ramazan Bayramı, resmi tatil ve arefe günü bilgilerini içermektedir. Belirtilen bu isimlerde sütunlar oluşturulmuş ve ilgili satıra denk gelen tarihte belirtilen parametreler var ise 1, yok ise 0 değerleri atanmıştır. Kullanılan takvimsel olaylar parametreleri aşağıda sırasıyla verilmiştir:

- Cumartesi günü sayısı
- Pazar günü sayısı
- Kurban Bayramı günü sayısı

- Ramazan Bayramı günü sayısı
- Resmi tatil günü sayısı
- Arefe günü sayısı

Tablo 3.4’te takvimsel olaylar veri setinin örnek verileri verilmiştir.

Tablo 3.4. Takvimsel olaylar veri seti örnek veriler.

Tarih	Ctesi	Pazar	KurBay	RamBay	ResTatil	ArfGunu
2019-08-30	0	0	0	0	1	0
2019-08-31	1	0	0	0	0	0
2019-09-01	0	1	0	0	0	0
2019-09-02	0	0	0	0	0	0
2019-09-03	0	0	0	0	0	0

3.1.5. Veri setlerinin birleştirilmesi

Önceki bölümde verilen 4 veri seti birleştirilerek üzerinde işlem gerçekleştirilecek olan analitik baz veri seti elde edilmiştir. İlk olarak doğal gaz tüketim bilgilerinin bulunduğu tablodaki okuma tarihi, önceki okuma tarihi ve abone numarası bilgileri dikkate alınarak her bir okuma dönemi için, ilgili tüketim zaman diliminde abonelerin sayaçlarına herhangi bir işlem yapıp yapılmadığı bilgisi eklenmiştir. Bu bilginin eklenmesi için “SayacIslem” adında bir sütun oluşturulmuş olup, tüketim verilerinin her bir satırı için kontroller sağlanarak işlem varsa “1”, yoksa “0” olacak biçimde ikili değerler yazdırılmıştır.

Meteorolojik veriler, günlük hava olayları bilgilerini içermektedir. Güneşlenme süresi, maksimum sıcaklık, minimum sıcaklık, ortalama sıcaklık, maksimum nem, minimum nem ve rüzgar hızı parametreleri, doğal gaz tüketim veri setindeki her bir okuma dönemi zaman dilimi arasında kalan tarihlerin ortalamaları alınarak nihai veri setine dahil edilmiştir. Benzer işlem, ilgili tarih aralığında kaç tane fırtına kaydı girildiğini belirtmek için toplama işlemi yapılarak ilgili okuma dönemindeki toplam fırtına kaydı sayısı elde edilmiştir.

Takvimsel olaylar veri seti, yılın günlerindeki tüm tatil günlerini belirtmektedir. Bu veriler, günlük olayları içermektedir ve nihai tüketim veri setinde her bir doğal gaz tüketim periyodu aralığında bu tatil günlerinden kaç tane olduğunu tespit etmek

amacıyla kullanılmıştır. Takvimsel olaylar veri setindeki veriler ile doğal gaz tüketim veri seti verileri birleştirme işlemine tabi tutulurken, ilgili tarih aralığındaki takvimsel olaylar verileri toplanmaktadır. Takvimsel olaylar veri setinde bulunan değişkenler doğal gaz tüketim verileri ile birleştirilerek nihai veri setine eklenme işlemi gerçekleştirilmiştir.

Tüm veri setlerinin birbiri ile ilişkili alanlarının birleştirilme işlemi gerçekleştirilerek nihai veri seti elde edilmiştir. Elde edilen nihai veri seti 18 sütun, 17.162.598 satır veri içermektedir. Bu veri seti, içerisinde 435.440 adet doğal gaz abonesinin birbirinden farklı sayılarda fatura bilgilerini içermektedir. Veri setlerini birleştirme işleminde sonra elde edilen nihai veri setinin parametreleri Tablo 3.5'te verilmiştir.

Tablo 3.5. Nihai veri seti değişkenler.

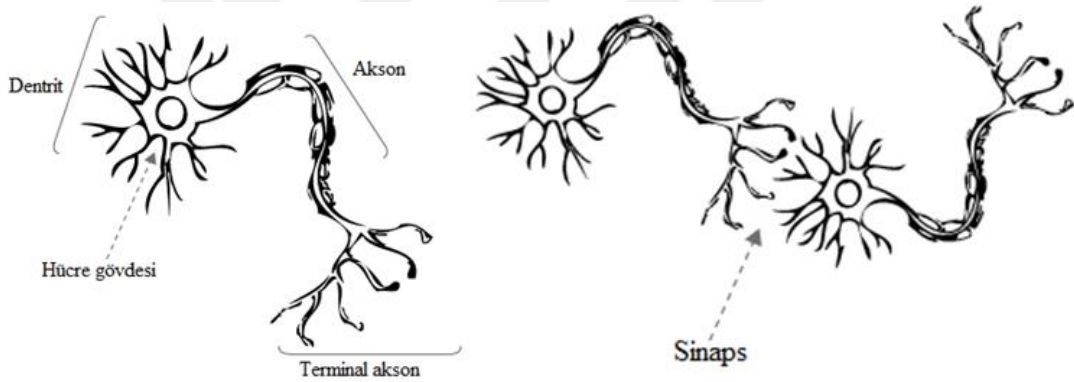
Değişken no	Açıklama
1	Abonenin benzersiz sözleşme numarası
2	Sayaca yapılan işlem sayısı
3	Faturanın kaç günlük tüketimi kapsadığı
4	Tüketim dönemi içerisindeki Cumartesi günü sayısı
5	Tüketim dönemi içerisindeki Pazar günü sayısı
6	Tüketim dönemi içerisindeki Kurban Bayramı sayısı
7	Tüketim dönemi içerisindeki Ramazan Bayramı sayısı
8	Tüketim dönemi içerisindeki Resmi Tatil günü sayısı
9	Tüketim dönemi içerisindeki Arefe günü sayısı
10	Günlerin ortalama güneşlenme süreleri ortalaması
11	Günlerin maksimum sıcaklık ortalamaları
12	Günlerin minimum sıcaklık ortalamaları
13	Günlerin ortalama sıcaklıklarının ortalamaları
14	Günlerin maksimum nem oranı ortalaması
15	Günlerin minimum nem oranı ortalaması
16	Tüketim aralığındaki fırtına kaydı olan gün sayısı
17	Günlerin rüzgâr hızı ortalamaları
18	Tüketim miktarı (m ³)

3.2. Kullanılan Yöntemler

Bu çalışmada, abonelerin doğal gaz tüketim tahminlerini gerçekleştirmek için yapay zeka yöntemleri kullanılmıştır. Tarihteki ilk bilgisayarlar, hesaplamalar yapabilen ve karmaşık matematiksel problemlere çözümler bulabilen makineler olarak kullanılmıştır. 1980'li yıllardan itibaren çıkarımlar yapma ve karar verme işlemlerinde kullanılmaya başlanmıştır. 2000'li yıllara gelindiğinde, donanım teknolojisindeki gelişmeler sayesinde bilgisayarların hesaplama yetenekleri artmıştır. Bu gelişme, öğrenilebilir makineler teorisi üzerindeki çalışmaları hızlandırmıştır.

Makinelere düşünme yeteneğinin kazandırılması konusu, ilk olarak 1950 yılında A. Turing tarafından yayımlanan bir makalede tartışılmıştır [30].

Yapay zekâ, insanlara özgü olan görme, duyma, karar verme, öğrenebilme ve yeni durumlara uyum sağlayıp kendini güncelleyebilme gibi özelliklerin bilgisayarlar tarafından yaptırılabilme çalışmalarının bütünüdür. Bilgisayarlara düşebilme yetisinin kazandırılması konusundaki çalışmaların çoğunda, insanların düşünebilme yeteneklerinden ve doğadaki canlıların davranışlarından esinlenilmiştir. Bu nedenle, yapay zekâ kavramını tanımlayabilmek için biyolojik zekayı iyi anlamak gerekmektedir. İnsanların belirli bir amaç doğrultusunda hareket etmesi, bulunduğu ortamlara uyum sağlayabilmesi ve düşünebilip mantıklı kararlar verebilme yetisi biyolojik zekâ olarak ifade edilebilir. İnsanlar, biyolojik sinir hücrelerine sahiptir. İnsan zekâsı, beyinde mevcut olan çok sayıdaki birbirine bağlı sinir hücreleri ile ilgilidir ve bu bağlantıların kuvveti zekaya etki eden bir durumdur. Şekil 3.1’de bir biyolojik sinir hücresinin sembolik gösterimi verilmiştir [31].



Şekil 3.1. Biyolojik sinir hücresinin sembolik gösterimi.

İnsan beyinde bilginin iletimi elektriksel sinyaller ile gerçekleşmektedir. Bir sinir hücresine gelen uyartılar, dentritler tarafından alınmaktadır. Dentritler tarafından alınan elektriksel sinyal, hücre gövdesine iletilir ve burada bir çıkış sinyali üretilip üretilmeyeceği eşik değeri ile karşılaştırılır. Bu eşik değeri, insan vücuduna dışarıdan fiziksel bir müdahale sonrasında verilecek tepkinin büyüklüğü olarak düşünülebilir. Eğer bir çıkış sinyali üretilecekse bu bilgi aksonlardan geçerek terminal aksona, oradan da bir diğer sinir hücresine aktarılır [32]. Bilgilerin bir sinir hücresinden diğerine aktarımın sağlandığı bağlantı noktalarına sinaps adı verilmektedir ve insan beyinde bilgilerin saklandığı bölgenin sinapslar olduğu düşünülmektedir. Yani

sinaps bağlantı noktaları ne kadar yüksek ise beynin hafıza kapasitesi o kadar yüksektir [37–39].

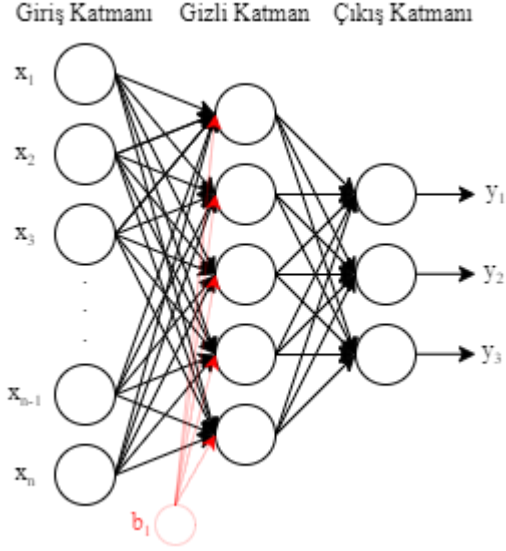
3.2.1. Yapay sinir ağları

Yapay sinir ağları (YSA), insan beyninde mevcut olan biyolojik sinir hücrelerinden ve çalışma prensibinden esinlenerek yapay olarak benzetimi sonucu ortaya çıkmıştır. YSA temel olarak veri girdilerinin yapıldığı giriş katmanı, gizli katman ve çıkış katmanından oluşmaktadır. Giriş katmanı, verilerin ağı verildiği katmandır. Veriler bu katmanda doğrudan gizli katmana aktarılır. Gizli katman, kendisine gelen verileri ağı ağırlık değerleri ile çarpılarak ve ilgili bias değerleri ile toplayarak nöronun çıkışına iletir. Veri, gizli katman nöronundan çıkarken kullanılan transfer fonksiyonuna parametre olarak girer ve buradan elde edilen sonuç çıkış katmanına aktarılır. Çıkış katmanı, işlem sonucunun çıktı olarak elde edildiği katmandır [36].

YSA'daki ağırlık değerleri, biyolojik sinir hücresindeki bilgilerin saklandığı bölgeyi temsil etmektedir. Bu nedenle YSA'da öğrenme, ağırlıklar sayesinde gerçekleşmektedir. YSA, değişen durumlar karşısında ağırlık değerlerini günceller ve yapılan bu işlem YSA'nın biyolojik sinir hücresi gibi kendisini yeni durumlara adapte etmesi olarak gösterilmektedir. YSA için değişen durumlar, farklı veri setleri olarak düşünülebilir.

Bir YSA modelinin performansını etkileyen en önemli unsurlar gizli katmandaki nöron sayıları ve transfer fonksiyonunun seçimidir [37]. Transfer fonksiyonu, ilgili nöronu aktif kılmak veya kılmamak adına belirli bir aralıkta çıktılar üretmektedir ve aktivasyon fonksiyonu olarak da anılmaktadır. Bu kontrol yapısı, biyolojik sinir hücresindeki elektriksel sinyallerin eşik değerine ulaşp ulaşmadığını sınamaya benzetilmektedir.

Şekil 3.2'de örnek bir yapay sinir ağının mimarisi verilmiştir [38].



Şekil 3.2. Örnek bir yapay sinir ağı mimarisi.

Görseldeki her bir yuvarlak sembol, bir sinir hücresini (nöron) temsil etmektedir. Giriş katmanındaki nöron sayısı, ağı girdi olarak verilecek parametre sayısına bağlı olarak değişmektedir. Her bir nöron arasındaki bağlantıları gösteren çizgilerde ise ağırlık değerleri bulunmaktadır.

Bu görselde 5 nöron içeren 1 adet gizli katman kullanılmıştır. Gizli katman sayısı ve sahip olduğu nöron sayıları problemin türüne göre ve ağı performansı test edilerek belirlenebilir. Kırmızı renkteki çizgiler ise gizli katman nöronlarına gelen bias (eşik değeri) değerleridir. Bu değer, nöronlara gelen değerleri düzenlemek için kullanılmaktadır. Çıkış katmanı, verilen girişlere karşılık ağıdan elde edilen çıktıların elde edildiği katmandır. Bu mimaride, 3 adet çıktı üreten yapay sinir ağı verilmiştir ve çıktı değerleri y_1, y_2 ve y_3 'tür.

YSA'da öğrenme, sinir ağları tarafından çözülmesi istenen problemin optimum çıktılarını üretecek biçimde ağırlıkların güncellenme işlemidir. Nöronlar arasında bulunan ağırlıklar, elde edilen hata değerine göre adım adım güncellenmektedir. Hata değeri ise ağıdan elde edilen çıktı ile üretilmesi beklenen çıktı arasındaki fark olarak tanımlanmaktadır.

3.2.1.1. Eğitici öğrenme

Çözülmesi istenen problem ile alakalı örneklerin girdileri ve çıktıları ağı verilmektedir. Bu modelden beklenen kendisine verilen girdilerden öğrenmesi ve çıktıları bilinmeyen girdiler için haritalama -eşleme- yapmasıdır [39]. Eğitim işlemi

gerçekleşirken veri örnekleri arasında işlemin istenen ve beklenen çıktısını gösteren bir etiketli veri bulunmaktadır. Şayet gerçekleştirilmesi istenen işlem bir sınıflandırma problemi ise, etiketli veriler arasında ilgili veri satırının hangi gruba dahil olduğunu belirten bir veriler bulunmaktadır. Sınıflandırma yerine bir regresyon işlemi ise, sonucu belirtlen bir sayısal büyüklük verilmektedir. Bu sayısal büyüklükler tüketim, fiyat, sıcaklık, maliyet gibi değerleri ifade edebilir [40].

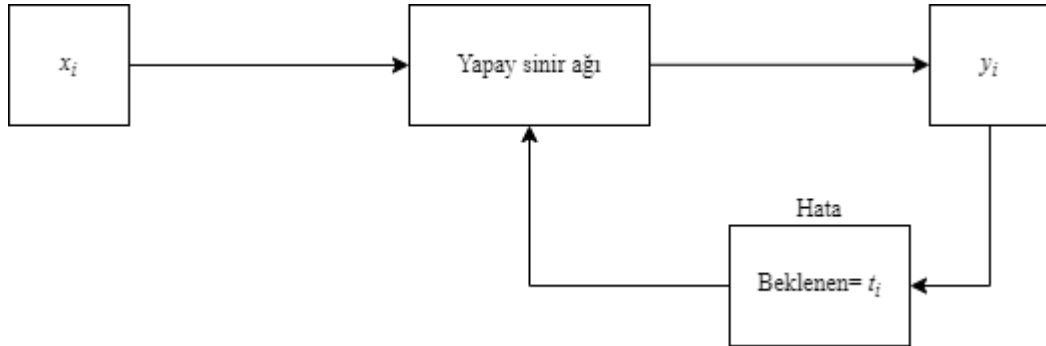
$x_i \in X$ giriş verilerinin kümesi, $y_i \in Y$ çıktı verilerinin kümesi ve $t_i \in T$ beklenen çıktı değerlerinin kümesi olduğunu kabul edelim. Burada her bir x_i girişlerine karşılık bir y_i çıktısı üretilmektedir. Daha sonra üretilen y_i çıktıları, beklenen t_i çıktıları ile karşılaştırılarak modelin doğruluğu gözlemlenebilmektedir.

Bu modelde tüm (x_i, y_i) giriş çıkış kümesi için çıktıların vektörü Denklem 3.1’de ki gibi ifade edilmektedir [41].

$$y = (y_i)_{i \in N}^T \quad (3.1)$$

Burada N , veri setindeki eleman sayısı; i , verinin indisi; y_i , i indisine sahip y çıktısı, T , çıktı dizisinin transpozu ve y ise çıktıların vektörünü ifade etmektedir.

Şekil 3.3’te eğitici öğrenmenin şeması verilmiştir [42].



Şekil 3.3. Eğitici öğrenme mimarisi.

Her bir iterasyon için çıktının elde edilmesinden sonra hata değeri Denklem 3.2’de verilen ifade ile hesaplanmaktadır.

$$\epsilon_i = y_i - t_i \quad (3.2)$$

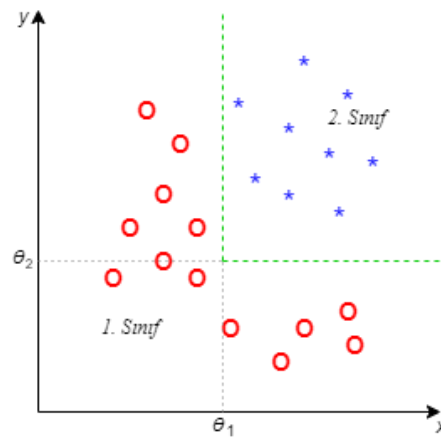
Burada ϵ hatayı, t beklenen değeri ve y gerçekleşen çıktı değerini ifade etmektedir. Geri yayılım algoritması ve türev tabanlı algoritmalar bu gruba ait birer öğrenme

algoritmalarıdır. Bu nedenle hata hesaplama işleminden sonra geriye yayılım yapılır ve hesaplanan bu hata üzerinden ağırlıklar güncelleştirilir. Hata hesaplama işleminde farklı yöntemler kullanılabilir. Örnek olarak geri yayılım algoritmasında gerçekleşen ve beklenen çıktılarının farkının karesi alınır. Elde edilen sonucun yarısı alınır ve bu işlem tüm çıkışlar için hesaplanarak tüm değerler toplanır. Böylece toplam hata değeri elde edilmektedir. Daha sonra bu hata değerine göre YSA'nın ağırlık değerleri güncellenmektedir. Bu konuya geri yayılım algoritması ve yapay sinir hücresi başlığı altında daha detaylı değinilecektir. Bu yöntem, literatürde mevcut olan bazı çalışmalarda danışmanlı öğrenme olarak da geçebilmektedir. Bunun sebebi, üretilmesi beklenen sonucun kendisinin de eğitim aşamasında ağırlık verilmesidir.

Eğitici öğrenme modeli, sınıflandırma ve regresyon olarak iki grup altında incelenebilir [43]. Bu tez çalışması kapsamında eğitici öğrenme modellerinden regresyon işlemi gerçekleştirilmiştir.

Sınıflandırma, bir veri kümesi içerisinde özniteliklere bağlı olarak ilgili veri satırının hangi sınıfa ait olduğunu belirleme işlemidir. Bir sınıflandırma probleminde etiketli veriler, sınıflandırma yapılacak kategorileri belirtmektedir. Bu kategori sayısı 2 ve üzeri olmalıdır. Sınıflandırma işlemi karar ağacı benzeri bir hiyerarşik ağaç yapısı düzeninde de olabilir [44]. Etiketli veriler, regresyon yapısında olduğu gibi birbirinden farklı sayısal değerler değildir. Sayısal değerler kullanılacaksa genel olarak 1 ve 0 şeklinde mantıksal ifadeler içermektedir.

Şekil 3.4'te bir sınıflandırma problemi için grafik verilmiştir. Grafikte sınıflandırma işleminin mantığı Denklem 3.3'te verilmiştir [45].



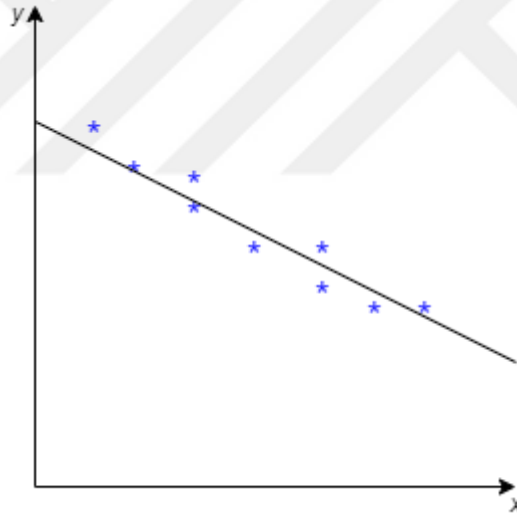
Şekil 3.4. Sınıflandırma işlemi.

$$\text{Eğer } x_i > \theta_1 \text{ Ve } y_i > \theta_1 \text{ İse Sınıf} = 2. \text{ Sınıf, Değilse Sınıf} = 1. \text{ Sınıf} \quad (3.3)$$

Burada x_i ve y_i değerleri, x ve y parametrelerinin birer örnekleimidir. θ_1 değeri ise sınıflandırma işlemi için eşik değerlerini ifade etmektedir.

Regresyon, birden fazla parametrenin tek bir değişkeni etkilediği durumlarda bağımlı değişkenin değerini tahmin etme işlemidir. Tahmin etme işlemi sonucunda tek bir sayısal değer elde edilmektedir [40]. Bu nedenle sınıflandırmadan farklı olarak yapılan tahmin işleminin başarısı, belirlenen kategorinin doğruluğuna veya yanlışlığına göre değil tahmin edilen ve gerçekleşmesi beklenen değerler arasındaki farkın hesaplanma işlemi ile gerçekleştirilmektedir.

Şekil 3.5'te tek girdisi olan bir regresyon işleminin örnek grafiği verilmiştir. Denklem 3.4'te ise Şekil 3.5'te mevcut olan bağımlı y değişkenine, bağımsız x ve w parametrelerinin etkisinin hesabı verilmiştir [45].



Şekil 3.5. Regresyon işlemi.

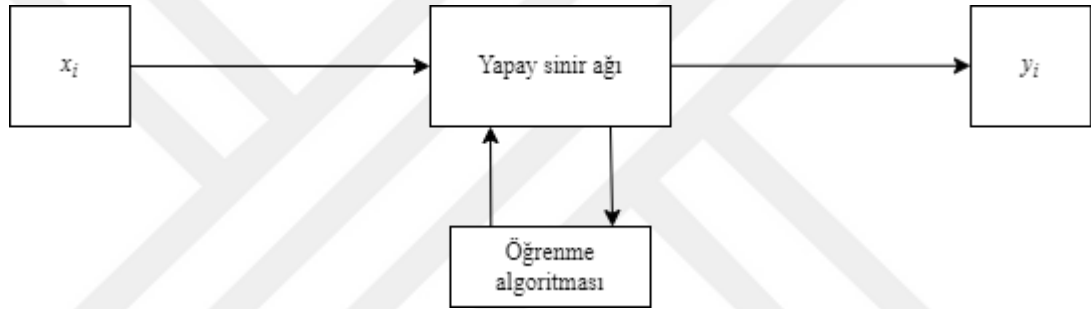
$$y_i = w_1 \cdot x_i + w_0 \quad (3.4)$$

Burada x_i değeri X girdiler kümesinin bir elemanı, w_i değişkenleri y çıktısına etki eden parametreler ve y_i ise çıktı değerleridir. w_i parametreleri YSA'daki ağırlık değerlerine benzetilebilir. Örnekte verilen regresyon eğrisi doğrusaldır. Verilerin türüne ve çözülmesi gereken problemin tipine göre regresyon eğrisinin derecesi ona göre seçilmelidir.

3.2.1.2. Eđiticisiz ğrenme

Veri seti ierisinde her zaman etiketli veri bulunmayabilmektedir. Bu durumda eđiticisiz ğrenme modelleri kullanılmaktadır ve herhangi bir ğretici bulunmadığı iin bu model denetimsiz ğrenme olarak da tanımlanmaktadır [46]. Herhangi bir ğrenme yntemine ihtiya olmayan bu yntemde ađ, kendisine girdi olarak verilen veri rneklemlerini alır ve verilerin dađılımına gre graplama iřlemi gerekleřtirir. Bu modelde ađa, sadece girdi parametreleri verilmektedir ve ilgili ađ ğrenme iřlemini kendisi gerekleřtirmektedir. Model parametrelerinin gncellenme iřlemi ise eđiticili modelde olduđu gibi girdi ve beklenen deđerler arasındaki farka (hataya) gre deđil, giriř verileri arasındaki istatistiksel dađılıma gre gerekleřtirilmektedir.

řekil 3.6’da eđiticisiz ğrenmenin mimarisi verilmiřtir [47].

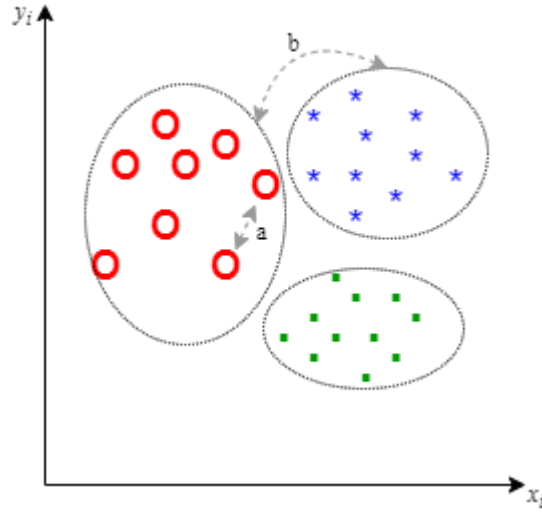


řekil 3.6. Eđiticisiz ğrenme mimarisi.

Burada her bir x_i girdisi iin y_i ıktıları belirlenen bir ğrenme algoritması yardımıyla hesaplanmaktadır.

En yaygın kullanılan eđiticisiz ğrenme yntemlerinden biri kmeleme yaklařımıdır. Kmeleme, veri dađılımı arasında birbirine benzeyen veri gruplarının kmelenmesi ve aynı gruba dahil edilmesi iřlemidir. Sınıflandırma iřleminde farkı herhangi bir hedef sınıf deđer olmamasıdır. Aynı kme ierisinde mevcut olan veri noktalarının birbirleri arasındaki benzerliğinin en yksek seviyede olması beklenmektedir. Kmeleme iřleminde, verilerin dađılımı gzetilerek bir kme sayısı belirlenir ve veriler belirlenen kme sayısınca kmelere dahil edilmektedir.

řekil 3.7’de kmeleme yaklařımının grseli verilmiřtir [48].



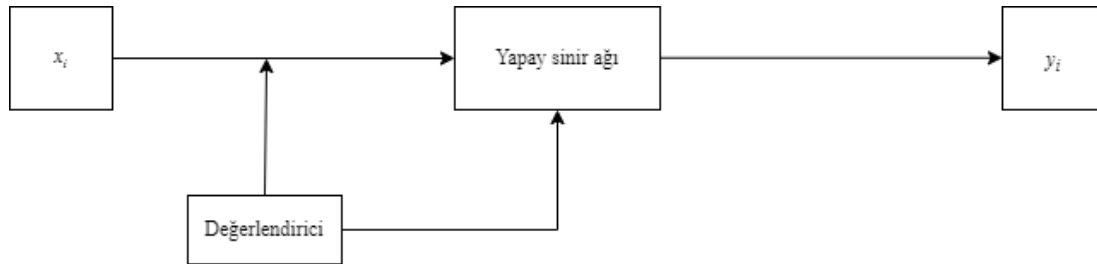
Şekil 3.7. Kümeleme yaklaşımı görseli.

Burada x_i ve y_i giriş parametlerine karşılık veri noktalarının dağılımına göre küme sayısı 3 olacak şekilde kümeleme işlemi gösterilmektedir. Kümeler arasındaki benzerliği b uzaklığı, küme elemanları arasındaki benzerliği a uzaklığı temsil etmektedir. Bu yaklaşımda gözetilen durum, veri noktaları arasındaki uzaklıktır.

3.2.1.3. Takviyeli öğrenme

Bu yöntemde üretilen sonucun doğruluğu veya yanlışlığı ağı takviye sinyali olarak geri bildirim şeklinde gönderilmektedir. Eğitici öğrenme yöntemine göre farkı, bu modellerde ağ kendi ürettiği sonucu değerlendirmemektedir. Yalnızca, verilen girdiler sonucunda elde edilen çıktıların beklenen çıktı değerlerine ne kadar yaklaştığı bilgisi üzerinde yoğunlaşmaktadır. Model, öğrenme işlemini kendisine gelen bu takviye sinyallerini dikkate alarak gerçekleştirmektedir.

Şekil 3.8’de takviyeli öğrenme modelinin mimarisi verilmiştir [49].



Şekil 3.8. Takviyeli öğrenme mimarisi.

Burada her bir x_i girdisi için elde edilen y_i çıktılarının beklenen t_i çıktılarına ne kadar yaklaştığı bilgisi değerlendirici tarafından tespit edilir ve ağa takviye sinyali gönderir. Bu nedenle ilgili model, ödül sistemi ile çalışan ağ olarak da belirtilmektedir. Ağ, değerlendirici tarafından kendisine gönderilen ödülleri en yüksek seviyeye çıkarmaya çalışmaktadır ve öğrenme bu şekilde gerçekleşmektedir. Bu özelliğinden dolayı ilgili model, literatürde pekiştirmeli öğrenme olarak da adlandırılmaktadır.

3.2.1.4. İleri beslemeli hesaplama

İleri beslemeli sinir ağları, giriş katmanından başlayarak çıkış katmanına doğru tek yönlü adım adım hesaplanarak sürdürülen işlemler yöntemidir. Bu YSA modelinde eğiticili öğrenme yöntemi kullanılmaktadır. Şekil 3.2’de verilen YSA mimarisi bir ileri beslemeli hesaplama örneğidir. İlgili YSA mimarisi dikkate alındığında, giriş katmanında bulunan girdi değerleri giriş katmanından ilk gizli katmana geldiğinde ilgili nörona olan bağlantısının sahip olduğu ağırlık değerleri ile çarpılır. Elde edilen çarpım değerleri bias değerleri ile toplanarak ilgili nöronun NET değeri elde edilmiş olur. YSA’da mevcut olan her bir nöron kendi içerisinde iki kısımdan oluşmaktadır. Net kısmı; x_i girdisi, w_i ağırlık değeri ve b_i bias değerleri olmak üzere Denklem 3.5’te verilmiştir.

$$NET_i = x_i \cdot w_i + b_i \quad (3.5)$$

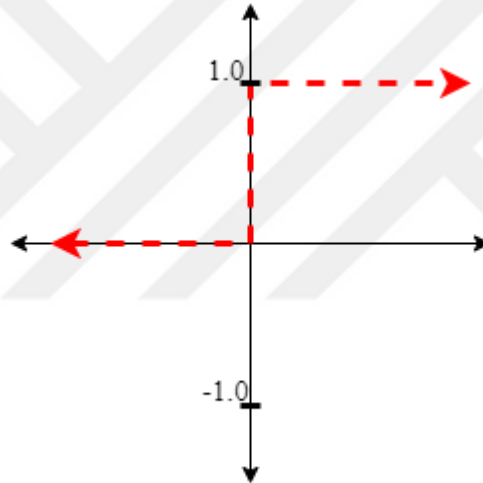
Şeklinde hesaplanan i . nöron için NET_i çıktı değeridir. Hesaplanan bu değer ilgili nöronun gerçek çıktısının hesaplanması için aktivasyon (transfer) fonksiyonuna parametre olarak gönderilmektedir. NET_i çıktısının transfer fonksiyonuna parametre olarak gönderildiği ve nöron çıktısının elde edildiği kısım ise Out olarak adlandırılmaktadır.

Transfer fonksiyonları, bir ağın ürettiği NET_i çıktısını beklenen çıktılar ile haritalama işlemini gerçekleştiren eşitliklerdir. Bir YSA modelinden, girdi çıktı setleri arasında doğru ve genelleyici bir yapı oluşturması beklenildiğinden, bu işlemi gerçekleştirmesi için transfer fonksiyonları kullanılmalıdır. Transfer fonksiyonları kullanılmadığı durumda ağın ürettiği çıktılar, lineer regresyon gibi doğrusal bir fonksiyon olacaktır. Bu durum ise ağın öğrenme becerisini oldukça düşürmektedir. Bir ağın, öğrenmesinin yüksek ve genelleyici çıktılar üretmesi için yüksek dereceden

fonksiyon gibi de davranması gerekmektedir. Ağa bu özelliği ise transfer fonksiyonları kazandırmaktadır. Kullanılacak transfer fonksiyonun türü, çözülmesi istenen problemin türüne ve verilerin durumuna göre değişkenlik göstermektedir. Bu nedenle doğru transfer fonksiyonunun seçimi, ağın performansını arttırmaktadır.

Basamak fonksiyonu, bir ağıın ürettiği NET_i çıktısını, bir eşik değeri ile karşılaştırmaktadır. Bu eşik değerinin altında veya üzerinde olma durumuna göre 0 veya 1 değeri sonucuna iletmektedir. Üretilen çıktı doğrusal olmadığından dolayı sınıflandırma problemlerinde kullanılması uygun görülmektedir. Denklem 3.6'da basamak fonksiyonunun eşitliği ve Şekil 3.9'da bu fonksiyonun grafiği verilmiştir.

$$OUT_i = \begin{cases} 0, & NET_i < 0 \\ 1, & NET_i \geq 0 \end{cases} \quad (3.6)$$

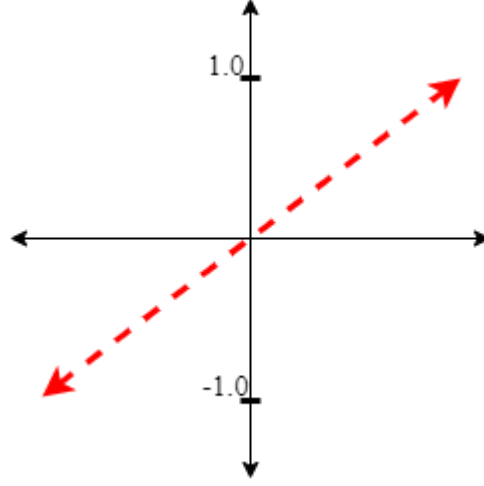


Şekil 3.9. Basamak fonksiyonu grafiği.

Doğrusal fonksiyon, ilgili nörondan çıkan NET_i değerinin aynısını OUT_i değeri olarak çıktı üretmektedir. Basamak fonksiyonundan farklı olarak, fonksiyon çıktısının grafiği doğrusaldır. Sınıflandırma işlemleri yerine çıkış değerinin tek bir sayısal değer olduğu bilinen problemlerin çözümlerinde kullanılmaktadır. Denklem 3.7'de doğrusal fonksiyonun eşitliği ve Şekil 3.10'da bu fonksiyonun grafiği verilmiştir.

$$OUT_i = f(NET_i) \quad (3.7)$$

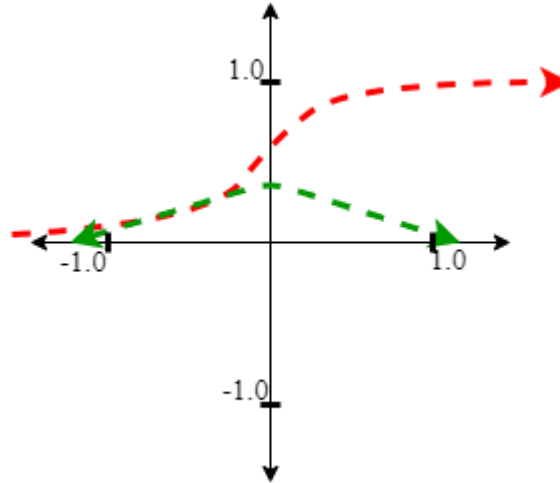
Bu nöronda $f(NET_i)$ değeri, giriş değerinin kendisine eşittir.



Şekil 3.10. Doğrusal fonksiyon grafiği.

Log-Sigmoid fonksiyonu, en yaygın kullanılan transfer fonksiyonları arasındadır. Bunun nedeni, çıktısının doğrusal olmayışı ve yalnızca ikili değerler almak yerine polinomsal olarak (0,1) aralığında değerler üretmektedir. Bu durum ise ağı, daha genelleyici çıktılar üretmesini sağlamaktadır. Denklem 3.8’de bu fonksiyonun eşitliği, Şekil 3.11’de fonksiyonun ve türevinin grafiği verilmiştir [50].

$$OUT_i = \frac{1}{1 + e^{-NET_i}} \quad (3.8)$$

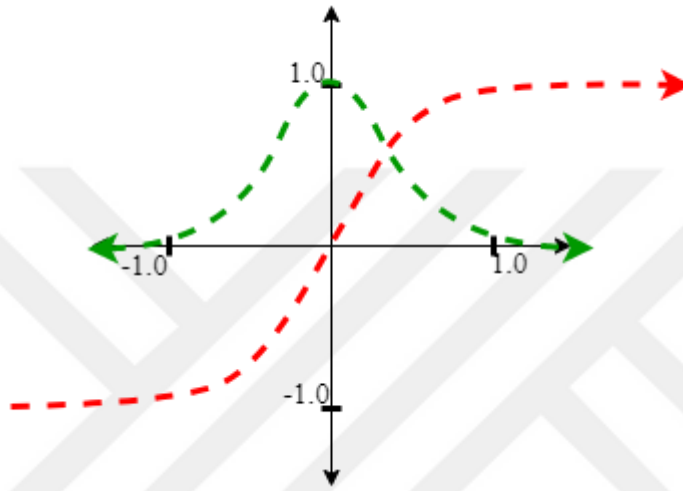


Şekil 3.11. Log-sigmoid fonksiyon grafiği.

Bu fonksiyonda öğrenmenin en hızlı olduğu bölge girdiler ekseninde (-1,1) aralığında gerçekleşmektedir.

Tan-Sigmoid fonksiyon, yapısal olarak log-sigmoid fonksiyonuna benzemektedir. Farklı olarak çıktı değerini (-1,1) aralığında üretmektedir. Diğer bir avantajı ise log-sigmoid fonksiyona göre türevi daha dik olduğundan daha fazla değerler alabilmektedir. Denklem 3.9'da bu transfer fonksiyonunun matematiksel gösterimi ve Şekil 3.12'de fonksiyonun ve türevinin grafiği verilmiştir [50].

$$OUT_i = \frac{2}{1 + e^{-2(NEt_i)}} \quad (3.9)$$



Şekil 3.12. Tan-sigmoid fonksiyon grafiği.

Transfer fonksiyonunun Sigmoid fonksiyonu olduğu varsayılırsa *Out* değerinin hesabı Denklem 3.10'da ki gibi hesaplanmaktadır.

$$OUT_i = 1 / (1 + e^{-NEt_i}) \quad (3.10)$$

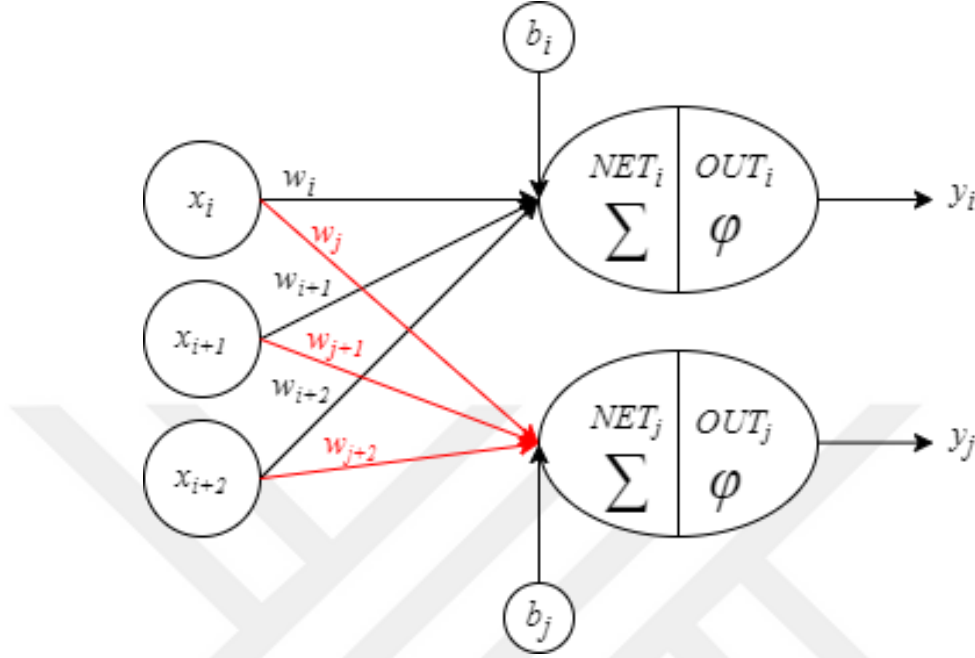
Bu denklemde OUT_i ; *i*. nöronun çıktı değerini ifade etmektedir. Bu değer bir sonraki katmana iletilir ve tüm işlemler *n* adet nöron için gerçekleştirilir. Bir sonraki katmanda bulunan NEt_{i+1} nöronuna bir önceki katmandan 2 adet nöron bağlı olduğu düşünülürse hesaplama işlemi Denklem 3.11'de ki gibi gerçekleştirilmektedir.

$$NEt_{i+1} = w_{i+1} \cdot OUT_i + w_{j+1} \cdot OUT_j + b_{i+1} \quad (3.11)$$

Benzer şekilde *i* + 1'nci katmanda bulunan NEt_{i+1} nöronunun çıktısı Sigmoid fonksiyonu kullanıldığı kabul edilirse Denklem 3.12'de ki gibi hesaplanmaktadır.

$$OUT_{i+1} = 1/1 + e^{-NET_{i+1}} \quad (3.12)$$

Şekil 3.13’de bir yapay sinir hücresinin *Net* ve *Out* kısımlarına odaklanılmıştır.



Şekil 3.13. Yapay sinir hücresi.

Burada x_i , x_{i+1} ve x_{i+2} girdiler olmak üzere giriş katmanında 3 nöron bulunmaktadır. Bu mimari dikkate alınarak y_i çıktısını üretecek olan nöronun NET_i kısmının hesaplanması genelleştirilmiş biçimde Denklem 3.13’de verilmiştir.

$$NET_i = \sum_{i=1}^n x_i \cdot w_i + b \quad (3.13)$$

Hesaplanan NET_i değeri, OUT_i değerinin hesaplanması için transfer fonksiyonuna parametre olarak verilir ve böylece nöron çıktısı elde edilmiş olmaktadır. Bu işlem Denklem 3.14’de verilmiştir.

$$OUT_i = \phi(NET_i) \quad (3.14)$$

Bu denklemde ϕ ifadesi transfer fonksiyonunu temsil etmektedir ve OUT_i değeri y_i değerine eşittir. Aynı işlemler y_j çıktısını üretecek olan nöron için de gerçekleştirilerek NET_j ve OUT_j değerleri elde edilebilmektedir.

3.2.1.5. Geri yayılım algoritması

Bir YSA modelinde ileri doğru hesaplama işlemi gerçekleştirilince Denklem 3.10'da belirtilen formül ile bir çıktı değeri elde edilmektedir. Tahmin edilen OUT_i çıktısı ve beklenen çıktı olan t_i değerleri arasında bir fark meydana gelmektedir. Bu fark hata değeri olarak adlandırılmaktadır ve bu hatanın hesaplanması Denklem 3.15'te verilmiştir.

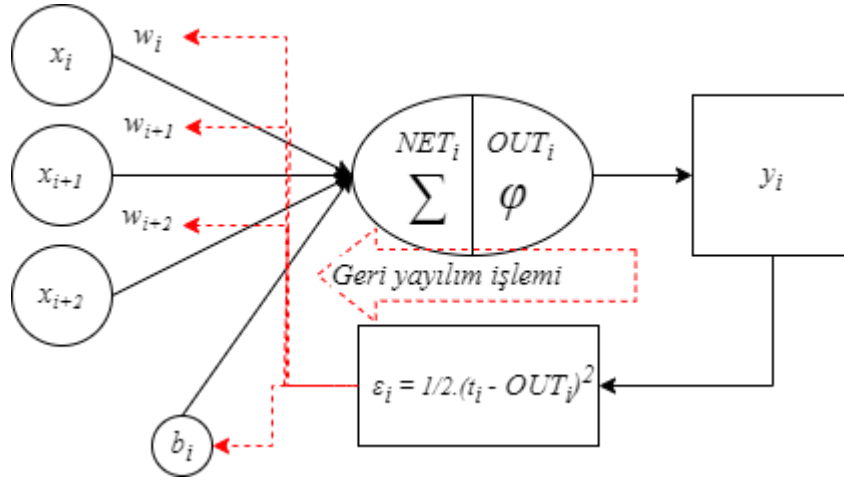
$$\varepsilon_i = \frac{1}{2}(t_i - OUT_i)^2 \quad (3.15)$$

Burada i . nöron için; t_i beklenen çıktı değerini, OUT_i tahmin edilen çıktı değerini ve ε_i ise hata değerini ifade etmektedir. Şekil 3.13'te verilen mimaride, çıktı nöron sayısı 2 olduğundan gerçekleşen tahmin değeri ile hedeflenen tahmin değeri arasındaki farkın karesi ikiye bölünmektedir. Benzer şekilde çıktı nöron sayısı n adet olan bir YSA mimarisi için ileri beslemeli hesaplama sonucunda elde edilecek toplam hata ε_{Toplam} olarak ifade edilebilir ve genelleştirilmiş değeri Denklem 3.16'da ki gibi hesaplanmaktadır.

$$\varepsilon_{Toplam} = \sum_i^n \frac{1}{n}(t_i - OUT_i)^2 \quad (3.16)$$

İleri hesaplama işlemi tamamlandıktan sonra elde edilen bu hata değeri, geriye doğru yayılmaktadır. Bu şekilde YSA'nın ağırlık değerleri güncellenir ve elde edilen hata değeri bu ağırlık değerleri üzerine geriye doğru dağıtılmaktadır. Bu nedenle, bu yöntemde geriye yayılım adı verilmektedir. Bu şekilde her bir iterasyonda ağırlık ve bias değerleri güncelleştirilir ve bir sonraki iterasyonda hata değerinin düşmesi amaçlanmaktadır. Bu algoritma, ileri beslemeli ağlar için en sık kullanılan algoritmalarından birisidir. Dikkat edilmesi gereken noktalardan birisi, türev tabanlı bir algoritma olduğu için kullanılan transfer fonksiyonun da türevi alınabilir bir fonksiyon seçilmiş olması gerekmektedir.

Şekil 3.14'te geri yayılım algoritmasının mimarisi verilmiştir [51].

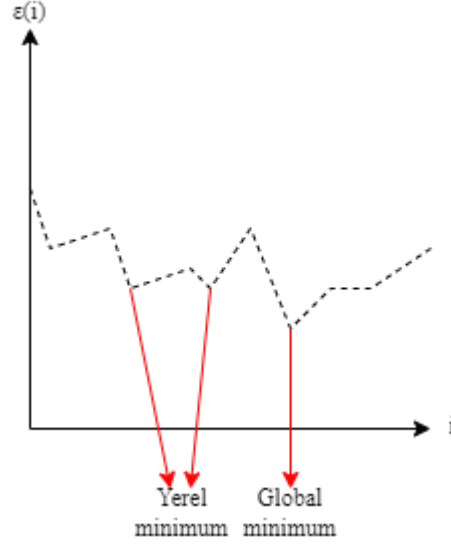


Şekil 3.14. Geri yayılım algoritması mimarisi.

Bu mimaride, 3 giriş ve 1 çıkış katmanına sahip olan tek katmanlı bir YSA verilmiştir. Geri yayılım algoritması ile elde edilen hatanın ağırlıklar üzerine dağıtılması işlemi, öğrenmenin gerçekleştiği kısmı oluşturmaktadır. Benzer şekilde, diğer bir yaklaşım ile tüm ağırlıkların gerçekleşen hata üzerindeki payları belirlenmektedir ve ağırlıklar bu mantığa göre yeni değerlerine sahip olmaktadır. Ağırlıklar güncellenirken belirli bir orana göre artmaktadır veya azalmaktadır. Bu miktara öğrenme katsayısı (λ) ismi verilmektedir ve bu katsayı, ağırlıkların değişim miktarlarını belirlemektedir. Bu değer çok küçük seçilirse öğrenme işlemi yavaşlamaktadır. Aksi şekilde büyük seçilmesi durumunda ise öğrenme daha hızlı olsa da ağırlar global minimum noktayı atlamasına neden olabilmektedir.

Gradyan tabanlı öğrenme algoritmaları, hata eğrisi üzerinde minimum hata noktasını bulmaya çalışmaktadırlar. Hata fonksiyonu üzerinde global ve yerel minimum noktaları bulunmaktadır. Bu öğrenme türünde yerel minimum noktalarda türev sıfır olacağından algoritma, yerel minimum noktaya takılıp global minimum olarak kabul edebilmektedir. Ağırlar yerel minimum noktalara takılmasını önlemek amacıyla momentum katsayısı (α) kullanılmaktadır.

Şekil 3.15'te bir hata eğrisi üzerinde ki global ve yerel minimum noktalar verilmiştir. $\varepsilon(i)$ hata değerleri ve i ilgili sinir ağırlarının iterasyon değerlerini göstermektedir.



Şekil 3.15. Hata eğrisi üzerindeki minimum noktalar.

Şekil 3.14'deki ağ mimarisi göz önünde bulundurulursa ilgili yapıda ağırlık değerlerinin güncellenmesi işlemi sırasıyla w_i , w_{i+1} ve w_{i+2} ağırlık değerleri için gerçekleştirilmektedir. Bu hesaplamada kullanılacak eşitlik Denklem 3.17'de verilmiştir.

$$w_i^{yeni} = w_i - \lambda \cdot \frac{\partial \epsilon_{Toplam}}{\partial w_i} \quad (3.17)$$

Eşitliğinde görüldüğü üzere, toplam hatanın güncellenmesi istenen ağırlık değerine göre kısmi türevi alınmaktadır ve belirlenen öğrenme katsayısı ile çarpılmaktadır. Daha sonra ilgili ağırlık değerinden bu elde edilen sonuç çıkartılır ve böylece ilgili ağırlığın değeri güncellenmiş olmaktadır. Bu formülasyon, tek katmanlı bir mimaride ve w_i ağırlığı için verilmektedir.

Hatanın geriye doğru yayılma işlemine ilişkin kısmi türev alma ve zincir kuralı Denklem 3.18'de verilmiştir.

$$\frac{\partial \epsilon_{Toplam}}{\partial w_i} = \frac{\partial \epsilon_{Toplam}}{\partial \epsilon_i} \cdot \frac{\partial \epsilon_i}{\partial OUT_i} \cdot \frac{\partial OUT_i}{\partial NET_i} \cdot \frac{\partial NET_i}{\partial w_i} \quad (3.18)$$

Burada, modelde kullanılan transfer fonksiyonunun sigmoid fonksiyon olduğu düşünülürse, $\frac{\partial OUT_i}{\partial NET_i}$ ifadesi sigmoid fonksiyonun türevini ifade etmektedir. Bunun nedeni OUT_i ifadesini elde ederken NET_i değişkenini kullanmış olmamızdır. Benzer mantıkla zincir kuralında, toplam hatadan başlamak üzere güncellenecek ağırlığa

kadar geriye doğru gidilmektedir ve ilgili değerin, o değeri elde ederken kullanılan parametreye göre kısmi türevi alınmaktadır. Buna göre $\frac{\partial OUT_i}{\partial NET_i}$ ifadesinin hesaplanmasında kullanılacak eşitlik Denklem 3.19’da verilmiştir.

$$(OUT_i)' = \frac{e^{-NET_i}}{(1 + e^{-NET_i})^2} = \frac{e^{-NET_i} + 1 - 1}{(1 + e^{-NET_i})^2} = \frac{1}{1 + e^{-NET_i}} - \frac{1}{(1 + e^{-NET_i})^2} \quad (3.19)$$

Eşitliğin sonucunda da görüldüğü üzere $(OUT_i)'$, OUT_i ’nin türevini ifade etmektedir ve bu türevin sonucu Denklem 3.20’de verilmiştir.

$$(OUT_i)' = OUT_i \cdot (1 - OUT_i) \quad (3.20)$$

Bu işlemler tüm ağırlık ve bias değerleri için hesaplanarak yeni değerleri bulunmaktadır.

3.2.2. Performans ölçüm metrikleri

Mevcut verileri kullanarak gelecekteki durumları belirleme işlemi, tahmin etmek olarak tanımlanmaktadır. Gerçekleşen tahminler ile beklenen çıktılar arasında oluşan farklar ise hata değerleridir. Geliştirilen bir modelin performansı bu hata değerine bağlıdır ve performansı ölçmek için çeşitli metrikler kullanılmaktadır [52-53]. Bu bölümde tahmin işlemlerinde modelin performansını ölçerken kullanılan metrikler verilmiştir. Bu çalışma kapsamında Ortalama Mutlak Yüzdesel Hata (MAPE) kullanılmıştır. Verilen tüm formüllerdeki er_i , hata değerini ifade etmektedir ve değeri Denklem 3.21’de verilmiştir. Burada OUT_i değeri modelden elde edilen çıktıyı ifade ederken, t_i değeri beklenen çıktı değerini ifade etmektedir.

$$er_i = OUT_i - t_i \quad (3.21)$$

3.2.2.1. Ortalama hata (ME)

Bir modelin tahmin ettiği değerler ile beklenen hedef değerleri arasındaki ortalama hatayı tanımlamaktadır. Eşitliği Denklem 3.22’de verilmiştir.

$$ME = \frac{1}{n} \sum_{i=1}^n er_i \quad (3.22)$$

3.2.2.2. Ortalama yüzdesel hata (MPE)

Bir modelde, tahmin edilen değerler ile gerçekleşmesi beklenen değerler arasındaki farkın yüzdesel hata değeridir. Hesaplanması aşamasında beklenen değerlere bölüldüğünden, bu değerler arasında 0 değeri varsa tanımsız durumda kalmaktadır. Denklem 3.23'te verilen eşitlik ile hesaplanmaktadır.

$$MPE = \frac{100}{n} \sum_{i=1}^n \frac{er_i}{t_i} \quad (3.23)$$

3.2.2.3. Ortalama mutlak hata (MAE)

Ortalama hataya benzemektedir fakat bu yöntemde hata değerlerinin kendisi yerine mutlak değerleri alınmaktadır. Bu sayede tahmin edilen değerler ile beklenen çıktı değerleri arasındaki mesafeyi vermektedir. Negatif ve sıfır değer içeren problemlerde kullanılması yönünden avantajlıdır. Denklem 3.24'te bu yöntemin eşitliği verilmiştir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |er_i| \quad (3.24)$$

3.2.2.4. Ortalama kare hata (MSE)

Tahmin değerleri ile beklenen çıktı değerleri arasındaki farkın kareleri toplamının veri sayısına bölümü ile elde edilmektedir. Özellikle regresyon analizi problemlerinde modelin performansını ölçmek için kullanılmaktadır. Değerinin sıfıra yaklaşması tahmin performansının iyileştini göstermektedir. Denklem 3.25'te ki eşitlik ile hesaplanmaktadır.

$$MSE = \frac{1}{n} \sum_{i=1}^n er_i^2 \quad (3.25)$$

3.2.2.5. Kök ortalama kare hata (RMSE)

Ortalama kare hatanın karekökü alınarak hesaplanmaktadır. Tahmin hatalarının regresyon modeli hattının etrafındaki yayılımına göre modelin performansını belirlemektedir. Değerinin sıfıra yakın olması, MSE ölçütünde olduğu gibi model

performansının iyi olduğu anlamına gelmektedir. Denklem 3.26’da verilen eşitlik ile hesaplanmaktadır.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n er_i^2} \quad (3.26)$$

3.2.2.6. Ortalama mutlak yüzdesel hata (MAPE)

Bu yöntemde, tahmin edilen değerler ile beklenen gerçek çıktılar arasındaki fark alındıktan sonra gerçek değere oranlanmaktadır. Daha sonra bu sonucu yüzdesel olarak ifade etmek için 100 ile çarpılmaktadır. Özellikle zaman serileri ve regresyon problemlerinde yaygın kullanılmaktadır. Beklenen çıktı değerlerine oranlama yapıldığından, bu değerler arasında 0 varsa tanımsız durumuna düşmektedir. Bu durumu engellemek için veriler arasında temizleme veya performansı etkilemeyecek şekilde beklenen çıktı değerleri yeniden düzenlenebilmektedir. MAPE değeri ne kadar düşük ise modelin performansı da o kadar yüksek olduğu anlamına gelmektedir. Denklem 3.27’de bu yöntemin hesaplanması için kullanılan eşitlik verilmiştir.

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{er_i}{t_i} \right| \quad (3.27)$$



4. TÜKETİM TAHMİNİ VE SONUÇLAR

Bu bölümde, bir önceki bölümde verilen veri setleri ve anlatılan yöntemler kullanılarak eksik sayaç okumalarından kaynaklanan kayıp doğal gaz belirlenmeye çalışılmıştır. Elde edilen bulgulardan faydalanılarak, tüketimlerinde anomali tespit edilen kullanıcılar incelenmiştir.

4.1. Modelleme

4.1.1. Abone gruplarının ve veri sayılarının belirlenmesi

Bölüm 3.1.5'te oluşturulan nihai veri setinde 435.440 adet doğal gaz abonesinin 17.162.589 adet fatura bilgileri yer almaktadır. Bu fatura verileri, her bir sayaç okuma işleminde oluşturulmuştur ve ilgili abonelerin tüketim miktarlarını içermektedir. Bu çalışmada, abone bazlı kayıp doğal gaz tüketimine bağlı tüketimdeki anomali tespit edilmeye çalışılmıştır. Öncelikle abonelerin tüketimlerini etkileyen parametrelerden takvimsel olaylar ve meteorolojik veriler 12 aylık tüketim periyodu içerisinde değişkenlik göstermektedir. Bu nedenle, bir abonenin verileri kullanılarak modelin olusturulmasi için en az 12 aylık bir tüketim verisine sahip olması gerekmektedir. Böylece ilgili model, tüketimleri etkileyecek olan tüm durumları görmüş ve genelleyici sonuçlar üretmesi için daha doğru eğitilmiş olacaktır. Bu durum göz önüne alınarak fatura sayısı 12'nin üzerinde olan aboneler verisetinde bulunmaktadır ve kullanılan abone sayısı 307.570 olmuştur. Bu bağlamda her bir abonenin fatura sayılarını belirlemek için veri tabanı üzerinde bir view oluşturulmuştur. Tablo 4.1'de ilgili view tablosunun örnek verileri verilmiştir.

Tablo 4.1. Abonelerin fatura sayıları.

Abone numarası*	Fatura sayısı
1	13
2	25
3	38
4	120

*: Veri gizliliği kapsamında sembolik ID atamaları yapılmıştır.

Abone bazlı modellerin elde edilmesi için öncelikle abone grupları belirlenmiştir. 12 aylık dönemler ve katları olacak biçimde abone grupları elde edilmiştir (Tablo 4.2). Bu yöntem, diğer tüketim dilimleri için uygulanarak aboneler on gruba ayrılmıştır. Tablodan görüldüğü üzere tüketim serisi arttıkça abone sayısında bir düşüş gerçekleşmektedir. Fakat veri sayısı, abonelerin fatura sayılarına bağlı olduğundan gruplar arasında hatırı sayılır bir fark meydana gelmemiştir.

Tablo 4.2. Abone grupları.

Grup adı	Fatura sayısı aralığı	Abone sayısı	Veri sayısı
1.Grup	$12 < FaturaSayisi < 24$	75.129	1.319.808
2.Grup	$24 \leq FaturaSayisi < 36$	53.614	1.562.951
3.Grup	$36 \leq FaturaSayisi < 48$	39.564	1.635.266
4.Grup	$48 \leq FaturaSayisi < 60$	28.862	1.542.718
5.Grup	$60 \leq FaturaSayisi < 72$	22.013	1.447.138
6.Grup	$72 \leq FaturaSayisi < 84$	20.082	1.559.379
7.Grup	$84 \leq FaturaSayisi < 96$	15.678	1.401.529
8.Grup	$96 \leq FaturaSayisi < 108$	15.063	1.529.355
9.Grup	$108 \leq FaturaSayisi < 120$	29.127	3.356.315
10.Grup	$120 \leq FaturaSayisi$	8.438	1.015.466

4.1.2. Modellerin oluşturulması ve kullanılan yaklaşım

Eksik sayaç okumalarından kaynaklanan kayıp doğal gazın belirlenmesinde en önemli kaynak tüketimlerdir. Aynı abone için tüketim davranışının değişikliğe uğraması beklenmeyen bir durumdur. Benzer sıcaklıklarda ve mevsimlerde, tüketim davranışında meydana gelen değişimler sırasıyla:

- Bina yalıtımları: Binaya yalıtım yapılması, pencerelerin değiştirilmesi vb.
- Hane halkında gerçekleşen değişiklikler: Yeni doğum, hane halkı sayısında azalma

olarak verilebilir. Fiziki olarak hatalardan meydana gelen değişimler sırasıyla:

- Sayacın hatalı okumaya başlaması,
- Sayaç üzerine fiziki müdahalede bulunulması,
- Sayaç okuma personeli hataları

olabilmektedir.

Bu okuma değişimleri ve hataların ne zaman başladığı her abone için bilinmesi imkansızdır. Bu sebeple ortak bir tüketim davranışı belirlenerek, bu davranışın hatalı olmadığı kabul edilmiştir. Bu durumun belirlenmesinde, en az 2 yıl ve üzerinde bir tüketime sahip olan abonelerde, eğitim için ilk yılın tüketim verilerini kullanmak yerine, ilgili abonenin tüm tüketimlerinin ortalamasına en yakın olan 12 aylık tüketim dönemi kullanılmıştır. Bu işlem için ilk olarak her bir abonenin toplam tüketimlerinin aritmetik ortalaması alınmıştır. İlgili abonenin verileri, on iki aylık veri içerecek şekilde alt gruplara ayrılıp her alt grubun kendi içerisinde aritmetik ortalaması alınmıştır. Daha sonra ilgili abonenin tüm tüketimlerinin ortalama değeri, on iki aylık ortalamaları ile karşılaştırılmıştır. Tüm tüketim verilerinin ortalamasına en yakın olan 12 aylık tüketim verileri, model oluşturmada eğitim verisi olarak seçilmiştir. Örnek olarak 5. grupta bulunan bir abonenin eğitim verisi seçilirken ilgili abonenin tüm tüketim ortalaması, ilk 5 yılın ayrı ayrı tüketim ortalamaları ile karşılaştırılmıştır. Bu işlemin amacı, veri setini ve kullanıcı davranışlarını en iyi temsil eden veri kümesini model eğitiminde kullanmaktır. Bu işlemde kullanılan eşitlik Denklem 4.1’de verilmiştir.

$$T_{ort} = \left(\sum_{i=1}^N T_i \right) \cdot \frac{1}{N} \quad (4.1)$$

Bu denklemde, T_{ort} ilgili abonenin tüketim ortalaması, T_i abonenin o ayki tüketim miktarı ve N ise abonenin sahip olduğu fatura sayısını ifade etmektedir.

Abonelerin her on ikişer aylık ortalama tüketimleri Denklem 4.2’deki gibi hesaplanmaktadır.

$$T_{Ort1.yıl} = \left(\sum_{i=1}^{12} T_i \right) \cdot \frac{1}{12} \quad (4.2)$$

Bu eşitlikte, $T_{Ort1.yıl}$ abonenin ilk 12 aylık zaman dilimindeki ortalama tüketimini ifade etmektedir. Bu işlem, tüm on ikişer aylık tüketimler için hesaplanmaktadır.

Daha sonra, abonenin tüketimlerinin aritmetik ortalaması ile ilgili yılların ortalama tüketiminin mukayesesi Denklem 4.3 ve Denklem 4.4'teki gibi yapılmaktadır.

$$T_{fark1} = |T_{ort} - T_{Ort1.yil}| \quad (4.3)$$

$$T_{fark2} = |T_{ort} - T_{Ort2.yil}| \quad (4.4)$$

Eşitliklerde verilen T_{fark1} ve T_{fark2} değerleri, ilgili yılın tüketiminin genel tüketime olan uzaklığını ifade etmektedir ve belirtildiği üzere tüm fark değişkenleri mukayese edilerek en küçük değere sahip olan yılın verileri eğitim için kullanılmıştır. Bu işlem Denklem 4.5'te verilmiştir.

$$T_{Egitim} = \min_{1 \leq fark_i \leq N} (T_{fark1}, T_{fark2}, \dots, T_{farkN}) \quad (4.5)$$

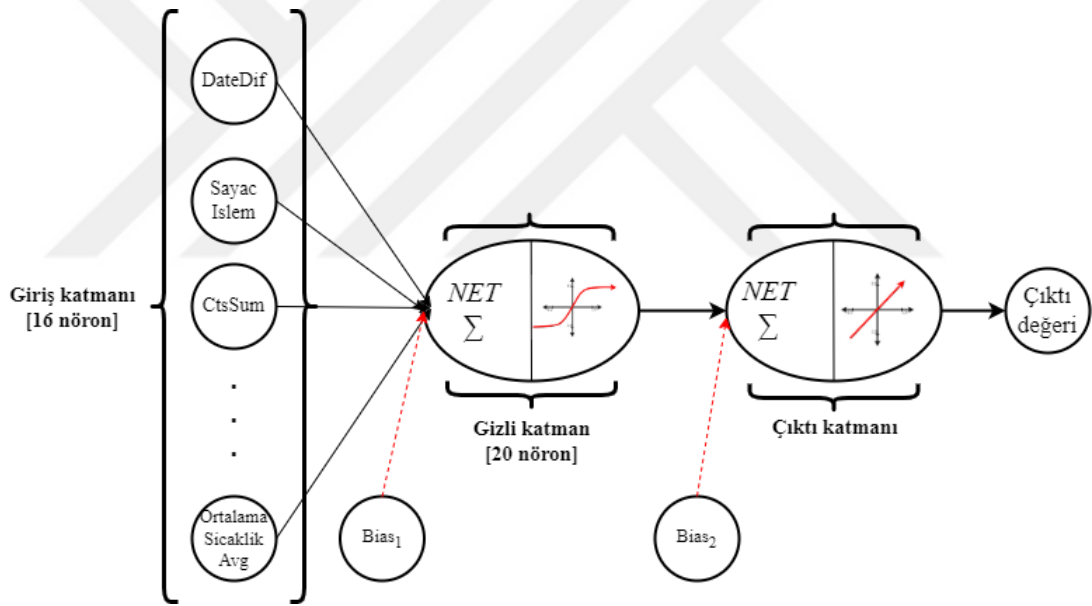
Bu işlemler sonucunda 307.570 adet abonenin verileri, doğal gaz tüketim tahmininde kullanılmak için uygun görülmüştür.

4.1.3. Modellerin eğitilmesi ve tüketim tahminleri

Bu çalışmada, 1 gizli katmana sahip ileri beslemeli bir YSA modeli oluşturularak önceki bölümde belirlenen 307.570 adet abonenin verileri kullanılarak abone bazlı doğal gaz tüketim tahmini gerçekleştirilmiştir. Her abonenin tüketim tahmini için ilgili abonenin verileri bir önceki bölümde belirtilen kural çerçevesinde kullanılmış olup toplamda 307.570 adet YSA modeli ve bu modellerin çıktıları elde edilmiştir. Nihai veri setinden abone numarası ve tüketim değerleri çıkartılarak geriye kalan değişkenler YSA'nın girdisi, tüketimler ise çıktı parametresi olarak seçilmiştir. Böylece giriş katmanındaki nöron sayısı, girdi parametresi sayısı ile aynı olup 16'dır. Gizli katmanda 20 adet nöron bulunmaktadır. Çıktı katmanı ise 1 adet çıkış nöronuna sahiptir. Çıktı katmanında elde edilen hedef değerler, doğal gaz tüketim miktarlarıdır.

Tüm abonelerin veri sayıları 16.369.925 satır ve abone numarası sütunu dahil 18 sütundan oluştuğundan veri seti üzerinde arama yaparak her bir abonenin verilerini tespit etmek ve YSA modelinde işlem yapmak yavaşlamaktadır. Bu durumu en aza indirmek için Tablo 4.2'de verilen abone grupları dikkate alınarak veriler ve abone numaraları 10 gruba ayrılmıştır. İlk olarak 1. grupta bulunan abone numaraları bir vektör olarak değişkene atanmıştır. Daha sonra 1. grupta bulunan 1.319.808 veri

arasından, abone numarası bilgilerinin bulunduğu vektörün ilk elamanının sahip olduğu veriler bir döngü yapısıyla abone numarası bilgileri karşılaştırılarak aranmış ve ilgili abonenin verileri ayrı bir matrise atanmıştır. Elde edilen matris üzerinden belirlenen sütun değerleri YSA modelinin eğitimi için giriş değerleri ve hedef çıktı değeri doğal gaz tüketim miktarı olarak belirtilmiştir. 1 yılın tüketim verileri ile model eğitimi gerçekleştirilip daha sonra ilgili abone için geriye kalan veriler ile test işlemi gerçekleştirilmiştir. Aynı işlemler, 1. Grup'ta bulunan diğer 75.128 abone için gerçekleştirilmiştir. Bu işlemler sonucunda elde edilen test işleminin ve eğitim işleminin MAPE değerleri, her bir abonenin tahmin edilen ve gerçekleşen tüketim grafikleri ve gerçek tüketimlerinin standart sapma değerleri kaydedilmiştir. Böylece 1. Grup için 75.129 adet YSA modeli ve sonuç değerleri elde edilmiştir. Bu işlemler diğer tüm gruplar için uygulanmıştır. Şekil 4.1'de bu işlemde kullanılan YSA modelinin mimarisi verilmiştir.



Şekil 4.1. Çalışmada kullanılan YSA mimarisi.

Gizli katmanda transfer fonksiyonu olarak tan-sigmoid fonksiyonu ve çıktı katmanında doğrusal fonksiyon kullanılmıştır. Öğrenme katsayısı, 0.01 olarak belirlenmiştir. Modelin eğitilmesi aşamasında, Denklem 4.5'te elde edilen $T_{Eğitim}$ tüketim veri setinin %100'ü eğitim için kullanılmıştır. Burada önemli olan nokta, eğitimde kullanılan verinin tüm veri setine oranı ortalama %22,54 olarak elde edilmiştir. Diğer bir ifadeyle verinin %77,45'i test verisi olarak kullanılmıştır. Eğitim, yapay zeka tekniklerinde eğitim veri seti %20 - %40 arasında değişirken, test veri seti %60 - %80 arasında değişmektedir. Bu çalışmada verinin büyük kısmı kayıp

ve hesaplanamayan tüketimleri belirleyebilmek için test veri setinde bulundurulmuştur. Daha sonra ilgili tüketicinin tüm verilerinden eğitim işleminde kullanılan veriler çıkartılmış ve geriye kalan veriler, test işleminde kullanılmıştır. Eğitim gerçekleştirildikten sonra, test verileri de verilerek abone bazlı modellerin hataları sınanmış ve MAPE değeri elde edilmiştir.

Çalışmada, sayaçlardan kaynaklı olan kayıp doğal gaz tüketimi araştırıldığı için model eğitimi iki şekilde gerçekleştirilmiştir. Bunlardan ilki; veri setinde belirtilen sayaç işlem sütunu için abonelerin sayaçlarına işlem gerçekleştirildiğinde o noktadan itibaren yenilenmiş bir sayaç ile tüketimlerine devam ettiklerini belirtmek ve modele göstermek amacıyla sayısal değerleri değiştirilmiştir. Bu model tezde “N” ile gösterilmiştir. Diğer işlem ise, sayaç işlem bölgelerini noktasal olarak göstermektir. Oluşturulan veri seti üzerinde zaten sayaç işlem bilgileri noktasal verildiğinden bu işlem için ilgili sütuna herhangi bir değişiklik yapılmamıştır. Bu model tezde “P” ile gösterilmiştir. Sayaç etkisi model eğitiminde kullanılan seri, Şekil 4.2’de verilmiştir.

FaturaDonemi	1	2	3	4	5	6
SayacIslem	0	0	0	0	1	1
FaturaDonemi	7	8	...	24	25	26
SayacIslem	1	1		1	2	2
FaturaDonemi	...	46	47	48		
SayacIslem		2	3	3		

Şekil 4.2. Sayaç işlemlerinin bağıl zamansal gösterimi.

Yukarıda verilen şekilde görüldüğü üzere abonenin sayacına işlem gerçekleştirildiği tüketim noktasından itibaren değerler güncellenmiştir. Bu işlem için tüm abonelerin sayaçlarına kaç kere işlem gerçekleştirildiği bilgisi Denklem 4.6 ile elde edilmiştir.

$$TSI = \sum_{i=1}^N Ti_{[s]} \quad (4.6)$$

Bu ifadede TSI , toplam sayac işlem sayısını göstermektedir. $Ti_{[s]}$, ilgili abonenin sayaç işlem sütununu ve N ise abonenin sahip olduğu toplam fatura sayısıdır. Sayaç işlem bilgilerini Şekil 4.2’de görüldüğü üzere bağıl olarak ifade etmek için her

abonenin sayaç işlem noktalarının indisleri elde edilmiştir. Bu işlem Denklem 4.7 ile gerçekleştirilmiştir.

$$SI_{i \in N} = \text{find}(Ti_{[s]} = 1) \quad (4.7)$$

Bu eşitlik ile, abonelerin sayaçlarına işlem gerçekleştirilen satırların indisleri elde edilmektedir ve $SI_{i \in N}$ vektörüne atanmaktadır. Elde edilen bu vektörden yararlanarak sayaç işlem bilgilerinin bağlı olarak güncellenme işlemi Denklem 4.8 ile gerçekleştirilmiştir.

$$Ti_{[s]} = \begin{cases} 0, & [Ti_{[1,s]}, Ti_{[SI_{[1]}-1,s]}] \\ 1, & [Ti_{[SI_{[1]},s]}, Ti_{[SI_{[2]}-1,s]}] \\ 2, & [Ti_{[SI_{[2]},s]}, Ti_{[SI_{[3]}-1,s]}] \\ 3, & [Ti_{[SI_{[3]},s]}, Ti_{[N,s]}] \end{cases} \quad (4.8)$$

Belirtilen eşitlikte örnek olarak TSI değeri 3 olan bir abone için değerler atanmıştır. Sayaç işlemlerinin modele noktasal olarak verildiği serinin bir örneği Şekil 4.3'te verilmiştir. Burada sayaç işlem sütununda bulunan verilere herhangi bir işlem uygulanmadan doğrudan modele dahil edilmiştir.

FaturaDonemi	1	2	3	4	5	6
SayacIslem	0	0	0	0	1	0
FaturaDonemi	7	8	...	24	25	26
SayacIslem	0	0		0	1	0
FaturaDonemi	...	46	47	48		
SayacIslem		0	1	0		

Şekil 4.3. Sayaç işlemlerinin noktasal gösterimi.

4.2. Tahmin Sonuçlarının Elde Edilmesi

Tüm abone grupları ve içerisinde mevcut olan tüm aboneler, belirtilen iki durum ile modelleri eğitilmiş ve testleri gerçekleştirilmiştir. Bu sayede, her bir abonenin tüketim tahminine bağlı olan test işlemlerinin MAPE değerleri üzerine sayacın etkileri belirlenmiştir. Bahsi geçen iki model, tüm abone grupları için Intel Core i7-6700K 4.00GHz işlemci ve 64 GB hafızaya sahip bir bilgisayarda çalıştırılmıştır. Tüm gruptaki aboneler için eğitim ve test işlemleri gerçekleştirilip sonuçlar

kaydedilirken, modellerin çalışma süreleri de elde edilmiştir. İki model için de aynı anda beş adet abone grubunun işlemleri paralel yürütülmüştür. Abone grupları içerisinde her bir abonenin verileri, veri seti içerisinde arandığından dolayı veri sayısı fazla olan abone gruplarının birlikte çalışması, tamamlanma süresini etkileyebilmektedir. Bu bağlamda, modellerin çalışma süreleri Tablo 4.3'te verilmiştir. Tabloda da görüldüğü gibi abone sayısı azaldıkça eğitim süresi de kısalmaktadır. Grup 9'da abone sayısı artış gösterdiğinden çalışma süresi de abone sayısına oranla artmıştır. Tez çalışmasındaki tüm eğitimler 20 günde tamamlanmış, işlem süresi olarak ise 59,31 gün sürmüştür.

Tablo 4.3. Modellerin çalışma süreleri.

Grup adı	Abone Grubu		Sayaç işlemleri bağıl olarak verilen model		Sayaç işlemleri noktasal olarak verilen model	
	Abone sayısı	Veri sayısı	Çalışma süresi (saniye)	Çalışma süresi (saat)	Çalışma süresi (saniye)	Çalışma süresi (saat)
1.Grup	75.129	1.319.808	751.440,65	208,734	724.517,72	201,255
2.Grup	53.614	1.562.951	412.731,07	114,648	636.313,94	176,754
3.Grup	39.564	1.635.266	265.623,568	73,784	457.165,14	126,990
4.Grup	28.862	1.542.718	186.042,412	51,678	290.415,75	80,671
5.Grup	22.013	1.447.138	135.759,783	37,711	189.219,89	52,561
6.Grup	20.082	1.559.379	99.694,454	27,693	113.929,71	31,647
7.Grup	15.678	1.401.529	75.495,96	20,971	79.448,45	22,069
8.Grup	15.063	1.529.355	82.851,222	23,014	69.356,70	19,266
9.Grup	29.127	3.356.315	197.639,36	54,900	265.597,30	73,777
10.Grup	8.438	1.015.466	41.440,310	11,511	50.162,10	13,934

Model çıktılarından elde edilen eğitim ve test işlemlerinin gruplar bazında ortalama MAPE değerleri Tablo 4.4'te verilmiştir. Tabloda MAPE değerleri

- Tüm aboneler,
- Sayacı yenilenen aboneler,
- Sayacı yenilenmeyen aboneler

olarak gruplandırılmıştır. Tüm gruplar içerisinde sayacı en az bir kere yenilenmiş olan abonelerin sayısı 69.515'tir. Bu da toplam abonelerin %22,60'ını oluşturmaktadır. Burada önemli olan nokta, birinci grup abonelerinin fatura sayıları düşük olduğu için test veri setinde bağıl ve noktasal MAPE değerleri tüm abone

türlerinde büyük farklılıklar göstermiştir. Çalışmadaki test işleminin tüm aboneler için ortalama MAPE değeri 2,27 olurken, bağıl ve noktasal modellerden elde edilen ortalama MAPE değerleri ise sırasıyla 2,42 ve 2,12'dir. Çalışma kapsamındaki abonelerden sayacı en az bir kere yenilenmiş abonelerin bağıl ve noktasal modelden elde edilen ortalama MAPE değerleri sırasıyla 2,43 ve 2,35 olmuştur. Sayacı daha önce hiç yenilenmemiş olan yani tüm aboneler içerisinde %77,40'lık dilimde bulunan abonelerin bağıl ve noktasal modelden elde edilen ortalama MAPE değerleri ise sırasıyla 2,37 ve 2,08 olmuştur. Tablo 4.4'teki veriler incelendiğinde ortalama MAPE değerlerinin artışına sebep olan abonelerin büyük çoğunluğunun grup 1 ve grup 2'ye dahil olduğu görülmektedir. Bunun sebebi ise yukarıda belirtildiği üzere bu iki gruba dahil olan abonelerin tüketim serilerinin az olmasıdır (Tablo 4.2). Tüm abonelerin gruplar bazında ortalama düzeltilmiş R^2 değerleri de tabloya eklenmiştir. Bu değer, "N" ve "P" modelleri tarafından gerçekleştirilen tüketim tahminlerinin abonelerin gerçek tüketimlerine olan yakınlığını ifade etmektedir ve 0 ile 1 arasında değişmektedir. Bu değerler incelendiğinde, ilgili gruplar içerisinde bulunan abonelerin test veri setlerinde bulunan tüketim sayısı arttıkça yapılan tahminlerin iyileştiği ve ortalama düzeltilmiş R^2 değerlerinin de yükseldiği görülmektedir (Tablo 4.4).

Tablo 4.4. Eğitim ve test veri setlerinin gruplar bazında ortalama MAPE ve düzeltilmiş R^2 değerleri.

Eğitim Veriseti							Test Veriseti						
Tüm aboneler		Sayacı Yenilenen Aboneler		Sayacı Yenilenmeyen Aboneler		Tüm aboneler		Sayacı Yenilenen Aboneler		Sayacı Yenilenmeyen Aboneler			
Abone Grubu	Bağıl	Noktasal	Bağıl	Noktasal	Bağıl	Noktasal	Bağıl	Noktasal	Ort.Adj .R ²	Bağıl	Noktasal	Bağıl	Noktasal
1.Grup	0,1137	0,1137	0,002	0,002	0,1194	0,1194	6,7334	3,8132	0,2709	4,7606	4,6782	6,8344	3,7689
2.Grup	0,0006	0,0006	0,0003	0,0003	0,0006	0,0006	2,158	2,2245	0,3433	2,2698	2,2025	2,1484	2,2264
3.Grup	0,0006	0,0006	0,0001	0,0002	0,0007	0,0007	2,0521	2,0909	0,3482	2,1678	2,1262	2,0377	2,0864
4.Grup	0,0011	0,0011	0,0009	0,0009	0,0011	0,0011	2,0527	2,0482	0,3408	2,1513	2,1023	2,0364	2,0393
5.Grup	0,0007	0,0007	0,0005	0,0006	0,0008	0,0008	2,0672	2,0557	0,3347	2,1878	2,1412	2,0395	2,0361
6.Grup	0,0004	0,0004	0,0007	0,0007	0,0003	0,0003	2,2685	2,1837	0,3289	3,3628	2,925	1,9596	1,9744
7.Grup	0,0008	0,0008	0,0009	0,0009	0,0007	0,0007	1,9616	1,9405	0,3361	2,2054	2,1836	1,8686	1,8477
8.Grup	0,0003	0,0003	0,0002	0,0002	0,0004	0,0004	1,932	1,9428	0,3350	2,2218	2,2298	1,7182	1,7311
9.Grup	0,0015	0,0015	0,0005	0,0005	0,0092	0,0092	1,7001	1,6828	0,3635	1,689	1,6654	1,7834	1,8134
10.Grup	0,0008	0,0008	0,0008	0,0008	0	0	1,334	1,3147	0,3962	1,3357	1,3139	1,2863	1,3347

Bağıl ve noktasal modellerden elde edilen sonuçların farkı alınarak her bir abone için bağıl modelden noktasal modele geçince tüketim tahmini başarısının değişme durumları incelenmiştir (Denklem 4.9).

$$MAPE_{Fark} = \left| MAPE_{TestBağıl} - MAPE_{TestNoktasal} \right| \quad (4.9)$$

Bu denklemde hesaplanan $MAPE_{Fark}$ değeri ile abonelerin test veri setlerindeki MAPE değerlerinin miktarı elde edilmiştir. Bu değişimler incelendiğinde üç farklı durum bulunmaktadır:

- Durum A: Sayacı yenilenen aboneler için sayaç işlemlerinin bağıl modelden noktasal modele geçişi sonrasında MAPE değeri artan aboneler.
- Durum B: Sayacı yenilenen aboneler için bağıl modelden noktasal modele geçiş yapınca MAPE değeri azalan aboneler.
- Durum C: Daha önce sayacına herhangi bir yenilenme işlemi uygulanmayan ve bağıl model - noktasal model MAPE değerindeki değişimleri düşük olan aboneler.

Tablo 4.5'te, A durumuna sahip farklı gruplardan örnek abonelerin iki modelden elde edilen test işlemlerinin MAPE değerleri verilmistir.

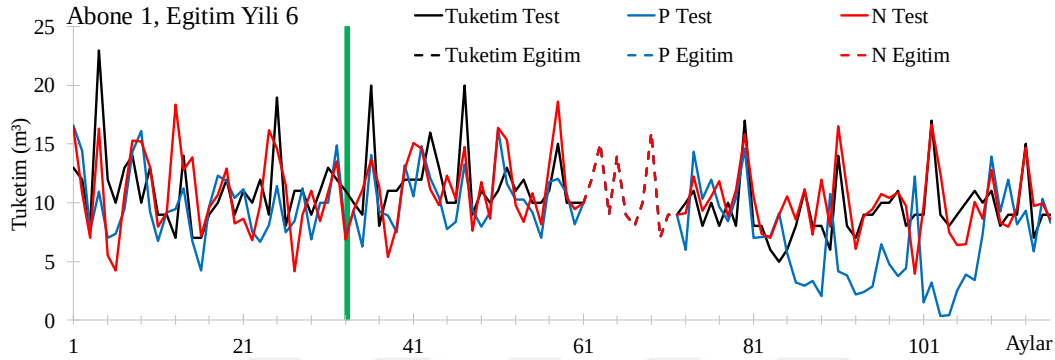
Tablo 4.5. Bağıl modelin daha başarılı olduğu abone örnekleri.

Grup no	Abone ID*	Bağıl model $MAPE_{Test}$	Noktasal model $MAPE_{Test}$	$MAPE_{Fark}$
9	1	0,2338	0,30996	0,07616
3	2	4,2915	14,6294	10,3379
6	3	0,16101	0,21907	0,05806
8	4	0,20665	0,32318	0,11653
7	5	0,27516	0,43495	0,15979
6	7	0,36496	0,96973	0,60477
4	8	0,36631	0,52936	0,16305

*: Veri gizliliği kapsamında sembolik ID atamaları yapılmıştır.

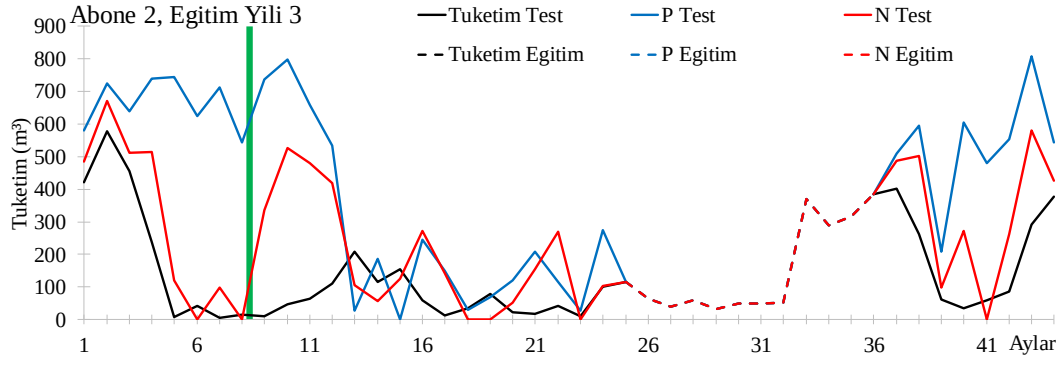
Şekil 4.4'te, yukarıdaki tabloda verilen 9 numaralı gruba ait olan ve 1 numaralı abone numarasına sahip abonenin iki model için elde edilen tüketim tahmin grafikleri verilmiştir. Bu abonenin sayaç işlem dönemi 33 olarak belirlenmiştir. Grafikte kesikli çizgiler eğitimde kullanılan veriyi ve eğitim verisetinin tahminlerini göstermektedir. Yine grafikte mavi renkte gösterilen “P” noktasal sayaç bilgisini, kırmızı renkte gösterilen “N” sürekli sayaç bilgisini göstermektedir. Verilen grafikte görüldüğü üzere ilgili abone için modeller arasındaki $MAPE_{Fark}$ değeri küçüktür. Bu

durum, sayaç deęişim bilgisinin modele verilmesi sonrasında doęal gaz tüketimi tahmin başarısında yüksek bir deęişimin meydana gelmediğini göstermektedir. Yenilenen sayaç tarafından gerçekleştirilen ölçümlerin doęru olduęu göz önünde bulundurulduğunda, bu duruma sahip abonelerde sayacın hata deęeri üzerine bir etkisinin olmadığı ve dolayısıyla eski sayaçlarında da bir anormallik bulunmadığı yönünde deęerlendirilmiştir. Bu durumda olan abonelerin sayaçları ve tüketim tahminleri normal kabul edilmiştir.



Şekil 4.4. 9 numaralı gruba ait örnek bir abone için baęıl ve noktasal modelin örnek çıktı grafikleri.

Şekil 4.5'te, Tablo 4.5 içerisinde bulunan 3 numaralı gruba ait ve 2 numaralı abone numarası ile verilmiş bir abonenin tüketim tahmini grafikleri verilmiştir. Verilen tüketim tahmini grafiklerinden görüldüğü üzere, ilgili aboneye sayaç yenilenme işlemi gerçekleştirildikten sonra baęıl modelin tahminleri iyileşirken, noktasal modelde bu durum söz konusu olmamıştır. Bunun nedeni, baęıl modelde yeni sayaç bilgisinin modele verilmiş olmasıdır. Sayaç yenilenme işleminden sonraki fatura dönemleri için gerçekleştirilen tüketim tahminleri, abonenin gerçek tüketimine daha yakın olduęun ve dolayısıyla gerçekleşen hata deęerleri düşük olduęu belirlenmiştir. İki model arasındaki MAPE deęeri farkları yüksek olduğundan bu duruma sahip aboneler anomali olarak belirlenmiştir.



Şekil 4.5. 2 numaralı gruba ait örnek bir abone için bağıl ve noktasal modelin örnek çıktı grafikleri.

Diğer bir durum olan B durumunda, sayacı yenilenen abonelerin noktasal modelde MAPE değerleri bağıl modele göre azalmıştır. Bu duruma sahip aboneler içerisinde kayıp doğal gaz tüketimi (UFG) potansiyeli taşıyan kullanıcıların bulunma riskinin yüksek olduğu düşünülmektedir. Bunun nedeni, sayaçların yenilenme bilgisinin modele dahil edilmesi sonrasında hata miktarında kaydedilen artıştır. Diğer bir ifade ile, hatasız ölçüm gerçekleştiren bir sayaca sahip abone için modellerin noktasal veya bağıl olması hata değerlerini anormal biçimde değiştirmeyecektir. Tablo 4.6’da, bu duruma sahip örnek abonelerin iki modelden elde edilen test işlemlerinin MAPE değerleri ve Denklem 4.9’da verilen eşitliğe göre hesaplanmış fark değerleri verilmiştir.

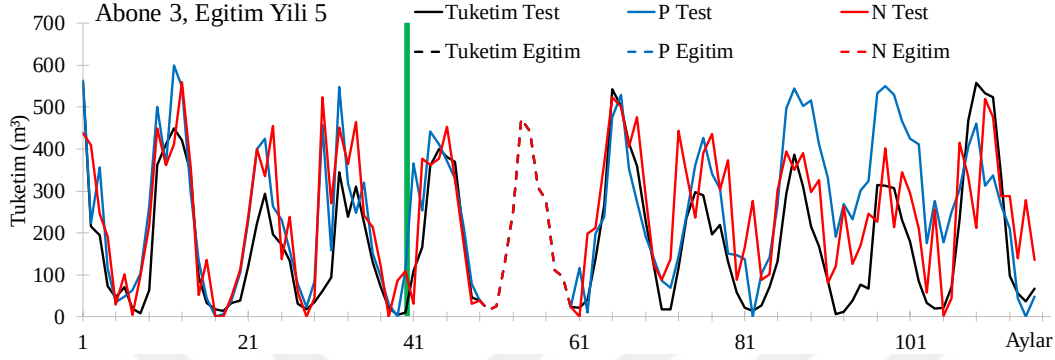
Tablo 4.6. Noktasal modelin daha başarılı olduğu abone örnekleri.

Grup no	Abone ID*	Bağıl model MAPE _{Test}	Noktasal model MAPE _{Test}	MAPE _{Fark}
9	3	1,9949	1,8723	0,1226
5	4	15,0843	0,9533	14,131
6	5	14,2873	1,2105	13,0768
4	6	13,8614	0,99063	12,87077
5	7	11,3839	0,87451	10,50939
10	8	11,237	0,89642	10,34058
4	9	10,3344	0,84979	9,48461

*: Veri gizliliği kapsamında sembolik ID atamaları yapılmıştır.

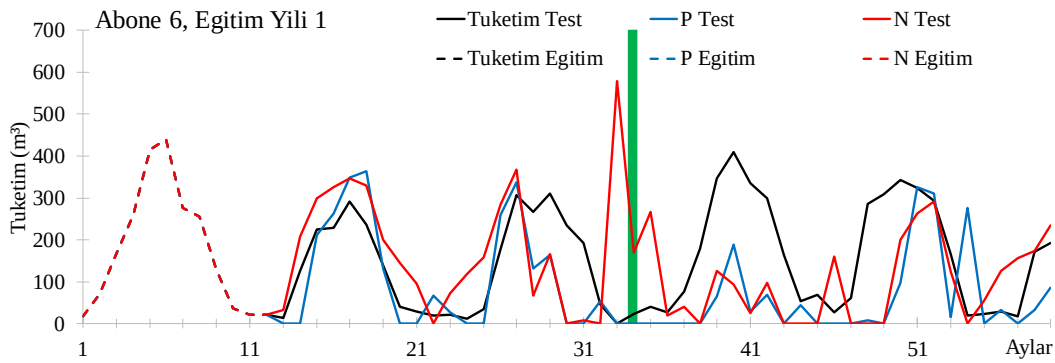
Şekil 4.6’da, yukarıdaki tabloda verilen 9 numaralı gruba ait olan ve 3 numaralı abone numarasına sahip abonenin iki model için elde edilen tüketim tahmin grafikleri verilmiştir. Bu abonenin sayaca son işlem yapıldığı dönem 40 olmuştur. Burada sayaç değişimi sonrasında tahmin hataları incelendiğinde N sürekli sayaç bilgisi bulunan modelin MAPE değerinin (2.08), P noktasal sayaç bilgisi bulunan MAPE

değerinden (2.20) daha düşük olduğu görülmüştür. Şekil 4.6’da verilen grafikte görüldüğü üzere ilgili abone için iki modelde gerçekleştirilen tüketim tahminlerinin hata değerleri arasındaki fark düşüktür. Şekil 4.4’te verilen abone ile benzer özellikleri taşımaktadır ve bu duruma sahip abonelerin UFG riski taşımadığı belirlenmiştir.



Şekil 4.6. 9 numaralı gruba ait örnek bir abone için bağıl ve noktasal modelin örnek çıktı grafikleri.

Şekil 4.7’de, Tablo 4.6 içerisinde bulunan 4 numaralı gruba ait ve 6 numaralı abone numarası ile verilmiş bir abonenin tüketim tahmini grafikleri verilmiştir. Sayaç işlem dönemi 34 olarak belirlenmiştir. Şekil 4.7’de verilen görseldeki grafiklerde görüldüğü üzere, ilgili abone için sayaç yenilenme bilgisinin noktasal verildiği model ile sayaç yenilenme bilgisinin sürekli verildiği modelin doğal gaz tüketim tahmininden elde edilen MAPE değeri sırasıyla 0.99 ve 13.86 olmuştur. Bu duruma sahip aboneler, modellerin hangisinin daha başarılı sonuç verdiği bakılmaksızın anormal olarak nitelendirilmiştir.



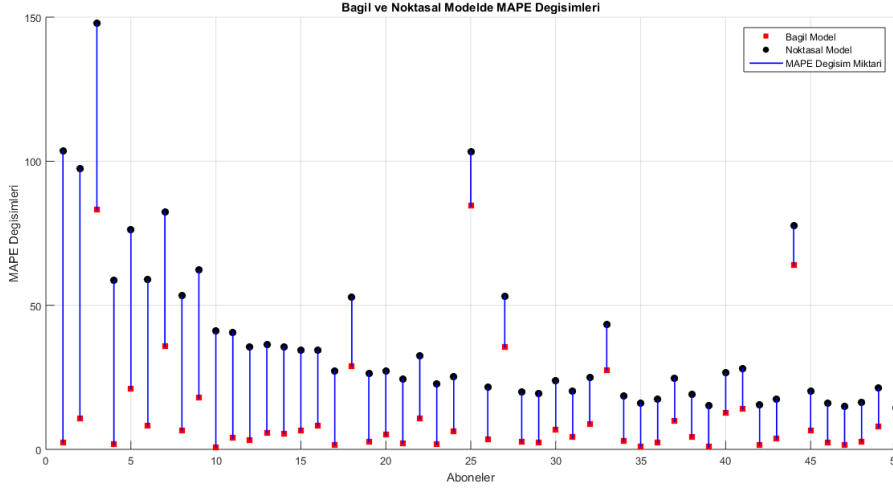
Şekil 4.7. 4 numaralı gruba ait örnek bir abone için bağıl ve noktasal modelin örnek çıktı grafikleri.

Son olarak, daha önce sayacına herhangi bir yenilenme işlemi uygulanmayan aboneler için gerçekleştirilen eğitim ve test işlemlerinde kullanılan sayaç işlem

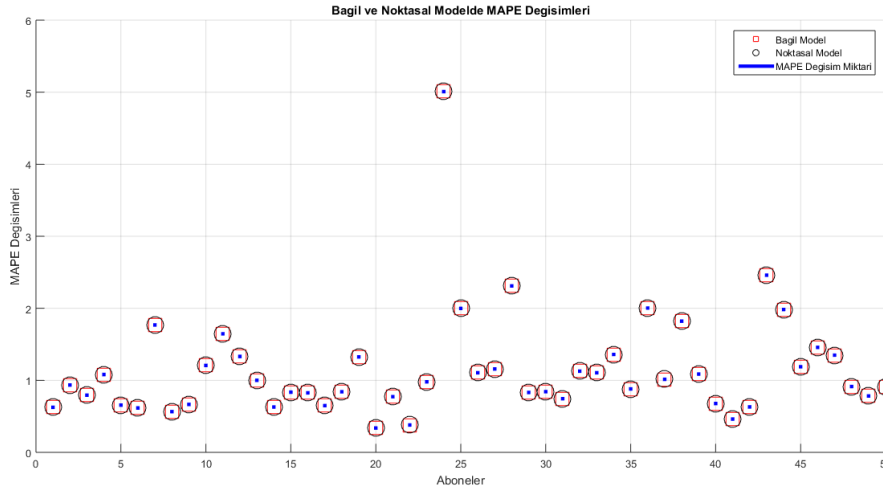
bilgisi verileri bağıl ve noktasal modelde birbirinin aynısı olacağından dolayı test işleminin MAPE değerlerindeki küçük değişimler göz ardı edilebilmektedir. Bu durum “C durumu” olarak adlandırılmıştır.

Verilen abone örnekleri ile benzer özellik taşıyan diğer tüm aboneler belirlenmiştir. Aboneler arasındaki anomali belirlenirken MAPE değerleri ve farkları göz önünde bulundurulduğundan, dönem sayısı düşük olan 1 ve 2. grupta bulunan aboneler (36 aya kadar tüketim bilgisi bulunanlar) değerlendirilmeye alınmamıştır. Bunun nedeni, bu iki grupta bulunan abonelerin eğitimleri diğer tüm abonelerde olduğu gibi 12 aylık tüketim serisi ile gerçekleştirilmiş olsa da, test işleminde veri sayısının az olmasıdır. Dolayısıyla test işlemi için tüketim serisi az olduğundan yapılan tahminler ve elde edilen MAPE değerleri, anomaliyi tespit edecek kadar yeterli olmamıştır.

A durumuna sahip olan abonelerin modellerden elde edilen MAPE farkları en yüksek olan ilk 50 abonenin ve son 50 abonenin grafiği büyükten küçüğe sıralı biçimde Şekil 4.8’de gösterilmiştir. Bu işlem, bağıl modelden elde edilen sonuçlar ile noktasal modelden elde edilen sonuçların aboneler bazında değişimlerini görebilmek adına gerçekleştirilmiştir ve ilgili grafikler fark büyüklüklerine göre dikey düşme şeklinde verilmiştir. Grafik üzerinde MAPE değişiminin en yüksek olduğu ilk 50 abone (a) görselinde, en düşük 50 abone ise (b) görselinde verilmiştir.



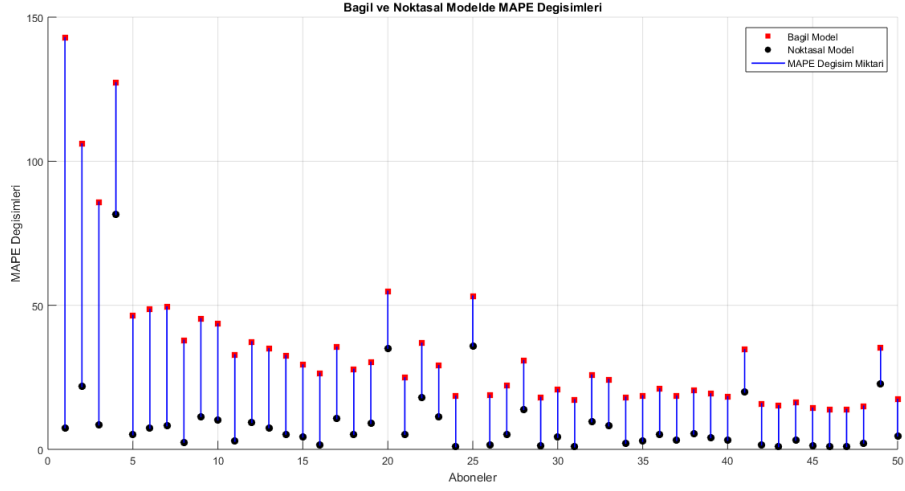
(a)



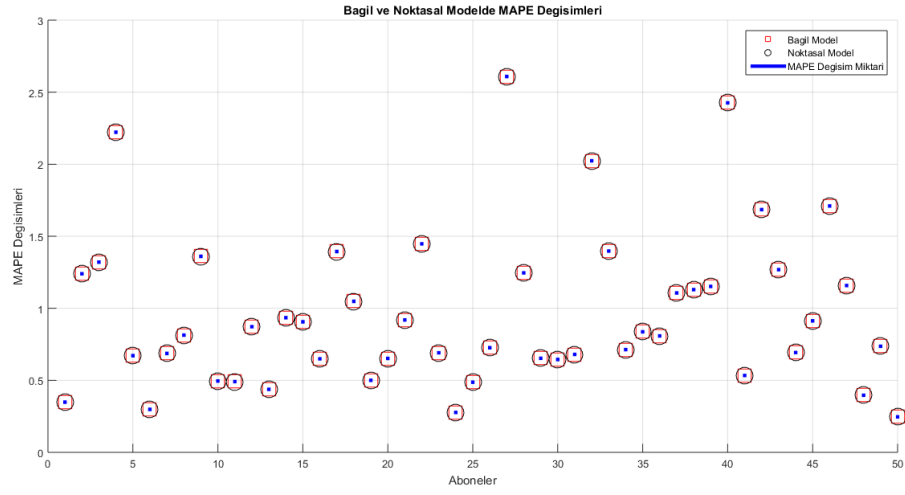
(b)

Şekil 4.8. A durumuna sahip aboneler için bağıl ve noktasal modelde MAPE değişimleri (a) Değişim miktarı en yüksek ilk 50 abone (b) Değişim miktarı en düşük 50 abone.

Şekil 4.8'dekine benzer işlem, B durumuna sahip aboneler için de gerçekleştirilmiştir ve bu abonelerin iki modelden elde edilen MAPE değerlerinin değişim durumları Şekil 4.9'da gösterilmiştir. Farkın en düşük olduğu abonelerin bulunduğu (b) görselinde, abone noktaları üst üste bindiğinden dolayı ilgili grafikte noktaların büyüklükleri değiştirilmiş ve dolulukları kaldırılmıştır. Bunun nedeni değişim miktarının çok küçük olduğunu grafik üzerinde daha iyi görebilmektir. Aynı işlem, noktasal modelden daha iyi sonuçlar alınan B durumuna sahip abonelerin verildiği grafikler için de gerçekleştirilmiştir.



(a)



(b)

Şekil 4.9. B durumuna sahip aboneler için bağıl ve noktasal modelde MAPE değişimleri (a) Değişim miktarı en yüksek ilk 50 abone (b) Değişim miktarı en düşük ilk 50 abone.



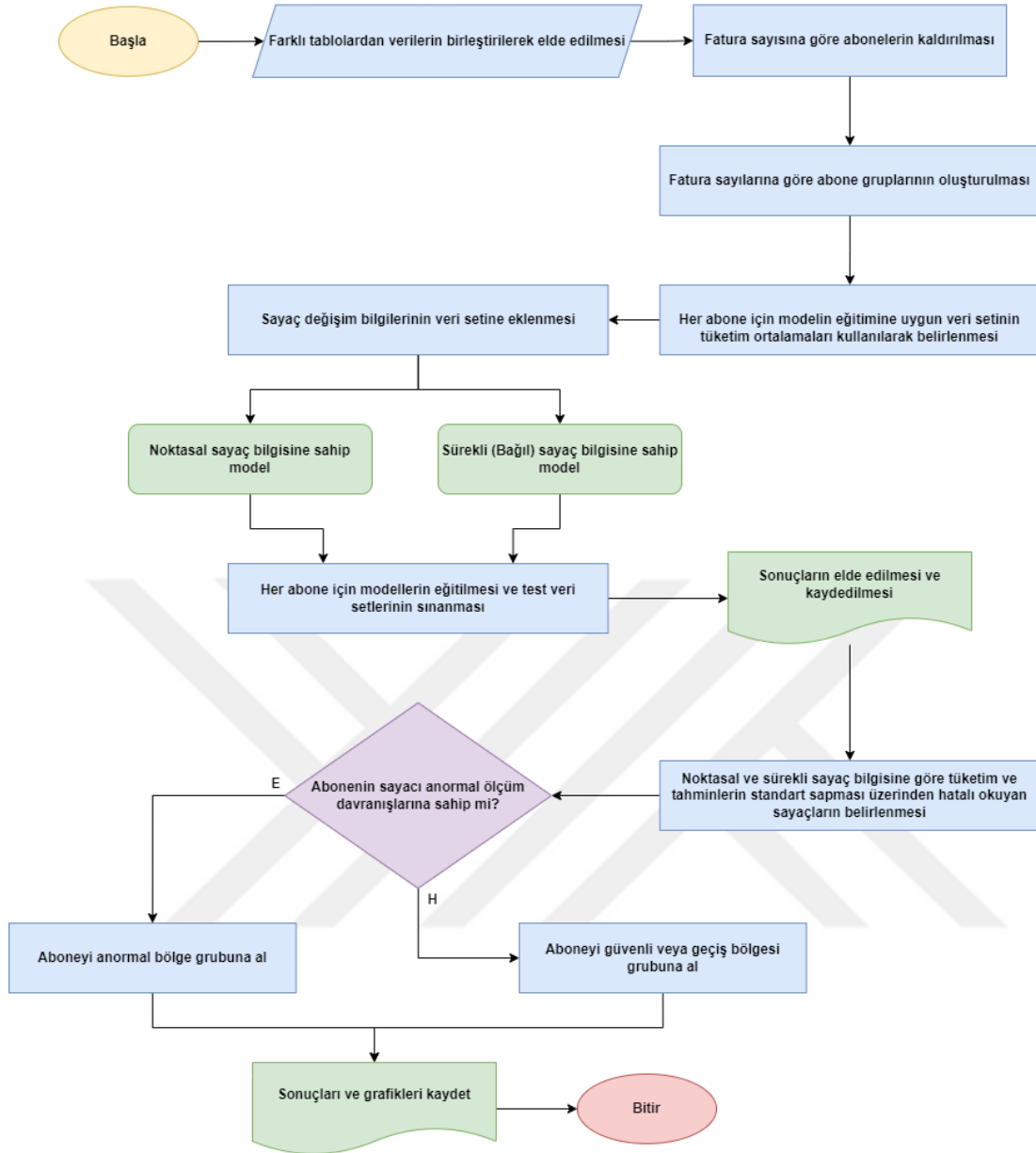
5. SONUÇ VE ÖNERİLER

Gerçekleştirilen tez çalışmasında kayıp doğal gaz tüketimi riskini taşıyan abonelerin belirlenmesi amaçlanmıştır. Bunun için çalışma kapsamında bulunan her bir abone için iki ayrı YSA modeli geliştirilmiş ve modellerin sonuçları ele alınmıştır.

Modellerin eğitilmesi aşamasında doğal gaz tüketim verileri, takvimsel olaylar verileri, sayaç bilgilerinin bulunduğu veriler ve meteoroloji verileri birleştirilerek kullanılmıştır. Böylece her bir abonenin tüketim dilimi içerisindeki tüketim alışkanlıklarının model tarafından öğrenilmesi amaçlanmış ve doğruluk oranı artırılmaya çalışılmıştır. Modellerden birincisi abonelerin sahip oldukları sayaçların yenilenme işleminden sonra yeni bir sayaç ile tüketimlerine devam ettikleri bilgisinin verildiği model, ikincisi ise sayaç yenilenme işlemlerinin ilgili fatura dönemi için noktasal olarak verildiği modeldir. Her bir abone için iki ayrı model ile doğal gaz tüketim tahminleri gerçekleştirilmiştir. Böylece yenilenen ve eski sayaçların doğal gaz tüketim tahmini başarısının üzerindeki etkisi elde edilmiştir. Aboneler tüketim serilerinin büyüklüklerine göre on ayrı gruba ayrılmıştır ve modellerin çalışma zamanları ilgili gruplardaki aboneler için yürütülmüştür. Daha sonra iki modelden elde edilen sonuçlar istatistiksel yöntemler ile analiz edilerek gaz okuma sayaçlarından kaynaklanan anormal tüketim davranışlarına sahip aboneler belirlenmeye çalışılmıştır.

5.1. Kayıp ve Hesaplanamayan Tüketime Sahip Abonelerin Belirlenmesi

Bu bölümde her abone için sayaç değişim bilgisine göre elde edilen model sonuçlarının kayıp ve hesaplanamayan tüketimin belirlenmesinde nasıl kullanıldığı gösterilmiştir. Gerçekleştirilen işlemler, Şekil 5.1'deki akış şemasında verilmiştir.



Şekil 5.1. Modeller ile gerçekleştirilen işlemlerin adımları.

Öncelikle noktasal ve bağıl sayaç bilgisi bulunan abonelerde sayaç kayıp ve okunamayan tüketim olmadığı varsayılmıştır. Bu durumda her iki durum için test verisetindeki MAPE değerlerinin benzer olması gerekmektedir. Bağıl sayaç bilgisi bulunan verisetinin MAPE değeri (“N”) ile, noktasal sayaç bilgisi bulunan verisetinin MAPE değeri (“P”) arasındaki ilişki basit doğrusal regresyon ile belirlenmeye çalışılmıştır. Burada elde edilmesi beklenen katsayı bire yakın olması gerekirken 0.68 olarak elde edilmiştir. MAPE değerlerinin tanımlayıcı istatistiklerine bakıldığında standart sapma, basıklık ve sivrilik değerlerinin anlamsız olduğu görülmüştür. Bunun sebebi MAPE değerlerinde uç değerlerin olmasıdır (Bknz maksimum 5396). Uç değerlerin sonuçlardan kaldırılması için art arda üç işlem

yapılmıştır. İlk olarak abone numarası en yüksek %5 sonuçlardan çıkartılmıştır. Abone numarası eskiden yeniye doğru artmaktadır. Son abonelerin tüketim dönem sayısı ve davranışlarının belirlenmesi daha zorlu olacağı için modelden kaldırılmıştır. Sonrasında, N ve P modellerinin en yüksek %5 MAPE değerlerine sahip verileri kaldırılmıştır. Böylece verisetinden %15 azaltılmıştır. Azaltılan veri kullanılarak tekrar basit regresyon modeli oluşturulmuş ve tanımlayıcı istatistikler çıkartılmıştır. Elde edilen regresyon modelinde katsayısının 0.68 değerinden 0.81 değerine yükseldiği görülmüştür. Bu durum, gerçeğe daha yakın sonuçların elde edildiğini göstermektedir. Devamında gerçekleştirilen tanımlayıcı istatistiklerde de standart sapma, sivrilik ve basıklık değerlerinin normal sınırlar içerisinde bulunduğu görülmüştür. Belirtilen doğrusal regresyon katsayıları Tablo 5.1’de, tüm ve azaltılmış verilerin tanımlayıcı istatistikleri Tablo 5.2’de verilmiştir.

Tablo 5.1. Tüm ve azaltılmış verinin regresyon katsayıları.

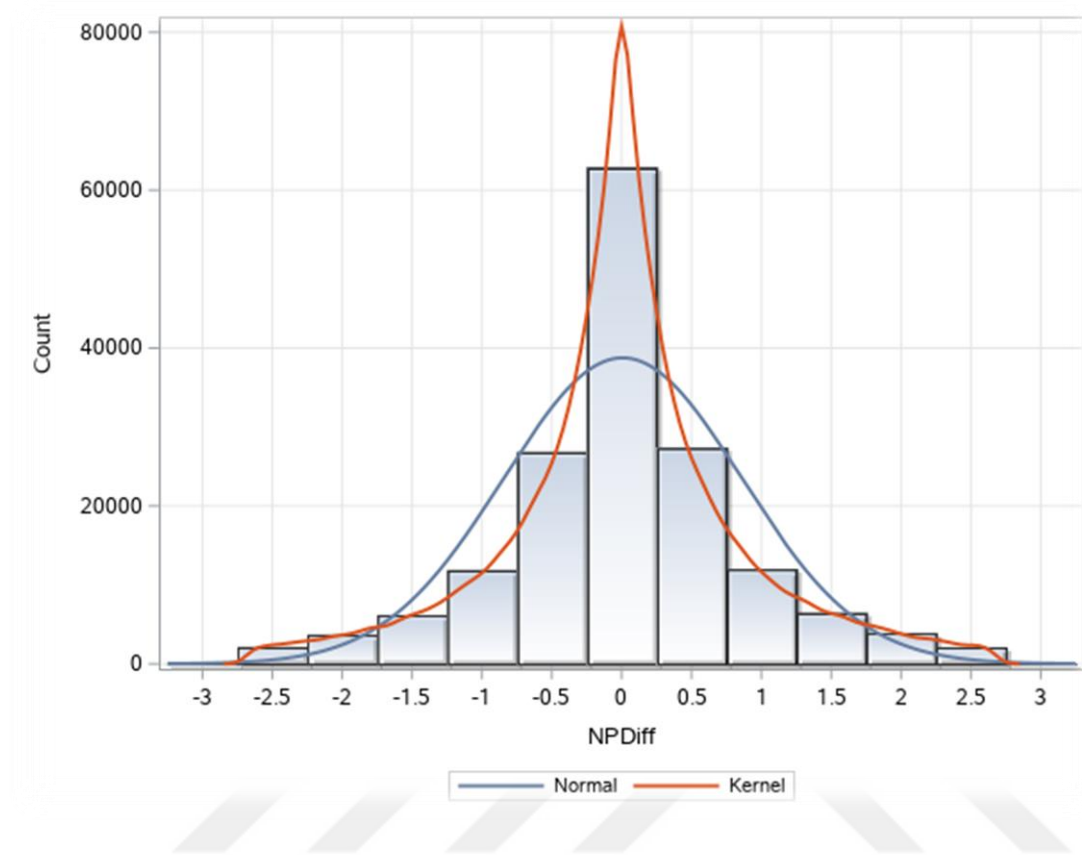
	Katsayılar	t Stat	P-value	Lower 95%	Upper 95%	R2	Gözlem Sayısı
Tüm Veri	0.68062	1060.545	0	0.679363	0.681878	0,8628	178827
Azaltılmış Veri	0.81796	680.250	0	0.815598	0.820312	0,7527	152004

Tablo 5.2. Tüm ve azaltılmış verinin tanımlayıcı istatistikleri.

	Tüm veri			Azaltılmış veri		
	N	P	NPFark	N	P	NPFark
Ortalama	1.969	1.961	0.008	1.518	1.461	0.057
Standart Hata	0.031	0.022	0.013	0.002	0.002	0.002
Medyan	1.310	1.302	0.000	1.226	1.212	0.008
Mod	1.081	1.154	0.000	1.081	1.041	0.000
Std. Sapma	13.039	9.461	5.527	0.933	0.829	0.894
Orn. Varyans	170.010	89.513	30.544	0.870	0.688	0.799
Basıklık	163818.022	123146.109	70705.626	1.744	0.752	2.512
Sivrilik	396.399	330.051	124.197	1.410	1.153	0.404
Aralık	5396.063	3633.426	2970.593	5.069	4.052	8.360
Minimum	0.078	0.122	-1208.000	0.078	0.127	-3.724
Maksimum	5396.141	3633.548	1762.593	5.148	4.179	4.635
Toplam	352125.454	350738.343	1387.111	230708.671	222016.828	8691.843
Eleman Sayısı	178827	178827	178827	152004	152004	152004

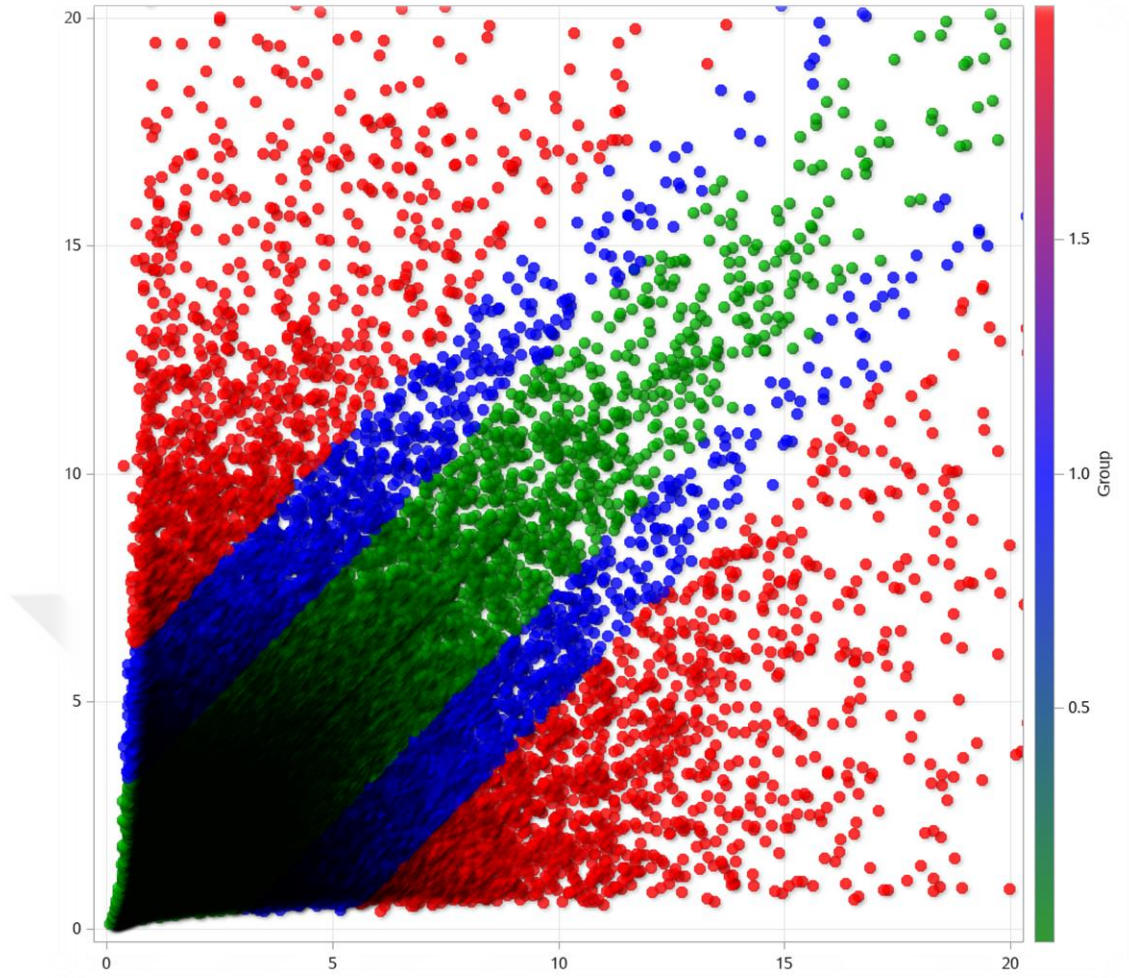
Azaltılmış veri ile MAPE farklarının dağılımı Şekil 5.2’de gösterilmiştir. Verinin normal dağılım gösterdiği de belirlenmiştir. Normal dağılım gereği %99,7’lik alan ± 3 standart sapma ile ifade edilmektedir. Azaltılmış veride bulunan standart sapmalar kayıp ve hesaplanamayan tüketim sınırlarını ifade etmekte kullanılacaktır. Azaltılmış veride bulunan ± 3 standart sapma aralığı normal davranışı gösteren sayaçlar olarak

ifade edilmiştir, bu alanın dışında kalan bölge ise iki alt gruba ayrılmıştır. Burada $[\pm 3 \pm 6]$ standart sapma aralığı ve ± 6 standart sapma üstü olarak ifade edilmiştir.



Şekil 5.2. Azaltılmış verinin MAPE farklarının dağılımı.

İlgili aralıkların sonuçları Şekil 5.3'te gösterilmiştir. Görseldeki her bir veri noktası bir aboneyi temsil etmektedir. Burada tüm abone verileri kullanılmıştır ve yeşil noktalar güvenli aralığı, mavi noktalar ara bölgeyi, kırmızı noktalar kayıp ve hesaplanamayan tüketime sahip aboneleri göstermektedir. Güvenli bölgede 165.180 abone bulunurken, geçiş bölgesinde 10.052 abone, kayıp ve hesaplanamayan tüketime sahip bölgede ise 3.595 abone bulunmaktadır.



Şekil 5.3. Abonelerin belirlenen bölgeler üzerinde dağılımı.

Çalışma kapsamına dahil edilen abonelerin büyük bir bölümünün yeşil (güvenli) bölgede yer aldığı görülmüştür (%92,36). Bu bölgede bulunan abonelerin önerilen iki modelden elde edilen doğal gaz tüketim tahmin değerleri birbirlerine oldukça yakındır. Bu durum ise, kullanılan sayaçların birbirlerine yakın davranışlar sergilediğini göstermektedir. Literatürde yer alan bazı çalışmalarda tek model birden fazla aboneye uygulanmakta ve tüketim tahmini başarısı düşük çıkan aboneler anormal tüketim davranışına sahip olarak nitelendirilebilmektedir. Fakat abonelerin tüketim davranışlarını yukarı veya aşağı yönde etkileyebilecek faktörler göz ardı edilmemelidir. Doğal gaz abonelerinin tüketim davranışları yıl içerisinde çeşitli sebeplerden dolayı değişkenlik gösterebilmektedir. Bu durumlara örnek olarak kışın gerçekleştirilen uzun veya kısa süreli seyahatler, hane halkı sayısında meydana gelen artış veya azalışlar, doğal gazı yakan cihazın değiştirilmesi veya ısınmak amacıyla farklı bir enerji kaynağına geçilmesi, apartman veya dairede bir tamirat tadilat işlemi

başlatılması olarak verilebilmektedir. Bahsi geçen durumlar gerçekleştiğinde ilgili abonenin normal tüketim davranışlarının değişmesi oldukça doğal bir durumdur. Bu çalışmada, kişisel tercihlerden veya zorunlu sebeplerden ötürü tüketim davranışlarında değişiklik meydana gelen abonelerin anormal bölgede yer almasını önlemek amacıyla yalnızca kullanılan doğal gaz ölçüm sayaçlarının etkisi ortaya çıkarılmaya çalışılmıştır.

Mavi bölge, modellerin MAPE farklarının verildiği dağılım grafiğinde normal sınırlar içerisinde yer almayan abonelerin başlangıç sınırındadır. Bunun sebebi modellerin MAPE farklarının bulunduğu verinin normal dağılım göstermesidir. Burada yer alan aboneler güvenli veya anormal olarak nitelendirilmemiş, geçiş bölgesi olarak ifade edilmiştir.

Kırmızı bölge, anormal tüketim davranışları sergileyen abonelerin bulunduğu bölgedir. Bu bölgede yer alan abonelerin ortak özellikleri, tüketim serilerinde mevcut olan sayaç parametresinin model çıktılarına farklı etkilerde bulunmasıdır. Bu durum ise modellerden elde edilen tüketim tahmini hata değerleri farklarının yüksek çıkmasına neden olmuştur. Elde edilen bulgulara göre, anormal bölgede bulunan aboneler tüm abonelerin %2,01'ini oluşturmaktadır. Bu bölgede bulunan aboneler arasında kayıp ve hesaplanamayan tüketime sahip abonelerin bulunma riski yüksektir.

Bu konuda gelecekte yapılacak olan benzer çalışmalar için model eğitiminde kullanılan bağımsız değişkenlere ilaveten ilgili doğal gaz abonelerinin coğrafi konumları gruplandırılarak eklenebilir. Bu bilgiden yola çıkarak tüm abonelerin gelir düzeyleri hakkında ortalama bir çıkarım gerçekleştirilebilir. Doğal gaz abonelerinin hanelerinde yaşayan kişi sayısı da model eğitiminde kullanılarak modellerin başarısı üzerine etkisi araştırılabilir. Ek olarak tüketimlerinde anormallik olduğu belirlenerek kayıp doğal gaz tüketimi riskinin yüksek olduğu düşünülen aboneler üzerinde öznitelik çıkarımı gerçekleştirilerek bu abonelerin ortak özellikleri belirlenebilir.

KAYNAKLAR

- [1] Ö. Evcim, “Derin öğrenme ve istatistiksel modelleme yöntemleri ile günlük doğalgaz tüketiminin tahmini ve karşılaştırılması”, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar ve Bilişim Mühendisliği Ana Bilim Dalı, Yüksek Lisans Tezi, 2019.
- [2] “T.C. Enerji Piyasası Düzenleme Kurumu, Doğal Gaz Piyasası Sektör Raporu Ocak 2022”, Ankara, 2022.
- [3] L. Botev ve P. Johnson, “Applications of statistical process control in the management of unaccounted for gas”, *J Nat Gas Sci Eng*, c. 76, sy February, s. 103194, 2020, doi: 10.1016/j.jngse.2020.103194.
- [4] K. Costello, “Lost and Unaccounted-for Gas: Practices of State Utility Commissions”, *Energy and Environment National Regulatory Research Institute*, sy 13, 2013.
- [5] A. Soltanisarvestani, A. A. Safavi, ve M. A. Rahimi, “The Detection of Unaccounted for Gas in Residential Natural Gas Customers Using Particle Swarm Optimization-based Neural Networks”, *Energy Sources, Part B: Economics, Planning, and Policy*, c. 18, sy 1, Ara. 2023, doi: 10.1080/15567249.2022.2154412.
- [6] A. Soltanisarvestani ve A. A. Safavi, “Modeling unaccounted-for gas among residential natural gas consumers using a comprehensive fuzzy cognitive map”, *Util Policy*, c. 72, s. 101251, Eki. 2021, doi: 10.1016/j.jup.2021.101251.
- [7] M. Shafiq, W. bin Nisar, M. M. Savino, Z. Rashid, ve Z. Ahmad, “Monitoring and controlling of unaccounted for gas (UFG) in distribution networks: A case study of Sui Northern Gas Pipelines Limited Pakistan”, 2018, c. 51, sy 11. doi: 10.1016/j.ifacol.2018.08.284.
- [8] F. Arpino, M. Dell’Isola, G. Ficco, ve P. Vigo, “Unaccounted for gas in natural gas transmission networks: Prediction model and analysis of the solutions”, *J Nat Gas Sci Eng*, c. 17, 2014, doi: 10.1016/j.jngse.2014.01.003.
- [9] M. Akpınar, “Evsel doğal gaz kullanımında konjonktür etkisinden arındırılmış tüketimin ARIMA ve çoklu regresyon yöntemleri ile tahmini”, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar ve Bilişim Mühendisliği Ana Bilim Dalı, Yüksek Lisans Tezi, 2014.
- [10] M. Akpınar, “İstatistik ve yapay zeka teknikleri ile enerji tüketiminin tahmini: Sakarya doğal gaz tüketiminin uygulaması”, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar ve Bilişim Mühendisliği Ana Bilim Dalı, Doktora Tezi, 2017.

- [11] N. Wei vd., “Data complexity of daily natural gas consumption: Measurement and impact on forecasting performance”, *Energy*, c. 238, 2022, doi: 10.1016/j.energy.2021.122090.
- [12] A. S. Anđelković ve D. Bajatović, “Integration of weather forecast and artificial intelligence for a short-term city-scale natural gas consumption prediction”, *J Clean Prod*, c. 266, 2020, doi: 10.1016/j.jclepro.2020.122096.
- [13] W. Qiao, Z. Yang, Z. Kang, ve Z. Pan, “Short-term natural gas consumption prediction based on Volterra adaptive filter and improved whale optimization algorithm”, *Eng Appl Artif Intell*, c. 87, 2020, doi: 10.1016/j.engappai.2019.103323.
- [14] H. Lu, X. Ma, ve M. Azimi, “US natural gas consumption prediction using an improved kernel-based nonlinear extension of the Arps decline model”, *Energy*, c. 194, 2020, doi: 10.1016/j.energy.2020.116905.
- [15] O. Laib, M. T. Khadir, ve L. Mihaylova, “Toward efficient energy systems based on natural gas consumption prediction with LSTM Recurrent Neural Networks”, *Energy*, c. 177, ss. 530-542, Haz. 2019, doi: 10.1016/j.energy.2019.04.075.
- [16] J. Song vd., “Estimate the daily consumption of natural gas in district heating system based on a hybrid seasonal decomposition and temporal convolutional network model”, *Appl Energy*, c. 309, s. 118444, Mar. 2022, doi: 10.1016/j.apenergy.2021.118444.
- [17] M. Brabec, O. Konár, E. Pelikán, ve M. Malý, “A nonlinear mixed effects model for the prediction of natural gas consumption by individual customers”, *Int J Forecast*, c. 24, sy 4, ss. 659-678, Eki. 2008, doi: 10.1016/j.ijforecast.2008.08.005.
- [18] M. E. Çelik, “Sinop ilinde doğal gaz tüketim profilinin araştırılması ve yapay sinir ağlarıyla modellenmesi”, Sinop Üniversitesi, Fen Bilimleri Enstitüsü, Disiplinlerarası Nükleer Enerji ve Enerji Sistemleri Ana Bilim Dalı, Yüksek Lisans Tezi, 2021.
- [19] S. Demirceylan, “Erzurum’da doğalgaz tüketim miktarının yapay sinir ağı algoritması kullanılarak tahmin edilmesi”, Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Ana Bilim Dalı, Yüksek Lisans Tezi, 2012.
- [20] E. M. Karaca, “Regresyon ve yapay sinir ağları yöntemi ile İstanbul ili doğalgaz tüketim tahmini”, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Ana Bilim Dalı, Yüksek Lisans Tezi, 2021.
- [21] Y. Yılmaz, “Daily natural gas consumption prediction by MARS and CMARS models for residential users in Ankara”, Orta Doğu Teknik Üniversitesi, Uygulamalı Matematik Enstitüsü, Bilimsel Hesaplama Ana Bilim Dalı, Yüksek Lisans Tezi, 2015.

- [22] B. Akın, “Yapay sinir ağlarıyla Konya bölgesinde kullanıcı doğal gaz tüketim öngörüsü”, İstanbul Teknik Üniversitesi, Enerji Enstitüsü, Enerji Bilim ve Teknoloji Ana Bilim Dalı, Yüksek Lisans Tezi, 2017.
- [23] B. Taşkın, “Ankara ili doğal gaz tüketiminin yapay sinir ağları ile öngörüsü”, İstanbul Teknik Üniversitesi, Enerji Enstitüsü, Enerji Bilim ve Teknoloji Ana Bilim Dalı, Enerji Bilim ve Teknoloji Bilim Dalı, Yüksek Lisans Tezi, 2018.
- [24] B. Kant, “Optimal doğal gaz tüketiminin tespitinde farklı yapay sinir ağı algoritmalarının kullanımı: Samsun ili örneği”, Ondokuz Mayıs Üniversitesi, Lisansüstü Eğitim Enstitüsü, Akıllı Sistemler Mühendisliği Ana Bilim Dalı, Yüksek Lisans Tezi, 2021.
- [25] T. Koturbash, “Determining the quality of natural gas and biomethane”, KTH Royal Institute of Technology, Energy Technology, Doctoral Thesis, 2021.
- [26] British Petroleum, “BP Statistical Review of World Energy 2022 71st edition”, 2022.
- [27] “T.C. Enerji Piyasası Düzenleme Kurumu, Doğal Gaz Piyasası Sektör Raporu Nisan 2022”, 2022.
- [28] İstanbul Büyükşehir Belediyesi, *UGETAM Doğal Gaz Sayaçları ve Çalışma Prensipleri*. İstanbul, 2020.
- [29] K. W. Costello, “Lost and unaccounted-for gas: Challenges for public utility regulators”, *Util Policy*, c. 29, 2014, doi: 10.1016/j.jup.2014.04.003.
- [30] A. M. Turing, “Computing machinery and intelligence”, *Mind*, c. LIX, sy 236, ss. 433-460, Eki. 1950, doi: 10.1093/mind/LIX.236.433.
- [31] V. Goncalves, K. Maria, ve A. B. F. da Silv, “Applications of artificial neural networks in chemical problems”, içinde *Artificial Neural Networks - Architectures and Applications*, 2013. doi: 10.5772/51275.
- [32] E. Ersoy ve Ö. Karal, “Yapay sinir ağları ve insan beyni”, *İnsan ve Toplum Bilimleri Araştırmaları Dergisi*, c. 1, sy 2, ss. 188-205, 2012.
- [33] A. C. Yazıcı, E. Ögüş, S. Ankaralı, S. Canan, H. Ankaralı, ve Z. Akkuş, “Yapay sinir ağlarına genel bakış”, *Türkiye Klinikleri Journal of Medical Sciences*, c. 27, ss. 65-71, 2007.
- [34] K. H. Durmuş ve C. Meriç, “Makina mühendisliğinde yapay sinir ağlarının (YSA) kullanımı”, *Mühendis ve Makina*, c. 46, ss. 47-56, 2005.
- [35] S. Haykin, “Learning processes; single-layer perceptrons; multilayer perceptrons. Neural Networks A Comprehensive Foundation.2nd ed.”, United States, 1999, ss. 14-68.
- [36] P. Hosseinzadeh Talaei, “Multilayer perceptron with different training algorithms for streamflow forecasting”, *Neural Comput Appl*, c. 24, sy 3-4, 2014, doi: 10.1007/s00521-012-1287-5.

- [37] W. Wang, P. H. A. J. M. V. Gelder, J. K. Vrijling, ve J. Ma, “Forecasting daily streamflow using hybrid ANN models”, *J Hydrol (Amst)*, c. 324, sy 1-4, 2006, doi: 10.1016/j.jhydrol.2005.09.032.
- [38] F. Bre, J. M. Gimenez, ve V. D. Fachinotti, “Prediction of wind pressure coefficients on building surfaces using artificial neural networks”, *Energy Build*, c. 158, 2018, doi: 10.1016/j.enbuild.2017.11.045.
- [39] P. Cunningham, M. Cord, ve S. J. Delany, “Supervised Learning”, içinde *Machine Learning Techniques for Multimedia*, Berlin, Heidelberg: Springer Berlin Heidelberg, ss. 21-49. doi: 10.1007/978-3-540-75171-7_2.
- [40] Z.-H. Zhou ve M. Li, “Semi-supervised learning by disagreement”, *Knowl Inf Syst*, c. 24, sy 3, ss. 415-439, Eyl. 2010, doi: 10.1007/s10115-009-0209-z.
- [41] E. Alpaydın, *Introduction to Machine Learning Second Edition Adaptive Computation and Machine Learning*. 2015.
- [42] S. Delepoulle, C. Renaud, ve P. Preux, “Photometric compression and interpolation for lightsource representation”, 2010.
- [43] K. K. Mak, K. Lee, ve C. Park, “Applications of machine learning in addiction studies: A systematic review”, *Psychiatry Research*, c. 275. 2019. doi: 10.1016/j.psychres.2019.03.001.
- [44] S. Kopácsi, R. Hudak, ve R. Ganguly, “Implementation of a classification server to support metadata organization for long term preservation systems”, *VOEB-Mitteilungen*, c. 70, sy 2, 2017, doi: 10.31263/voebm.v70i2.1897.
- [45] E. Alpaydın, “Yapay Öğrenme”, içinde *Boğaziçi Üniversitesi Yayinevi*, 2018.
- [46] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, ve K. J. Kim, “A survey of deep learning-based network anomaly detection”, *Cluster Comput*, c. 22, 2019, doi: 10.1007/s10586-017-1117-8.
- [47] A. Baldominos, E. Albacete, Y. Saez, ve P. Isasi, “A scalable machine learning online service for big data real-time analysis”, içinde *IEEE SSCI 2014 - 2014 IEEE Symposium Series on Computational Intelligence - CIBD 2014: 2014 IEEE Symposium on Computational Intelligence in Big Data, Proceedings*, 2015. doi: 10.1109/CIBD.2014.7011537.
- [48] A. E. Ezugwu, A. K. Shukla, M. B. Agbaje, O. N. Oyelade, A. José-García, ve J. O. Agushaka, “Automatic clustering algorithms: a systematic review and bibliometric analysis of relevant literature”, *Neural Computing and Applications*, c. 33, sy 11. 2021. doi: 10.1007/s00521-020-05395-4.
- [49] X. Tang vd., “Joint modeling of dense and incomplete trajectories for citywide traffic volume inference”, içinde *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, 2019. doi: 10.1145/3308558.3313621.

- [50] S. T. Pérez Suárez, C. M. Travieso González, ve J. B. Alonso Hernández, “Design methodology of an equalizer for unipolar non return to zero binary signals in the presence of additive white Gaussian Noise using a time delay neural network on a field programmable gate array”, *Sensors (Switzerland)*, c. 13, sy 12, 2013, doi: 10.3390/s131216829.
- [51] S. Anam, “Rainfall prediction using backpropagation algorithm optimized by Broyden-Fletcher-Goldfarb-Shanno algorithm”, içinde *IOP Conference Series: Materials Science and Engineering*, 2019, c. 567, sy 1. doi: 10.1088/1757-899X/567/1/012008.
- [52] M. Akpınar ve N. Yumusak, “Forecasting household natural gas consumption with ARIMA model: A case study of removing cycle”, içinde *AICT 2013 - 7th International Conference on Application of Information and Communication Technologies, Conference Proceedings*, 2013. doi: 10.1109/ICAICT.2013.6722753.
- [53] M. Akpınar ve N. Yumusak, “Estimating household natural gas consumption with multiple regression: Effect of cycle”, içinde *2013 International Conference on Electronics, Computer and Computation, ICECCO 2013*, 2013. doi: 10.1109/ICECCO.2013.6718260.



ÖZGEÇMİŞ

Ad-Soyad : Kürşad ARSLAN

ÖĞRENİM DURUMU:

- **Lisans** : 2020, İnönü Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği
- **Yükseklisans** : Devam ediyor, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar ve Bilişim Mühendisliği Anabilim Dalı

MESLEKİ DENEYİM VE ÖDÜLLER:

- Şubat 2022’de Bandırma Onyedi Eylül Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi Yazılım Mühendisliği Bölümü’nde araştırma görevlisi olarak çalışmaya başladı ve görevine devam etmektedir.
- 2019 yılında Kale Kalıp Makine ve Kalıp Sanayi A.Ş.’de stajyer bilgisayar mühendisi olarak çalıştı.