T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

ELEKTRİK GÜÇ DAĞITIMINDA AKILLI SAYAÇ VERİLERİ için ANOMALİ TESPİTİ ve TAHMİNLEME

Serhat YARAT

DANIŞMAN
Doç. Dr. Zeynep ORMAN

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

İSTANBUL, 2022

TEZ KABUL VE ONAYI

Serhat YARAT tarafından, **Doç. Dr. Zeynep ORMAN** danışmanlığında hazırlanan **"ELEKTRİK GÜÇ DAĞITIMINDA AKILLI SAYAÇ VERİLERİ için ANOMALİ TESPİTİ ve TAHMİNLEME"** başlıklı bu çalışma, jürimiz tarafından **27/12/2022** tarihinde yapılan sınav sonucunda **oy birliği** ile başarılı bulunarak **Yüksek Lisans Tezi** olarak kabul edilmiştir.

Tez Jürisi

		İmza	**Sonuç**
DANIŞMAN	Doç. Dr. Zeynep ORMAN İstanbul Üniversitesi-Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı		☒ Kabul ☒ Ret
ÜYE	Doç. Dr. Selçuk SEVGEN İstanbul Üniversitesi-Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı		☒ Kabul ☐ Ret
ÜYE	Dr. Öğr. Üyesi Şengül BAYRAK İstanbul Sabahattin Zaim Üniversitesi Yazılım Mühendisliği Anabilim Dalı		☒ Kabul ☐ Ret

İTHAF

Annem, Babam ve Kardeşlerime ithaf ediyorum…

BÜTÇE DESTEKLERİ

ELEKTRİK GÜÇ DAĞITIMINDA AKILLI SAYAÇ VERİLERİ için ANOMALİ TESPİTİ ve TAHMİNLEME

Bu tez çalışması için herhangi bir kurumdan bütçe desteği alınmamıştır.

TEŞEKKÜR

Tez çalışma süreci boyunca gösterdikleri yardım ve desteklerden dolayı danışmanım Doç. Dr. ZEYNEP ORMAN'a teşekkürlerimi sunarım. Ayrıca, hali hazırda çalıştığım GTech (G Teknoloji Bilişim San. Tic. A.Ş.) şirketindeki Büyük Veri ve Analitik ekibinin değerli desteklerine de teşekkür ederim. Beni her zaman seven ve destekleyen aileme de şükranlarımı sunarım.

Aralık 2022 **Serhat YARAT**

İÇİNDEKİLER

Sayfa No

1. GİRİŞ 1
1.1. ARAŞTIRMA SORUSU 6
1.2. SORUNUN KAPSAMI 6
1.3. TEZ TASLAĞI 7
2. KAVRAMSAL ÇERÇEVE 8
2.1. AKILLI ELEKTRİK SAYAÇLARI 8
2.2. ZAMAN SERİLERİ 9
2.3. ANOMALİLER VE ANOMALİ TESPİTİ 10
2.3.1. Anomali Tespitinde Başarı Ölçümü 15
2.4. ZAMAN SERİLERİNDE TAHMİN YAKLAŞIMLARI 17
2.4.1. Zaman Serileri Üzerinde Tahmin için Başarı Ölçümü 18
3. YÖNTEM 20
3.1. VERİ KÜMESİ 20
3.2. KULLANILAN ALGORİTMALAR 23
3.2.1. İzolasyon Ormanı (Isolation Forest - IF) 23
3.2.2. Yerel Aykırı Değer Faktörü (Local Outlier Factor - LOF) 23
3.2.3. Ekstrem Gradyan Arttırma (eXtreme Gradient Boosting – XGBoost) 25
3.2.4. FbProphet 25
3.3. ANOMALİ TESPİTİ VE TAHMİNLEME 25
3.3.1. Anomali Tespitinde Kullanılan Yapılar 26
3.3.2. Tahmilemede (forecasting) Kullanılan Yapılar 30
4. BULGULAR 32
4.1. ANOMALİ TESPİTİ İLE İLGİLİ BULGULAR 32
4.2. TAHMİNLEME İLE İLGİLİ BULGULAR 36
5. TARTIŞMA 40
6. SONUÇ VE ÖNERİLER 42
KAYNAKLAR 43
EKLER 49

Ek-1.Yöntemlerin Modellemesindeki Önemli Kod Blokları 49
Ek-1.1. IF 49
Ek-1.2. LOF 49
Ek-1.3. FbProphet 49
Ek-1.4. XGBoost 49
Ek-2. Veri Kümesi ile İlgili Metrikler 50

ŞEKİL LİSTESİ

Sayfa No

Şekil 2. 1: Akıllı Sayaçların Global Marketteki Değeri [25]. ---------- 9

Şekil 2. 2: Bağlamsal anomali için örnek Iğdır şehrinin 2022 sıcaklık değerlerinin gösterimi [44]. ---------- 11

Şekil 2. 3: Farklı anomali tespit yöntemleri ve veri kümesindeki etiketlerin kullanılabilirliği [42,43]. ---------- 13

Şekil 2. 4: Bazı anomali tespiti algoritmaları [42,45].---------- 14

Şekil 2. 5: Anomali algılama değerlendirmesi için performans metrik ağacı [48], [49].------ 16

Şekil 3. 1: Kullanılan veri kümesinin nokta dağılımı. 20

Şekil 3. 2: Kullanılan veri kümesinin grafiksel dağılımı. 21

Şekil 3. 3: Bir akıllı sayacın aylık tüketim dağılımı. 21

Şekil 3. 4: Bir akıllı sayaca ait aylık tüketim verisinin eğilim, mevsimsellik ve gürültü ayrışımını veren STL ayrışması dağılımı. 22

Şekil 3. 5: İzolasyon Ormanı [67]. 23

Şekil 3. 6: Tez çalışmasındaki anomali tespiti akışı. 29

Şekil 3. 7: Bu çalışmanın çıktısı üzerinde ileri ki zamanlarda yapılabilecek çalışma örneğinin akışı. 30

Şekil 3. 8: Bu çalışmada kullanılan veri üzerinde tahminleme akışı. 31

Şekil 4. 1: Çalışmanın akış şeması. 32

Şekil 4. 2: Bir sayaca ait veri dağılımı ve olası anomali noktaları. 33

Şekil 4. 3: Bir Akıllı Sayaç (Sayaç_13) için IF ile Aylık Tüketim Değeri Bazında Anomali Tespiti. 35

Şekil 4. 4: Bir Akıllı Sayaç (Sayaç_13) için LOF ile Aylık Tüketim Değeri Bazında Anomali Tespiti. 35

Şekil 4. 5: Bir Akıllı Sayaç (Sayaç_13) için FbProphet ile Aylık Tüketim Değeri Bazında Anomali Tespiti. 36

Şekil 4. 6: Kullanılan Akıllı Sayaçların FbProphet ile Ortalama Olarak Anomali Önemi. 36

Şekil 4. 7: Kullanılan Akıllı Sayaçlar Bazında FbProphet ile 12 Aylık Tüketim Değeri Tahmininde MAPE Değeri Karşılaştırılması. 37

Şekil 4. 8: Kullanılan Akıllı Sayaçlar Bazında XGBoost ile 12 Aylık Tüketim Değeri Tahmininde MAPE Değeri Karşılaştırılması. 38

Şekil 4. 9: Bir Akıllı Sayaç için FbProphet ile Aylık Tüketim Tahmini.....38

Şekil 4. 10: Bir Akıllı Sayaç için XGBoost ile Aylık Tüketim Tahmini. 39

Ek-2. 1: Veri kümesinin sweetviz_report ile gösterimi. 50

TABLO LİSTESİ

Sayfa No

Tablo 3. 1: Bir Akıllı sayaca ait aylık tüketim veri yapısı.....22

Tablo 3. 2: neighbors.LocalOutlierFactor davranışı [69].....24

Tablo 4. 1: Anomali için Kullanılan Algoritmaların Hiper Parametre Listesi. 33

Tablo 4. 2: Kullanılan yaklaşımlarının anomali çıktıları.....34

Tablo 4. 3: FbProphet ve XGBoost'nin MAPE değeri ile karşılaştırması. 37

SİMGE VE KISALTMA LİSTESİ

Simgeler **Açıklama**

ds : Tarih

y : Gerçek tüketilen Değeri

y_hat : Tahmin Edilen Tüketilen Değeri

yhat_upper : Tahmin Edilen Üst limit Tüketilen değeri

yhat_lower : Tahmin Edilen Alt Limit Tüketilen Değeri

Kısaltmalar **Açıklama**

ABADT **:** Açı Bazlı Aykırı Değer Tespiti (Angle Based Outlier Detection)

AUC **:** ROC eğrisinin altında kalan alan (Area Under Curve)

ARIMA **:** Otoregresif Hareketli Ortalamalar(Autoregressive Integrated Moving Average)

AR **:** Otomatik Regresyon (Autoregressive)

ARMA **:** Otoregresif Hareketli Ortalama (Autoregressive Moving Average Model)

BBADF **:** Bağlantı Bazlı Aykırı Değer Faktörü(Connectivity Based Outlier Factor)

CNN-RNN -BiLSTM **:** Çift Yönlü Uzun Kısa Vadeli Bellek tabanlı Tekrarlayan Sinir Ağı ile birleştirilmiş Evrişimli Sinir Ağı (Convolutional Neural Network combined with Bidirectional Long Short Term Memory based Recurrent Neural Network)

CNN **:** Evrişimsel Sinir Ağları (Convolutional Neural Networks)

CAGR **:** Yıllık bileşik büyüme oranı (Compound Annual Growth Rate)

DNKADT **:** Derece Numarası Kullanarak Aykırı Değer Tespiti (Outlier Detection Using Indegre Number)

DN **:** Doğru Negatif (True Negative)

DOK **:** Derin Otomatik Kodlayıcılar(Deep Auto-Encoders)

DP **:** Doğru Pozitif (True Positive)

EPDK **:** Enerji Piyasaları Dengeleme Kurumu

FFNN **:** İleri Beslemeli Sinir Ağı (Feed-forward Neural Network)

FbProphet **:** Facebook Prophet Algrotiması

GRU **:** Kapı Özyinelemeli Geçitler (Gated Recurrent Units)

GTBA : Güçlü Temel Bileşen Analizi (Robust Principal Component Analysis)

HABADT : Hızlı Açı Bazlı Aykırı Değer Tespiti (Fast Angle Based Outlier Detection)

HBADP : Histogram Bazlı Aykırı Değer Puanı (Histogram Based Outlier Score)

IOT : Nesnelerin interneti (Internet of things)

KNN : K-En Yakın Komşu **(**K-Nearest Neighbour)

LSTM : Uzun Kısa Süreli Bellek (Long Short-Term Memory)

LOF : Yerel Aykırılık Faktörü (Local Outlier Factor)

MAPE : Ortalama Mutlak Yüzde Hatası (Mean Absolute Percentage Error)

MAE : Ortalama Mutlak Hatası (Mean Average Error)

MSE : Ortalama Kare Hatası (Mean Squared Error)

ML : Makine Öğrenmesi (Machine Learning)

MLP : Çok Katmanlı Yapay Sinir Ağları (Multi Layer Perceptron)

MA : Hareketli Ortalama (Moving Average)

RMSE : Ortalama Hata Kareleri Toplamı Kökü (Root Mean Square Error)

RNN : Özyinelemeli Sinir Ağı (Recurrent Neural Network)

ROC : İşlem Karakteristik (Receiver Operating Characteristic) Eğrisi

SVM : Destek Vektör Makineleri (Support Vector Machine)

SARIMA : Mevsimsel Otoregresif Entegre Hareketli Ortalama (Seasonal ARIMA)

STL : LOESS kullanarak Mevsimsel-Trend ayrıştırması (Seasonal-Trend decomposition using LOESS)

STLF : Kısa Vadeli Yük Tahmini (Short Term Load Forecasting)

T-S-DVM : Tek-Sınıf Destek Vektör Makineleri (One-Class SVM)

vd. : ve diğerleri

YN : Yanlış Negatif (False negative)

YP : Yanlış Pozitif (False Positive)

YYAF : Yerel Yoğunluk Aykırı Faktör (Local Density Outlier Factor)

XGBoost : Ekstrem Gradyan Arttırma (Extreme Gradient Boosting)

ÖZET

YÜKSEK LİSANS TEZİ

ELEKTRİK GÜÇ DAĞITIMINDA AKILLI SAYAÇ VERİLERİ için ANOMALİ TESPİTİ ve TAHMİNLEME

Serhat YARAT

İstanbul Üniversitesi-Cerrahpaşa

Lisansüstü Eğitim Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı

Danışman : Doç. Dr. Zeynep ORMAN

Nüfus yoğunluğu ve ekonomik büyümenin etkisiyle enerji talebi hızla artmaktadır. Bu talep karşısında enerji ve elektrik şebekeleri daha fazla zorlukla karşı karşıya gelmektedir. Enerji tüketiminin sıkı bir şekilde izlenmesi ve kontrol altında tutulması önem arz etmektedir. Enerji dağılımını düşündüğümüzde akıllı sayaçlar bu enerjinin kontrolünde kilometre taşı rolü oynamaktadır. Enerji tüketim ölçümlerinin yapıldığı sayaçlarda meydana gelebilecek herhangi bir elektrik kesintisi, bir hata veya yanlış ölçüm, dağıtım şirketlerinden son kullanıcılara kadar birçok tarafı etkilemektedir. Enerji sektöründeki bu tür anomalilerin tespiti için veri analitiği çalışmaları ve büyük veri teknolojileri, sensörlerden ve sayaçlardan toplanan zaman serisi verilerini gerçek zamanlı veya toplu olarak analiz ederek verimliliği ve tasarrufu arttırmayı amaçlayan net ve eyleme geçirilebilir çıktılar üretmede önemli rol oynamaktadır.

Bu çalışmada, akıllı sayaçlarla ölçülen aylık tüketim değerlerine dayalı olarak enerji tüketimindeki olası anomalilerin tespit edilmesi ve farklı yöntemler kullanılarak gelecek tüketiminin tahmin edilmesi amaçlanmaktadır. Bu tez çalışması, akıllı sayaç verileri üzerinde enerji sektöründe kullanılabilecek bir takım makine öğrenmesi yöntemlerinin bir analizini içerdiğinden, dağıtım şirketlerinin olası finansal kayıplarını en aza indirebilmelerine, enerji kayıplarını önleyebilmelerine ve diğer bazı olumsuz etkileri azaltabilmelerine yardımcı olacaktır.

Sonuç olarak; enerji sektöründe genel aydınlatma sayaçları üzerinde yapılan uygulamalarda İzolasyon Ormanı (Isolation Forest-IF), Yerel Aykırı Değer Faktörü (Local Outlier Factor-LOF) ve FbProphet algoritmalarının anomali tespitinde olası uç anomali noktalarını başarılı bir şekilde tespit edebildiği ve FbProphet algoritmasının XGBoost algoritmasına göre sayaç verileri üzerinde zaman serileri ile yapılan tahminlemelerde ortalama olarak daha iyi sonuç verdiği tespit edilmiştir.

Aralık 2022 , 69 sayfa.

Anahtar kelimeler: Tahminleme, Anomali Tespiti, Zaman Serileri, Elektrik Güç Dağıtımı, Akıllı Sayaç

ABSTRACT

M.Sc. THESIS

ANOMALY DETECTION and PREDICTION FOR SMART METER DATA in ELECTRICAL POWER DISTRIBUTION

Serhat YARAT

İstanbul University-Cerrahpaşa

Institute of Graduate Studies

Department of Computer Engineering
Computer Engineering Program

Supervisor : Assoc. Prof. Dr. Zeynep ORMAN

With rapidly increasing energy demand as a result of increasing population density and economic growth, energy and electricity grids are now facing more challenges. This means that energy consumption should be kept under strict monitoring and control. When we consider energy distribution, smart meters play a milestone role in the control of this energy. Any power outage and an error or an incorrect measurement in the meters in which energy consumption measurements are made affect many stakeholders, from the distribution company to the end user. These unexpected behaviors can occur due to outages, faulty network infrastructures, unusual consumption situations of users, cyber-attacks, energy fraud, etc. Identifying such anomalies in the energy sector, data analytics studies, and big data technologies play an important role to produce clear and actionable outputs that aim to increase efficiency and savings by analyzing time series data collected from sensors and meters in real-time or batch.

In this study, it is aimed to detect possible anomalies in energy consumption and forecasting consumption using different machine learning methods, based on monthly consumption values measured by smart meters. Since this thesis contains an analysis of the relevant methods in the energy sector on smart meter data, it will help distribution companies to minimize possible financial losses, prevent energy losses, and reduce other negative effects.

As a result, it was found in this study's experiments and observations on general lighting meters in the energy sector that the Isolation Forest (IF), the Local Outlier Factor (LOF), and FbProphet were successful in detecting the potential extreme anomaly points. Additionally, it has been found that when using the FbProphet and XGBoost algorithms to forecast data from this time series, FbProphet outperforms XGBoost.

December 2022, 69 pages.

Keywords: Anomaly Detection, Forecasting, Electrical Power Distribution, Time Series, Smart Meters

1. GİRİŞ

Elektrik enerjisi günlük yaşamımızda aydınlatmadan barınakların ısınma ve soğutmasına kadar birçok alanda önemli bir rol almaktadır. Enerjiye bağımlılığın artmasıyla birlikte zaman içerisinde elektrik üretim ve dağıtım sistemleri yüksek teknolojili bir endüstri haline gelmiştir. Diğer endüstrilerden farklı olarak, elektrik endüstrisinin temel farklılaştırıcısı ürünün kendisidir. Elektrik enerjisinin depolanması yüksek maliyetli ve çok uzun süreli değildir. Dolayısıyla, üretim ve tüketim neredeyse gerçek zamanlı olarak gerçekleşir. Sektörün bu özelliği, bu sektördeki veri bilimi tabanlı yenilikler için temel bir etki yaratmaktadır. Dünya genelinde elektrik piyasasında arz ve talep arasında her zaman bir boşluk oluşabilmektedir. Piyasaya gerekli arzı sağlayarak boşluğu doldurmak ve hizmet verimliliğini artırmak için enerji alanında faaliyet gösteren elektrik şirketleri, tahmin teknikleri kullanarak gelecekteki talebi tahmin etme, talebi karşılama ve elektrik tüketimini optimize etme üzerine çalışmalar yapmaktadır. Bu alanda ele alınan bu gibi problemler için enerji yönetim sistemleri (EMS) yapıları şeklinde çözümler geliştirilebilir [1].

Elektrik enerjisi alanında veri bilimindeki modeller ile elektrik tüketimindeki bireysel kullanıcılar ve diğer kullanıcı grupları için tek modelleme çözümü oluşturulması oldukça güçtür. Modeller, elektriğin tüketim kalıplarına, enerji kayıplarına neden olan dış etkenlere, tüketici gruplarına ve endüstri tarzı gibi tüketim tiplerine bağlıdır. Veri analitiğinin bu alanda getirdiği çözümlerde akıllı sayaçların kullanımı önemli bir avantaj sağlamaktadır. Genel olarak, elektrik tüketimine yönelik yapılan tahmin çalışmalarında veri ile ilgili çalışmaya konu edilen bölgenin sosyo-ekonomik ve iklim koşulları gibi parametrelerin de göz önünde bulundurulması gerekmektedir. Uygulanan yönteme bağlı olarak verilerde çeşitli faktörlerin herhangi birinden veya birkaçından etkilenen birçok model oluşabilmektedir. Örneğin, normal bir evin elektrik tüketiminin en fazla olduğu zamanlar; bayramlar ve yeni yıl kutlamaları gibi tatiller veya özel günlerle ilgilidir. Bu dönemlerde ev kullanıcıları yıl boyunca kullanılan tüketimden daha fazla enerji tüketme eğilimindedir; tatil günlerinden önce bu günler için tüketim şekli, günlerin akşam ve gece gibi kısımlarına göre değişir. Bir binada yaşayanların profesyonel profiline dayalı verilerde bazı örüntüler görülebilir. Tipik olarak, öğrencilerin ve gençlerin kaldığı bir yurt, gece boyunca nispeten yüksek elektrik tüketimine sahip olacaktır; aynı zamanda, gündüz çalışanların

yaşadığı konutlar akşam ve sabah erken saatlerde yüksek enerji tüketimine sahiptir. İnsan yaşamının bu davranışsal, sosyal ve ekonomik yönleri tüketim kalıplarını oluşturur, dolayısıyla bu özellikleri doğrudan veya dolaylı olarak tespit etmek, veri bilimcisinin elektrik endüstrisinde hem tüketim hem de yük tahmini senaryoları için etkili bir model oluşturmasına yardımcı olur [2].

Akıllı sayaçlar ile kullanıcı hareketlerini ve sayaç durumlarını gerçek zamanlı olarak izlemek ve daha verimli enerji yönetimi için kararlar almak mümkündür. Akıllı ölçüm sistemi, şebekeyi besleyen elektriği veya şebekeden tüketilen elektriği ölçebilen elektronik bir sistemdir. Bu cihazlar, modern elektrik şebekelerinde geleneksel sayaçlardan daha fazla bilgi sağlar. Ancak akıllı cihazların, sistemi arızalara ve siber saldırılara karşı daha duyarlı olması gibi bazı dezavantajları da vardır [3]. Bu dezavantajları tehlikeli ve maliyetli olan enerji hırsızlığına ve şebeke dolandırıcılığına yol açabilir. Bunun yanı sıra, kesintiler, hatalı ağ altyapıları, kullanıcıların olağan dışı tüketim durumları gibi daha çok dikkate alınması gereken durumlar da bulunmaktadır.

Genel olarak, akıllı sayaçlar, enerji sektöründe elektrik tüketimi veya farklı bir alanda saatlik, günlük veya aylık büyük miktarda veri toplar. Elektrik alanında elektrik şirketleri, enerji dağıtımında birçok anomali ile karşı karşıya kalmakta ve bu anomaliler akıllı sayaçlar ile tespit edilebilmekte ve bu sayaçların verileri dikkate alınarak geleceğe yönelik tahminler yapılabilmektedir. Anomali tespiti, çeşitli araştırma ve uygulama alanlarında iyi çalışılmış önemli bir problemdir. Nesnelerin interneti (Internet of things - IoT) dünyasında bu kavramın önem kazanmasıyla birlikte birçok anomali tespit yönteminin şu anda bu sistemlerle aktif hale getirilmesi ve ardından anomalileri tespit edip bulunan çıktıları yorumlaması için önemli çalışmalar gerekmektedir. Cook vd., IoT verileri üzerinde anomali tespit tekniklerini uygularken karşılaşılabilecek zorluklar hakkında bir literatür incelemesi yapmışlardır [4].

Elektrik sektöründe, elektrik hırsızlığı, dolandırıcılık, kesinti vb. çeşitli alt problemler mevcuttur. Elektrik Hırsızlığı, elektrik tedarikçilerine zarar veren ve ekonomik kayıplara neden olan elektrik hizmetlerinde önemli bir sorundur. Elektrik hırsızlığını tespit etmek ve kontrol etmek için önemli çalışmalar vardır. Chandel ve Thakur, daha önce elektrik hırsızlığını tespit etmek için uygulanan yöntemlerden Destek Vektör Makinesi (Support Vector Machine - SVM) ve çok sınıflı (multiclass) Destek Vektör Makinesi yöntemlerini uygulayarak yeni önerdikleri yöntem ile karşılaştırmışlardır [5]. Bu çalışmadaki karşılaştırmalı sonuçlar doğrultusunda önerilen Çift Yönlü Uzun Kısa Vadeli Bellek tabanlı Tekrarlayan Sinir Ağı ile birleştirilmiş

Evrişimli Sinir Ağı (Convolutional Neural Network combined with Bidirectional Long Short Term Memory based Recurrent Neural Network - CNN-RNN-BiLSTM) modelinin elektrik hırsızlığı algılama için verimli çalıştığını göstermişlerdir. Bu tür hibrit yaklaşımlar bu alanda farklı problemlere de çözüm getirmektedir. Xu vd., zaman serisi anomali tespiti için yeni bir hibrit algoritma olan LSTM-GAN-XGBOOST yapısını önermişlerdir. Çalışmalarında, zaman serisi verilerinin zaman boyutlu özelliklerini çıkarmak için uzun kısa süreli bellek (LSTM), normal verilerin derin özelliklerini etkili bir şekilde çıkarmak için üretken rakip ağları (GAN) ve anomali puanlarını dışa aktarmak ve ayıklanan özellikleri sınıflandırmak için Ekstrem Gradyan Arttırma (Extreme Gradient Boosting - XGBOOST) algoritmalarını kullanmışlardır [6].

Chiosa vd., akıllı ölçüm altyapılarının yaygınlaşmasından yararlanan binalar ile ilgili veriler üzerinde verimsizlikleri veya anormal enerji davranışlarını tespit ederek, binalarda enerji yönetim sürecini geliştirmek için sayaç düzeyindeki verilerden faydalı bilgiler çıkarabilen daha sağlam analiz çerçevelerinin geliştirilmesine odaklanmışlardır [7]. Çalışmalarında, tüm bina sayacı düzeyinde anormal enerji tüketimini otomatik olarak tespit etmek ve ardından anormal modellerden sorumlu alt yükler üzerinde bir teşhis gerçekleştirmek için yenilikçi bir anomali tespiti ve teşhisi (anomaly detection and diagnosis - ADD) metodolojisi önermişlerdir. Binalardaki kullanılan akıllı sayaç verilerinden yararlanarak yapılan bir diğer çalışmada [8] ise problem bazlı olarak anomali tespit yaklaşımları uygulanmıştır.

Vafeiadis vd., çalışmalarında büyük bir Sloven elektrik dağıtım şirketi şebekesinin akıllı sayaçlarındaki anormal olayların tespiti ile ilgili tüm akıllı sensörlerden gelen veriler üzerinde anomali tespiti üzerine bir çalışma yapmışlardır [9]. Önerdikleri sistem, akıllı sayaçlardan gelen verileri analiz ederek; istatistiksel, makine ve derin öğrenme tekniklerini ve özellikle Otomatik Kodlayıcıları (Autoencoders) kullanarak, dağıtım ağındaki her akıllı sayaçtaki anomalileri tespit etmektedir.

Khan, yük tahmini ve talep tarafı yönetiminin (demand side management -DSM) verimli kullanımı için akıllı sayaç veri boyutu ve karmaşıklık gibi problemleri ele almıştır [10]. Daha doğru ve ayrıntılı yük profili oluşturmayı, azaltılmış profil karmaşıklığını ve gelişmiş yük tahmini doğruluğu ve DSM çözümlerini amaçlayan akıllı sayaç verilerinin analizi için yeni bir kümeleme tabanlı yaklaşım sumaktadır.

Shaukat vd., anomali türleri, anomalilerin veri türleri, zaman serisinin veri türleri, zaman serisi verilerinin bileşenleri, anomali bağlamının sınıflandırılması ve zaman serisi anomalilerin tespiti için kullanılan sınıflandırma yöntemleri ile ilgili bir literatür özeti çalışması yapmışlardır [11].

Al-Ghaili vd., gelişmiş ölçüm altyapısı ve akıllı şebeke sistemlerinde veri manipülasyonunu tespit etmek için kullanılan anomali teknikleri üzerine mevcut çalışmaları gözden geçirerek bir araştırma sunmuşlardır[12]. Ayrıca, mevcut çalışmalar tarafından kullanılan çeşitli ölçüm yöntemleri ve yaklaşımları ayrıntılı inceleyerek çalışmaları değerlendirmişlerdir.

Akıllı sayaç verilerinde anomaliler tespit edilerek gerçek zamanlı olarak takip edilebilmektedir. Shah, elektrik talep verilerindeki anomali problemi için LSTM ve İleri Beslemeli Sinir Ağı (Feedforward Neural Network - FFNN) tabanlı yöntemleri elektrik talebi verisi üzerinde uygulayıp karşılaştırmıştır [13]. Verilerde düzenli bir model mevcut olduğunda LSTM modelinin etkili olduğunu belirtilerek, yapılan çalışmanın sonuçlarının LSTM'nin FFNN'ye üstünlüğünü iddia etmek için istatistiksel olarak anlamlı olmadığı şeklinde değerlendirmişlerdir.

Steinborn, araştırmasında elektrik tüketiminde çok fazla eğitim verisi olmadığında Kısa Vadeli Yük Tahmini ayarlarında yüksek tahmin doğruluğunun nasıl elde edileceği sorusuna cevap vermeye yönelik çalışma yürütmüştür [14]. Bu çalışmada sunulan deneylerle, 30 dakikalık aralıklarla ölçülen 350 hanenin toplam elektrik talebinin bir zaman serisi olarak kullanılması ile farklı model türleri için model performansının nasıl etkilendiği analiz edilmiştir. Kardi vd., çalışmalarında zamansal ve gecikmeli özelliklere ek olarak çeşitli hava özelliklerini de dikkate alarak elektrik tüketim verileri üzerinde LSTM ve Otomatik Kodlayıcı (Autoencoder) yöntemleri ile yerel (local) ve küresel (global) anomaliler üzerinde çalışmışlardır [15].

Enerji alanında sınıflandırma, anomali tespiti ve tahmini için Türkiye'de yapılan bazı çalışmalar da bulunmaktadır. İşyapar, çalışmalarında Türkiye'de faaliyet gösteren bir enerji dağıtım şirketinin abonelerine ait verilerin makine öğrenmesi teknikleri kullanılarak sınıflandırılmasını ele almıştır [16]. Tajraq, Özyinelemeli Sinir Ağı (Recurrent Neural Network - RNN), LSTM ve Kapı Özyinelemeli Geçitler (Gated Recurrent Units - GRU) yöntemlerini kullanarak Türkiye'deki enerji tüketim verileri üzerinde tahminler yaparak bu modellerin

performanslarını karşılaştırmıştır [17]. Bu çalışmada, elektrik tüketim verileri sıcaklık ve tatiller gibi faktörler ile zenginleştirilmiştir. Kısa vadeli tüketim tahmini için daha iyi bir model olarak belirlenen GRU modeli seçilmiştir.

Dağ ve Alamin çalışmalarında, elektrik tüketimini hızlı ve güvenirliği yüksek şekilde tahmin etmek amacıyla, bellek üzerinde çalışan veri tabanları ile makine öğrenmesi konusunu ele almışlardır [18]. Çalışmada çeşitli makine öğrenimi algoritmalarını kullanarak Londra hanelerinde SmartMeter enerji tüketimi verileri üzerinde elektrik tüketim tahmini yapmışlardır. Bu çalışmada R-karenin değerlendirme ölçütüne göre en iyi algoritmanın Torbalama Algoritması (Bagging Algorithm) olduğunu belirtmişlerdir.

Çoğu ülkede olduğu gibi Türkiye'de de enerji denetimi yapan kurumlar mevcuttur. Türkiye'de EPDK (Enerji Piyasaları Dengeleme Kurumu) tarafından belirlenen kurallar doğrultusunda, üreticilerin günlük üretim tahminlerini raporlamaları gerekmektedir, bu kurallar rüzgâr enerjisi ve diğer enerji alanları içinde geçerlidir ve uygun çözümler sunulmalıdır. Aksoy çalışmasında, rüzgâr ile ilgili 8 türbin verisi kullanarak derin öğrenme yöntemleri ve makine öğrenme yaklaşımları uygulayarak rüzgâr gücü üretim tahmini yapmıştır [19]. Çalışmasında istatistiksel metotlar ve hibrit metotlar gibi yöntemlere odaklanmıştır.

Mwansa, evlerdeki davranış eğilimlerini öğrenmek ve anlamak için gelişmiş makine öğrenimi algoritmalarını kullanarak akıllı sayaç okumalarından ev davranış kalıplarını modelleme ve tahmin etme çalışması yapmıştır [20]. Bu araştırmada kullanılan veriler, 5 konuttan oluşan bir akıllı ev ortamının dijital simülasyon modelinden oluşturulmuştur. Çalışma sonunda oluşturulan veri kümesiyle en yüksek sonuç, %87,6 Eğri Altında Kalan Alan (Area Under the Curve – AUC) değerine ulaşan İki Sınıflı Karar Ormanı Sınıflandırıcısı (Two-Class Decision Forest Classifier) kullanılarak elde edilmiştir.

De Mattos Neto vd., enerji tüketimi zaman serileri verileri üzerinde yaptıkları çalışmada enerji tüketimi tahmini için Nöral Tabanlı Topluluklar (Neural-Based Ensembles) önermiş ve temel olarak Mevsimsel Otoregresif Entegre Hareketli Ortalama (Seasonal ARIMA-SARIMA) modelini ele almışlardır [21].

Resulaj vd., Norveç'teki elli beş hanede kısa vadeli yük tahmini için makine öğrenimi algoritmalarını kullanmışlardır [22]. Bir yıldan fazla elektrik tüketimi verilerini içeren bir veri kümesini hava durumu verileri, atmosferik veriler ve takvim değişkenleri gibi verilerle zenginleştirerek, İleriye Yönelik Sinir Ağı modelini ve Rastgele Orman (Random Forest)

modelini eğitmek ve değerlendirmek için kullanmışlardır. Değerlendirme sonucunda genel olarak, uygulanan Sinir Ağı modelinin, Rastgele Ormandaki karşılığına göre daha yüksek doğruluk elde ettiği belirtilmiştir.

1.1. ARAŞTIRMA SORUSU

Anomali tespit yöntemleri ve yük tahminleme teknikleri, veri türüne ve probleme özeldir. Örneğin, görüntülerdeki anomalileri tespit etmek için kullanılan algoritmalar, veri akışlarında kullanılan yaklaşımlardan farklıdır. Bu araştırma, akıllı sayaç zaman serisi verilerinde anomali tespiti ve tahminleme için popülerliği artan yöntemlere odaklanarak enerji tüketiminde önlemler alabilmemize olanak sağlayacak çıktılar üretmektedir. Veri üzerinde hem anomali tespiti hem de tahminleme için çözümler sağlayan yöntemler olduğu gibi sadece birine çözüm getiren yöntemlerde mevcuttur.

Bu tezde incelenen temel araştırma sorusu şu şekildedir:

- Elektrik sayaçlarının aylık tüketim değeri içeren zaman serisi verileri üzerinde olası anomali tespiti ve tüketim tahmini için amaca uygun İzolasyon Ormanı (Isolation Forest -IF), Aykırı Değer Faktörü (Local Outlier Factor-LOF), FbProphet ve XGBoost yöntemlerin uygulanmasında elde edilen sonuçların karşılaştırılması ve hangi yöntemin daha başarılı olduğu gibi sorulara cevap arayarak probleme çözüm getirmeyi amaçlamaktadır.

1.2. SORUNUN KAPSAMI

Bu çalışmada, genel aydınlatma sayaçlarıyla ölçülen aylık tüketim değerleri üzerindeki olası anomalilerin tespit edilmesi ve tüketiminin tahmin edilmesi üzerinde durulacaktır. Buna göre;

- Zaman serisi verisi üzerinde tespit edilen anomaliler ile oluşan veri kümesi üzerinde kullanılabilecek yeni bir yönteme yer verilmemiştir. Mevcut yöntemler kullanılarak tahminleme çalışmaları yürütülmüştür. İleri çalışmalarda bu konu üzerinde detaylı çalışma yapılması planlanmaktadır. Bu alanda az sayıda çalışma bulunmaktadır ve örnek olarak Fadwa Maatug çalışmasına başvurabilir [23]. Maatug, anomali etiketleriyle veri kümesini oluşturduktan sonra, gelecekteki güç tüketimlerini anormal veya normal olarak sınıflandırmak için yeni bir yöntem önermiş ve sonuçlarını farklı

sınıflandırma algoritmaları ile karşılaştırmıştır. Bu çalışmada veri üzerinde FbProphet ile tahminleme ve anomali tespiti, XGBoost ile tahminleme, scikit-learn kütüphanesinde bulunan LOF ve IF algoritmaları ile anomali tespiti yeni bir yaklaşım olarak ele alınmış ve değerlendirilmiştir.

- LOF ile yapılan yenilik tespiti (novelty detection) bu çalışmada kapsam dışı bırakılmıştır.
- Bu çalışmada genel aydınlatma sayaçlarındaki aylık bazlı tüketim değerleri üzerindeki olası anomalilerin tespit edilmesi problemi, zaman serilerinde anomali tespiti olarak ele alınmıştır.
- Elde edilen veri kümeleri sadece zamana bağlı sayaç bazlı aylık tüketimi içermektedir. Sıcaklık ve nem gibi ilişkili özellikler mevcut olmadığı için bu tezdeki uygulamalar sadece aylık tüketim değeri ile sınırlandırılmıştır.
- Bu çalışmada analizler için Python programla dili kullanılmıştır.

1.3. TEZ TASLAĞI

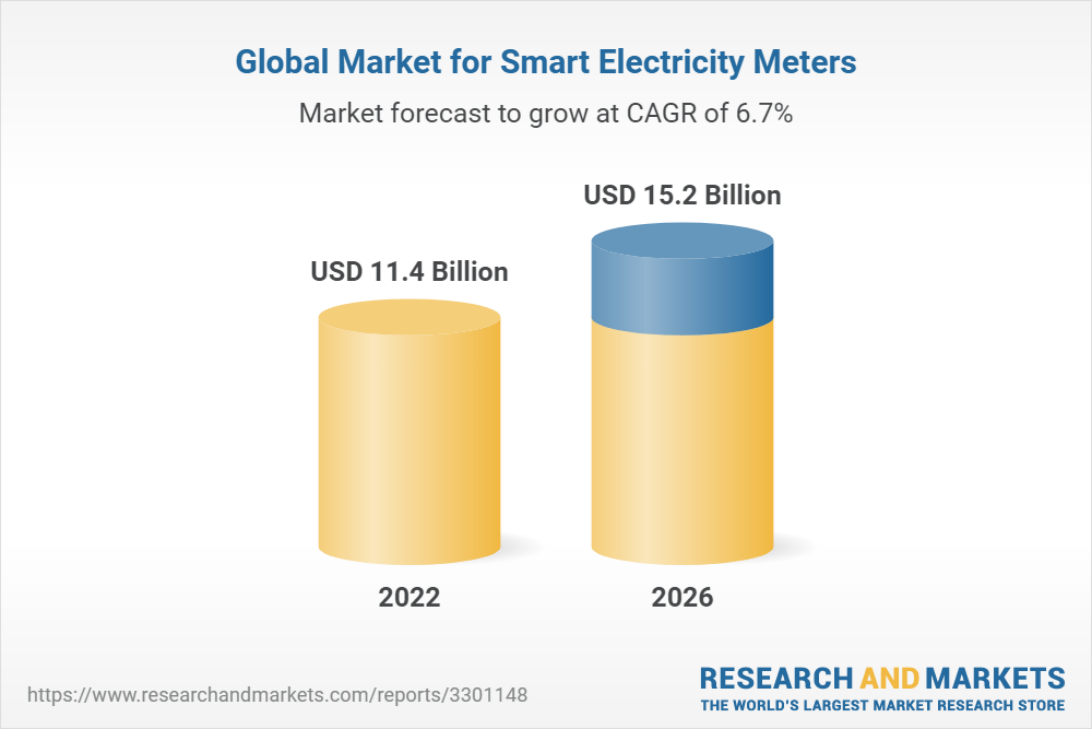
Bu tez, ampirik bir tez taslağını izleyen bölümlerde yapılandırılmıştır. Giriş bölümünde problemin genel amacı literatürdeki yeri üzerinde durulmuştur. Kavramsal Çerçeve bölümünde, sayaçların önemi, yeri, sayaç verilerine ait zaman serilerinde tahminleme ve anomali tespiti gibi konseptlerin literatürdeki yeri ve temel yapıları tanıtılmıştır. Yöntem bölümünde bu çalışmada kullanılan yöntemler tanıtılarak veri kümesine açıklık getirilmiştir. Bulgular bölümde çalışmada önerilen yöntemlerin uygulanmasından elde edilen sonuçlara yer verilmiştir. Tartışma, Sonuç ve Öneriler bölümlerinde çalışmanın değerlendirme sonuçlarını ve gerçekleştirilen deneylerin nihai sonucu tartışılmıştır.

2. KAVRAMSAL ÇERÇEVE

Bu bölüm, akıllı elektrik sayaçlarının marketteki yeri, zaman serilerinde veriler üzerinde anomali tespiti ve yük tahminleme tekniklerini kapsayacak şekilde bir inceleme sunar.

2.1. AKILLI ELEKTRİK SAYAÇLARI

Akıllı elektrik sayaçları, elektrik enerjisi ölçüm uygulamalarında kullanılmak üzere özel olarak tasarlanmış elektrik tüketimini ölçen elektronik ölçüm cihazlarıdır. Akıllı bir elektrik sayacı, bir kamu hizmeti müşterisinin enerji tüketim modellerini otomatik olarak yakalar, doğru ve güvenilir faturalandırma için bilgileri iletirken manuel sayaç okuma ihtiyacını önemli ölçüde azaltır. Akıllı elektrik sayaçlarının kullanımı, bu pazarlardaki müşteriler için ayrıntılı düzeyde faturalama verilerine ve doğru oranlara olan sürekli ihtiyaç nedeniyle zamanla kamu hizmetinin birkaç büyük müşterisinden konut, ticari ve endüstriyel müşteriler de dahil olmak üzere tüm müşteri sınıflarını kapsayacak şekilde genişlemiştir. Akıllı elektrik sayaçları, yenilikçi özelliklerinin yanı sıra, elektrik kesintilerini belirleme ve müdahale etme, enerji hırsızlığını önleme, yenilikçi hizmet modellerini kullanıma sunma, yeni ve yenilikçi tarife planlarını uygulama, aboneliği uzaktan etkinleştirme ve devre dışı bırakma gibi kamu hizmetlerine çeşitli yüksek kaliteli avantajlar da sağlamaktadır [24] – [26].



Şekil 2. 1: Akıllı Sayaçların Global Marketteki Değeri [25].

Akıllı elektrik sayaçlarının küresel pazardaki değeri Şekil 2.1 ile verilen değerlerden izlenebildiği üzere 2022 yılının ortalarında 11,4 Milyar ABD doları olarak tahmin edilirken, 2026 yılına kadar 15,2 Milyar ABD doları büyüklüğe ulaşacağı ve analiz süresi boyunca %6,7'lik bir CAGR'de (Compound annual growth rate -Yıllık bileşik büyüme oranı) büyüyeceği tahmin edilmektedir [25].

2.2. ZAMAN SERİLERİ

Zaman serisi, zaman içinde sürekli ölçümlerle alınan bir dizi gözlemden bir başka deyişle veri noktalarından oluşmaktadır. Genellikle, gözlemler eşit aralıklarla alınır. Zaman içinde sırayla toplanan bu verilerde, çeşitli bileşenler belirli bir zaman noktasında gözlemlenen herhangi bir zaman serisi değerine katkıda bulunur. Zaman serilerinde tahmin yöntemlerini seçmek için yararlı bir soyutlama, bir zaman serisini sistematik ve sistematik olmayan bileşenlere ayırmaktır.

- Sistematik (Systematic): Tutarlılığı veya tekrarı olan, tanımlanabilen ve modellenebilen zaman serilerinin birleşenleridir.
- Sistematik Olmayan (Non-systematic): Zaman serisinin doğrudan modellenemeyen serilerinin bileşenleridir.

Belirli bir zaman serisinin seviye, eğilim, mevsimsellik olmak üzere üç sistematik bileşeni ve gürültü adı verilen sistematik olmayan bir bileşenle beraber toplamda 4 bileşenden oluştuğu belirtilmektedir. Bu bileşenler şu şekilde tanımlanır [27], [28]:

- Düzey (Level): Serideki ortalama değer.
- Eğilim (Trend): Serideki artan veya azalan değer.
- Mevsimsellik: Seride tekrar eden kısa vadeli döngü.
- Gürültü: Serideki rastgele değişim.

2.3. ANOMALİLER VE ANOMALİ TESPİTİ

Anomaliler, veri kümesindeki diğer veri noktaları arasında öne çıkan ve verilerdeki normal davranışı doğrulamayan aykırı değerler olarak tanımlanan veri noktalarıdır. Bu veri noktaları veya gözlemler, veri kümesinin normal davranış kalıplarından çıkar. Anomali Tespiti, beklenen davranışa uymayan verilerde örüntü bulma sorunu olarak ele alınır [28] – [30]. Genel olarak anomalilerle ilgili iki ana özellik öne çıkmaktadır:

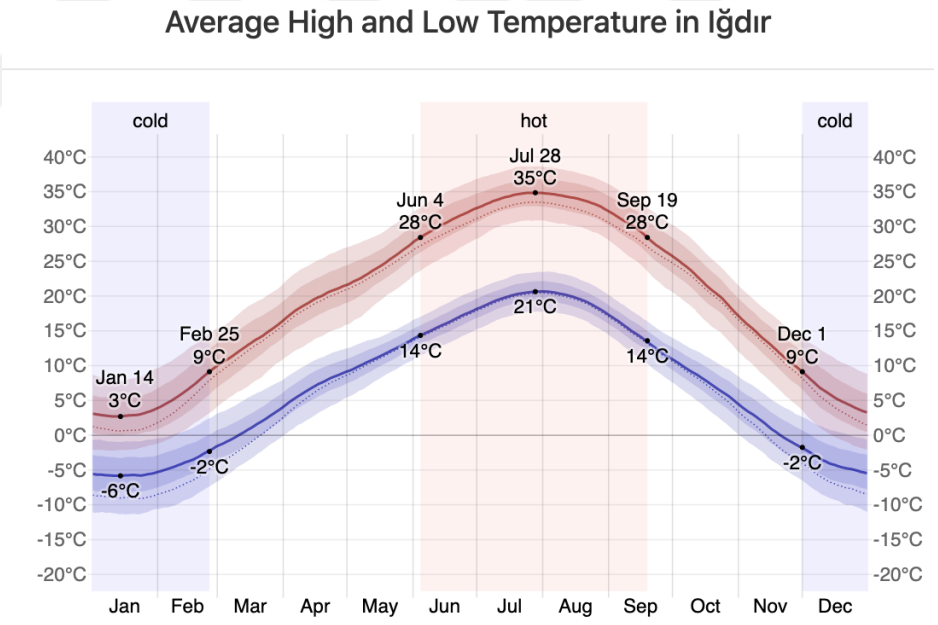
- Anomalilerin dağılımı, verilerin genel dağılımından önemli ölçüde sapmaktadır.
- Veri kümesinin büyük çoğunluğu normal veri noktalarından oluşmaktadır. Anomaliler, veri kümesinin yalnızca çok küçük bir bölümünü oluşturur.

Anomali algılama algoritmaları, izinsiz giriş tespiti (intrusion detection), dolandırıcılık tespiti (fraud detection), veri sızıntısını önleme (data leakage prevention), veri kalitesi, gözetim ve izleme (data quality, surveillance and monitoring) gibi birçok uygulama alanında kullanılmaktadır. Bunlar çok çeşitli ve farklı özellikleri olan uygulamalardır; bazıları çok hızlı, neredeyse gerçek zamanlı anomali tespiti gerektirirken, bazıları bir anomaliyi kaçırmanın yüksek maliyeti nedeniyle çok yüksek performans gerektirir. Anomali tespit teknikleri en yaygın olarak, kötü niyetli girişimlerin/işlemlerin çoğu itibari vakadan farklı olduğu dolandırıcılığı tespit etmek için kullanılır. Anomaliler farklı şekillerde ortaya çıkabilir. Genel olarak, üç farklı anomali türü vardır [27], [28], [31]:

1. *Nokta Anomalileri (Point Anomalies):* Bir nokta verisi içinde bulunduğu veri kümesinden önemli ölçüde saparsa, nokta anomalisi olarak kabul edilir. Örneğin, diğer işlemlerden farklı olarak büyük bir kredi işlemi nokta anomalisidir. Bu nedenle, değeri $[X_{t-k}, X_{t+k}]$, k ∈ R aralığındaki tüm noktalardan önemli ölçüde farklıysa ve k yeterince

büyükse, bir X_t noktası nokta anomalisi olarak kabul edilir. Bu en basit anomali türüdür ve anomali tespiti üzerine yapılan araştırmaların çoğunluğunun odak noktasıdır.

2. *Kolektif Anomaliler (Collective Anomalies):* Tek tek noktaların anormal olmadığı, ancak bir dizi noktanın anomali olarak etiketlendiği durumlar vardır. Örneğin, bir banka müşterisi haftanın her günü banka hesabından 500$ çektiğini varsayalım, Müşteri için ara sıra 500$ çekmek normal olsa da bir dizi para çekme işlemi anormal bir davranıştır. Anormal bir durum birçok örnekten oluşan bir küme olarak temsil ediliyorsa buna kolektif anomali denir.
3. *Bağlamsal Anomali (Contextual Anomaly):* Bazı noktalar belirli bir bağlamda normal olabilirken başka bir bağlamda anormallik olarak algılanabilir: Şekil 2.2 ile gösterilen örnekte, Türkiye'deki Iğdır şehrinde yazın günlük sıcaklığın 35°C olması normal olarak görülürken, kış aylarında aynı sıcaklık anormal olarak kabul edilir. Bağlamsal anomalilerde nokta normal olarak görülebilir, ancak belirli bir bağlam dikkate alındığında nokta bir anomali haline gelir.



Şekil 2. 2: Bağlamsal anomali için örnek Iğdır şehrinin 2022 sıcaklık değerlerinin gösterimi [44].

Nokta anomalilerini tespit edebilen bazı yaklaşımlar ile kolektif veya bağlamsal anomaliler tamamen tespit edilemeyebilir. Bu doğrultuda problem ve veriye uygun yöntemleri uygulamak gerekmektedir. Bu nedenle belli bir anomali tespit tekniğini her alana uygulamak mümkün değildir. Zaman serisi verilerinde zaman, tüm dizideki bir örneğin konumunu belirleyen

bağlamsal bir niteliktir. Bu tez çalışmasında veri yapısından ötürü bağlamsal anomaliler üzerinde çalışılmıştır.

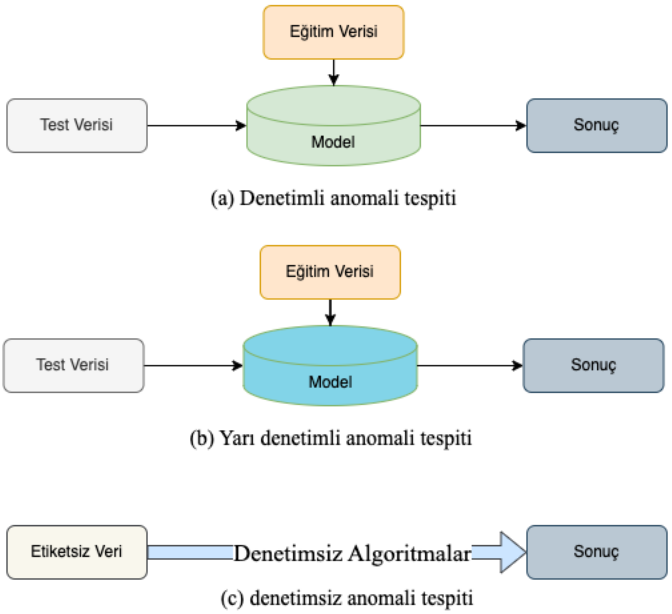
Zaman serisi verilerindeki anomalileri tespit etmek için istatistiksel yaklaşımlar, makine öğrenimi yaklaşımları veya derin öğrenme yaklaşımları uygulanmaktadır. Bu yaklaşımlar, Braei vd. ve Aggarwal'in çalışmalarında iki ana kategoriye ayrılmıştır [28], [32] : *(i) Tahminine dayalı anomali tespiti* ve *(ii) Olağandışı şekillerine dayalı anomali tespiti*. Braei vd., zaman serisi verileri üzerinde birkaç farklı yaklaşımı kullanarak nicel bir karşılaştırma sunmaktadır [28]. Bu yaklaşımların bazıları aşağıdaki gibi kategorize edilmiştir:

- *İstatistiksel yaklaşımlar kullanarak anomali tespiti:* Otomatik Regresyon (Autoregressive- AR), Otoregresif Hareketli Ortalamalar (Autoregressive Integrated Moving Average - ARIMA) Modeli vb.
- *Klasik makine öğrenimi yaklaşımlarını kullanarak anomali tespiti:* LOF, İzolasyon Ormanı vb.
- *Sinir ağlarını kullanarak anomali tespiti:* Çok Katmanlı Yapay Sinir Ağları (Multi Layer Perceptron - MLP), Evrişimsel Sinir Ağları (CNN), LSTM vb.

Ayrıca, algoritmaların hesaplama süresinin yanı sıra her bir yöntemin doğruluğunu analiz ederek, belirli bir veri türü için hangi yöntemin uygun olduğuna dair bazı genel kavramlara da açıklık getirmişlerdir. Braei bir başka çalışmasında, belirtilen kategorilerden 20 tek değişkenli ve 24 çok değişkenli anomali tespit yöntemini inceleyerek bir karşılaştırma sunmuştur [27]. Bu alandaki anomali tespiti konusunda daha birçok çalışma mevcuttur ve bu çalışmaların sayısı giderek artmaktadır [33]- [40].

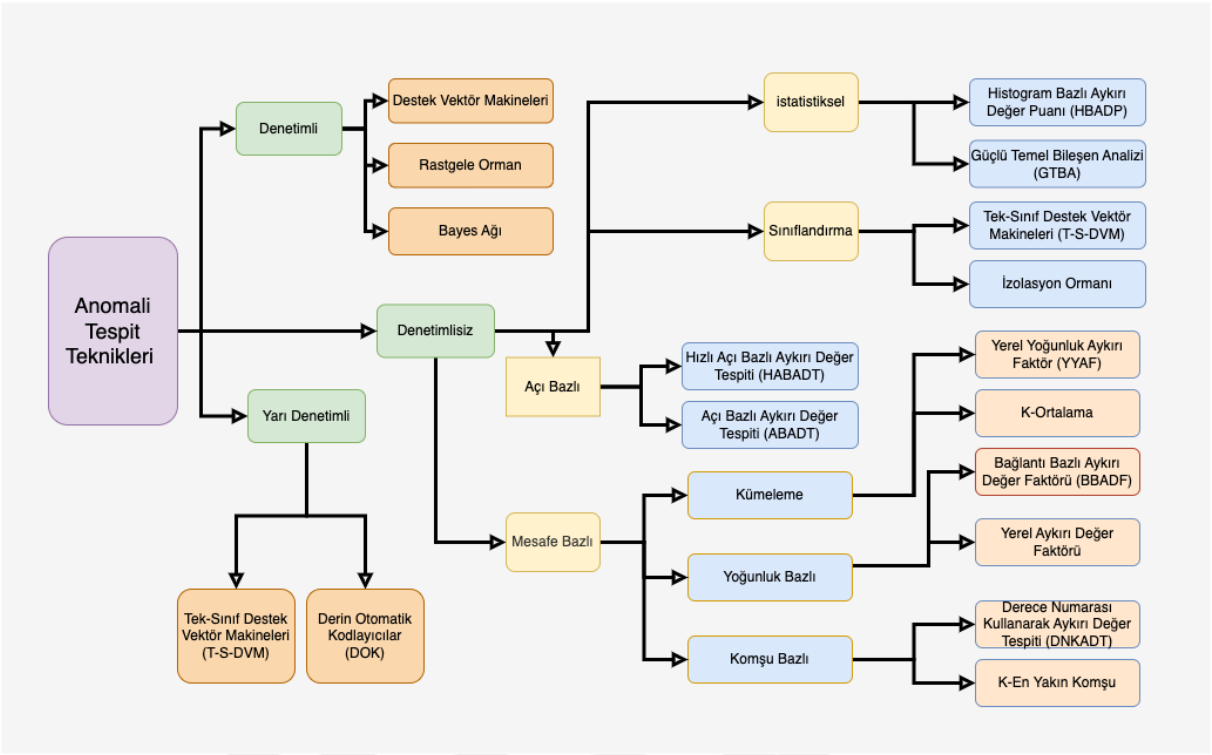
Anomali tespiti; siber güvenlik saldırı tespiti, dolandırıcılık tespiti, arıza tespiti, sistem sağlığı izleme, sensör ağlarında olay tespiti, ekosistem bozulmalarını tespit etme, makine kullanarak görüntülerde kusur tespiti, tıbbi teşhis gibi uygulamalarda beklenmedik davranışları tespit etmek için çok çeşitli uygulama alanlarında kullanılmaktadır. Ayrıca, daha sağlam veri bilimi modelleri oluşturmak için de anomali tespiti teknikleri kullanılabilir [32], [41]. Anomaliler, özellikleri bakımından normalden farklıdır ve normal örneklere kıyasla bir veri kümesinde nadirdir. Makine öğrenme teknikleri, denetimli (supervised), denetimsiz (unsupervised), ve yarı-denetimli (semi-supervised) olarak 3 farklı teknik altında sınıflandırılır. Zaman serilerinde anomali tespiti için genellikle denetimsiz makine öğrenmesine

dayalı yöntemler kullanılmaktadır. Bu çalışma kapsamında da denetimsiz yöntemler ele alınmıştır.



Şekil 2. 3: Farklı anomali tespit yöntemleri ve veri kümesindeki etiketlerin kullanılabilirliği [42,43].

Şekil 2.3 ile gösterilen yapıları şu şekilde detaylandırabiliriz: *(a) Denetimli anomali tespiti (Supervised anomaly detection)*, eğitim için tam olarak etiketlenmiş bir veri kümesi kullanır. Burada denetimli anomali tespiti hem normal hem de anormal verilere erişime sahiptir. Bu nedenle, denetimli bir anomali tespit probleminin özü, anormalliği normalden ayırmaya çalışan bir sınıflandırma problemidir. *(b) Yarı denetimli anomali tespiti (semi-supervised anomaly detection)*, anomali içermeyen bir eğitim veri kümesi kullanır. Daha sonra, anomaliyi tespit etmek için normal modelden test verilerindeki sapmalar kullanılır. *(c) Denetimsiz anomali tespit algoritmaları (usupervised anomaly detection)*, verilerin çoğundan sapan örnekleri tespit etmek için yalnızca verilerin içsel bilgilerini kullanır. Bu yöntemde denetimsiz anomali tespiti, verilen veri kümesinin tam etiketlerine erişim sağlamaz. Veri örnekleri arasında paylaşılan kalıpları belirleyerek ve aykırı değerleri gözlemleyerek anomali tespiti sağlar. Bu nedenle, denetimsiz anomali tespiti görevi, aykırı değer tespiti ile yoğun bir şekilde ilişkilidir.



Şekil 2. 4: Bazı anomali tespiti algoritmaları [42,45].

Sekil 2.4'te bazı anomali tespiti yöntemlerinin bir sınıflandırması yapılmıştır. Bu sınıflandırmayı farklı metrikler için de yapmak mümkündür. Bazı algoritmaların sınıflandırılma yeri birden fazla olabilmektedir. Örneğin LOF algoritmasının denetimli yapısı da vardır.

Schmid vd., zaman serilerindeki anomalileri tespit eden teknikler için 158 kaynaktaki kullanımı inceleyerek sınıflandırmışlardır. Bu sınıflandırmada algoritmalar şu başlıklar altında gruplandırılmıştır [46]:

- Derin Öğrenme- Deep Learning
- Stokastik Öğrenme- Stochastic Learning
- Klasik Makine Öğrenimi- Classic ML
- Aykırı Değer Tespiti- Outlier Detection
- İstatistik (Regresyon ve Tahmin)- Statistics (Regression and Forecasting)
- Veri Madenciliği- Data Mining
- Sinyal Analizi (Signal Analysis)

Domingues vd., anomali tespitinde denetimsiz makine öğrenimi algoritmaların kullanımına yönelik bir özet getirmişlerdir [47]. Anomali tespiti bağlamında, sentetik ve gerçek

veri kümeleri üzerinde 14 algoritmanın ortalama kesinlik, sağlamlık, hesaplama süresi ve bellek kullanımı karşılaştırmışlardır.

Bu tez çalışmasında anomali tespiti için tahmine dayalı algoritmalardan olan FbProphet, iyi bilinen ve yaygın kullanılan anomali tespiti algoritmaları olan LOF ve IF kullanılmıştır.

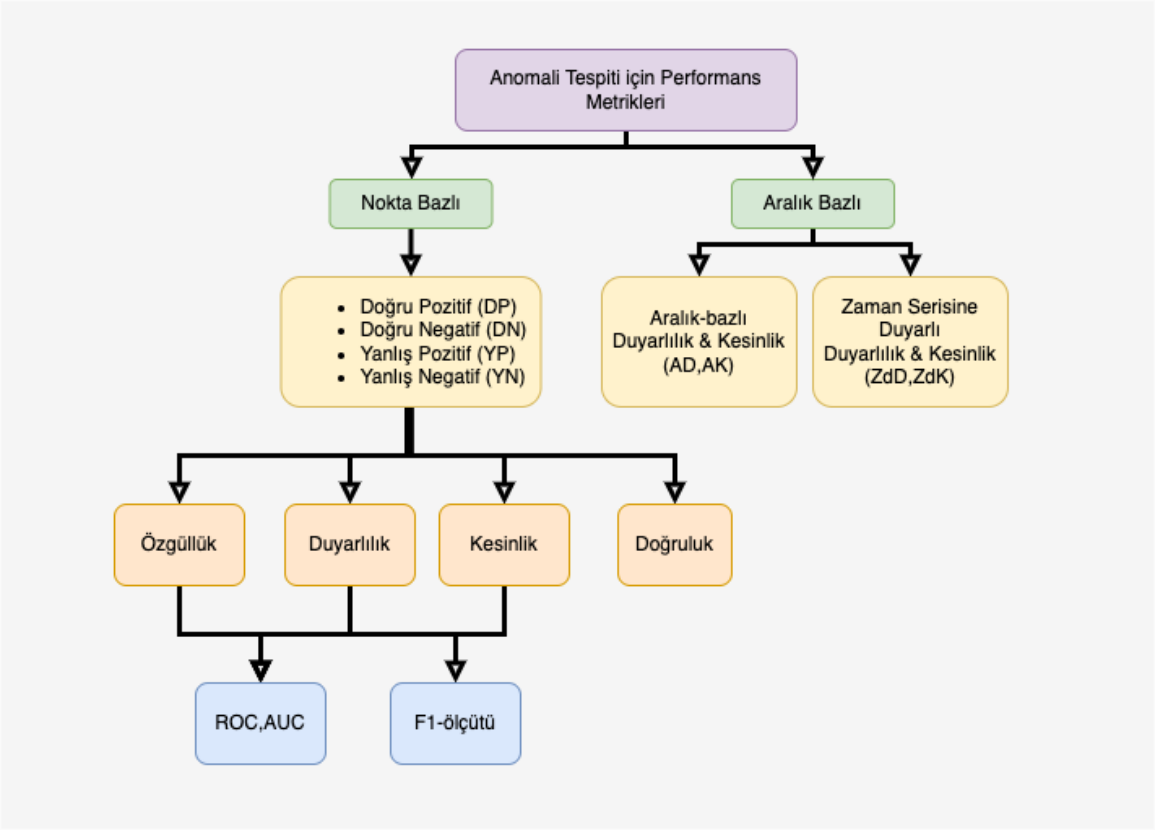
2.3.1. Anomali Tespitinde Başarı Ölçümü

Anomali tespitinde kullanılan algoritmanın başarısını ölçmek için tespit edilen anormal verilerin ne kadar gerçeği gösterdiğinin incelenmesi gerekmektedir. Her model için anormal olarak tespit edilen değerler içerisinde normal veriler de bulunabilir. Burada başarı ölçümdeki temel kabul anomali tespiti için tasarlanan model ya da uygulama gerçeği yansıttığı oranda başarılı sayılmasıdır. Anomali tespiti için tahminleme tabanlı olmayan tekniklerde bazı anomali tespit tekniklerin önemli bir yönü, anomalilerin raporlanma şeklidir [29]. Burada üretilen çıktılar aşağıdaki iki türden biridir:

- Puanlar (Scores): Puanlama teknikleri, o örneğin bir anomali olarak kabul edilme derecesine bağlı olarak, test verilerindeki her bir örneğe bir anomali puanı atar. Bu nedenle, bu tür tekniklerin çıktısı, sıralanmış bir anomaliler listesidir. Bir analist, ilk birkaç anomaliyi analiz etmeyi veya anomalileri seçmek için bir eşik değeri kullanmayı seçebilir.
- Etiketler (Labels): Bu kategorideki teknikler, her test örneğine bir etiket (normal veya anormal) ataması yapar.

Puanlamaya dayalı anomali tespit teknikleri, en alakalı anomalileri seçmek için alana özgü bir eşik kullanmaya izin verir. Test örneklerine ikili etiketler sağlayan teknikler, böyle bir seçim yapmaya doğrudan izin vermez, ancak her teknikteki parametre seçimleri aracılığıyla dolaylı olarak kontrol edilebilir.

Anomali tespiti değerlendirmesi için performans ölçütleri, Şekil 2.5'te gösterildiği üzere nokta-bazlı (point-based) ve aralık-bazlı (range-based) yöntemlere ayrılabilir. Nokta-bazlı yöntemlerde, yalnızca anormal ve normal yanıt etiketleri doğrudur ve yanıt aynıysa puan verilir. Buna karşılık, aralık-bazlı yöntemler yalnızca anormal ve normal etiketlere bölünmekle kalmaz, aynı zamanda performans ölçütlerine göre farklı puanlar da verir. Bu, değerlendiricinin farklı puanlar vermek için farklı ağırlıklar atamasını sağlar.



Şekil 2. 5: Anomali algılama değerlendirmesi için performans metrik ağacı [48], [49].

Şekil 2.5'te belirtilen değerlendirme yöntemi verinin uygunluğuna göre kullanılabilmektedir. Genel anlamda, anomali tespiti ile veri kümesindeki her bir değer için tahmin yöntemi ile normal ya da anormal veri etiketlemesi yapılmaktadır. Ancak, bu iki temel sınıfta yapılan tespitin gerçekte doğru olup olmaması faktörü göz önünde bulundurulduğunda yetersiz kalmaktadır. Burada Doğru Pozitif (TP-True Positive), Yanlış Pozitif (FP-False Positive), Doğru Negatif (TN-True Negative) ve Yanlış Negatif (FN-False Negative) olmak üzere dört sınıf kullanılmaktadır. Şekil 2.5'te belirtilen bazı değerleri aşağıdaki gibi tanımlanabilir:

- Doğru Pozitif (DP): Anomali olarak tespit edilen veriler gerçekte doğrudur.
- Yanlış Pozitif (YP): Anomali olarak tespit edilen veriler gerçekte yanlıştır.
- Doğru Negatif (DN): Normal olarak tespit edilen veriler gerçekte doğrudur.
- Yanlış Negatif (YN): Normal olarak tespit edilen veriler gerçekte yanlıştır.
- Doğruluk (Accuracy): Doğru pozitiflerin ve negatiflerin toplamının tüm veri kümesine oranıdır.
- Duyarlılık (Recall): Uygulamanın doğru pozitifleri tespit etmede ne derece başarılı olduğunu göstermektedir.

- Özgüllük (Specificity): Tespit edilen negatiflerin ne derece yanlış olduğunun göstergesidir.
- Kesinlik (Precision): Tespit edilen pozitiflerin ne derece doğru olduğunun göstergesidir.
- F1-ölçütü (F1-score): Duyarlılık ve kesinlik ölçümlerinin harmonik ortalamasıdır.

Anomali tespiti değerlendirmesi için kullanılan performans ölçütlerinin probleme özel olarak değerlendirilmesi gerekmektedir. Bunun yanında zaman serileri özelinde önerilen bazı metrikler de bulunmaktadır. Hwang vd., zaman serileri üzerinde yaptıkları anomali tespit çalışmasında değerlendirme aşamasında kesinlik ve duyarlılık ölçütlerinin daha dikkate değer metrikler olduğunu ifade etmişlerdir [49]. Gerçek dünya veri kümesini ve çeşitli örnekleri kullanarak yapılan değerlendirmelerde kesinlik ve duyarlılık metriklerinin mevcut metriklere kıyasla zaman serisi verileri için daha uygun olduğunu göstermektedirler. Kim vd., çalışmalarında yine aynı metriklerin zaman serilerinde performans ölçütü olarak kullanılmasını önermişlerdir [48].

Bu tez çalışmasında kullanılan veri kümesi bu değerlendirme yapısına uygunluk göstermemektedir. Çalışma sonucunda elde edilen çıktıya bu yöntemlerin uygulanacağı şekilde herhangi bir yöntem ile kesinlik, duyarlılık ve f1-ölçütü sonuçları değerlendirilmemiş ve bu araştırmada kapsam dışında bırakılmıştır. Bu gibi çalışmalarda sistem ve veri yapısı hakkında yeterli bilgi ve deneyime sahip bir uzmanın geri bildirimleri anomalilerin tespitinde katkıda bulunabilir. Bu çalışmada anomali anlamında olası anomalilerin tespit edilmesi için literatürde sıklıkla kullanılan bazı yöntemler uygulanmıştır.

2.4. ZAMAN SERİLERİNDE TAHMİN YAKLAŞIMLARI

Zaman serisi tahmini, stratejik kararlar vermeye yardımcı olacak tahminler yapmak için istatistik ve modelleme kullanarak zaman serisi verilerini analiz etme sürecidir. Zaman serisi verilerinde yaygın olarak dalgalanan değişkenler ve kontrolümüz dışındaki faktörler olduğundan tahminlerin olasılığı hızlıca değişebilir, bu değişkenlerden ötürü tahminin sonucu kesin değildir. Elimizdeki veriler ne kadar kapsamlı olursa, tahminler o kadar doğru olabilir. Akıllı sayaç verileri, akıllı şebekelerde yük tahmini için önemli değerler getirmektedir. Bu doğrultuda akıllı sayaç verileri ile üretimden son kullanıma kadar çeşitli aşamalardaki veriler üzerinde analiz yapmak ve tahminlerde bulunmak mümkündür. Akıllı sayaçların yüksek ayrıntı düzeyi, tahmin modelinin performansını pozitif yönde etkilemektedir. Literatürde elektrik

enerjisi tüketim tahmini için çeşitli yöntemler ve modeller kullanılmaktadır. Veri üzerinde tahminin türüne bağlı olarak sadece birkaç saat veya 20 yıl gibi uzun vadeli süreler için tahminler yapılabilmektedir. Bu doğrultuda veri üzerinde yapılan bu tür tüketim tahminleri üç alana ayrılmaktadır [13], [22], [50]:

- **Kısa vadeli tahmin**, bir hafta öncesine kadar bir gün tahmin ufkuna sahiptir. Jeneratörler ve elektrik tesislerinin üretim birimlerini talebe göre planlamaları için çok önemlidir. Aralık olarak birkaç dakika ile 1 gün arası tahminler yapar.
- **Orta vadeli tahminin** iki yıla kadar bir tahmin ufku olduğu kabul edilir ve bakım onarım amaçları için kullanılır. Aralık olarak 1gün ile 1 yıl arası tahminler yapar.
- **Uzun vadeli tahmin**, gelecekte birkaç yıllık bir tahmin ufkuna sahiptir. Bu tür bir tahmin, güç şebekesinin genişletilmesinin planlanmasında iletim sistemi operatörleri ve elektrik tesisleri için ilgi çekicidir. Aralık olarak 1 yıldan fazla olan tahminler yapar.

2.4.1. Zaman Serileri Üzerinde Tahmin için Başarı Ölçümü

Modellerin performansını değerlendirmek için ortalama mutlak hatası (Mean Average Error -MAE), ortalama hata kareleri toplamı kökü (Root Mean Square Error - RMSE) , ortalama kare hatası (Mean Squared Error - MSE) ve ortalama mutlak yüzde hatası (Mean Absolute Percentage Error - MAPE) gibi metrikler uygulanmaktadır [51]:

$$\text{MAPE}=\frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \tag{2.1}$$

$$\text{MAE}=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \tag{2.2}$$

$$\text{MSE}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{2.3}$$

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{2.4}$$

(2.1)-(2.4) arası formüllerinde verilen y_i gerçek değerleri, $\hat{y}_i$ modelin tahmin ettiği değerleri ve n değeri de gözlem sayısını temsil etmektedir.

Bu çalışmada modelin performansı değerlendirilirken gerçek değerlerin modelin ürettiği tahminlerle uyumluluğu incelenmiştir. Tahmin edilen değerlerin gerçek değerlere olan uzaklıklarına göre farklı metrikler üretilebilir. Bu metriklerden biri olan MAPE, zaman serisi

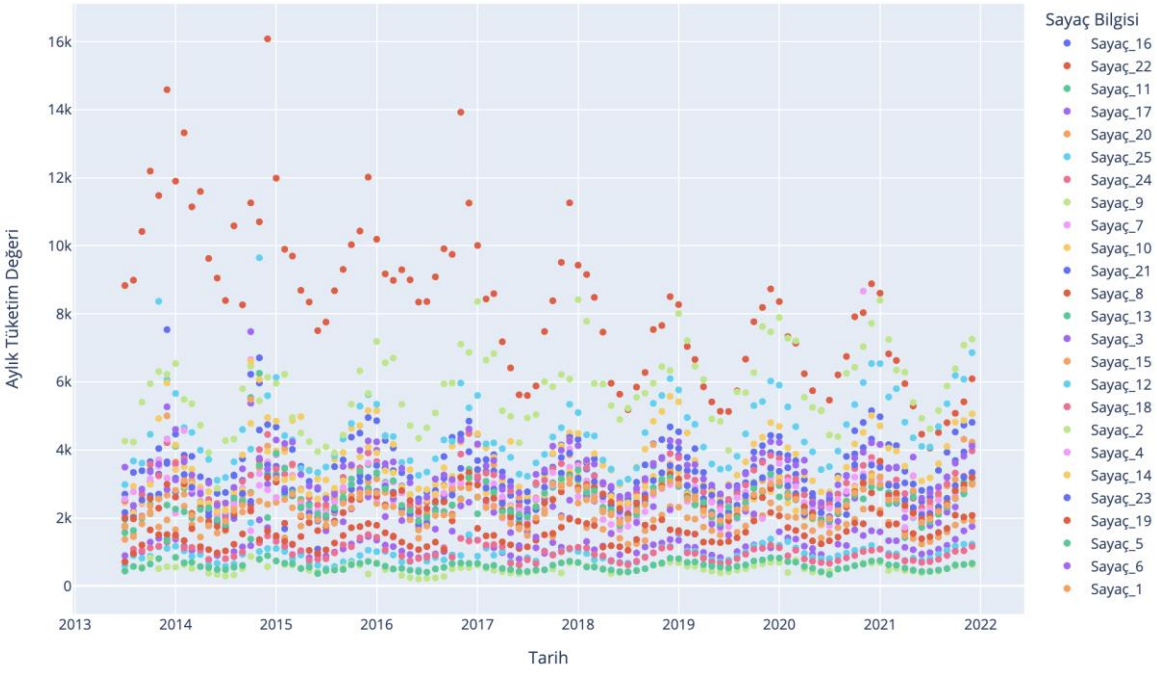
modellerinde tahminlerin doğruluğunu ölçmek için sıklıkla kullanılır. Bu metrik, tahmin değeri ile gerçek değer arasındaki mutlak yüzdelik farkın ortalamasını hesaplamaktadır. Bu çalışmada modelin başarısını yorumlamak için MAPE yöntemi kullanılmıştır.

3. YÖNTEM

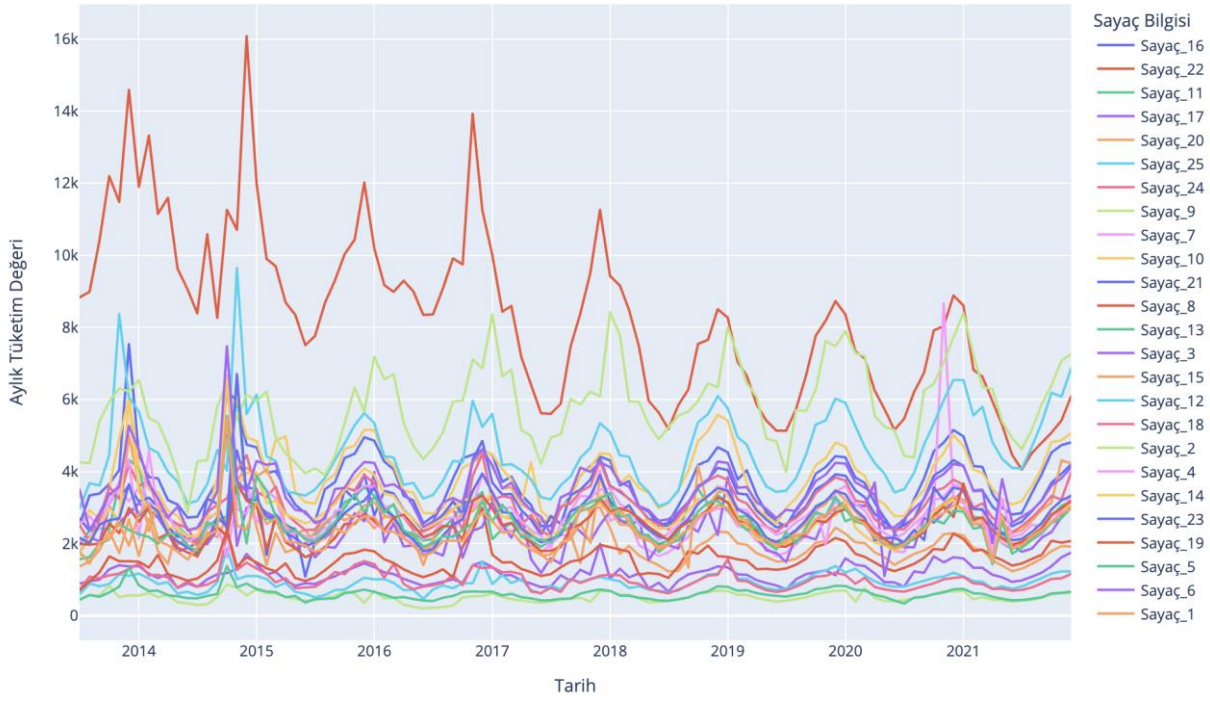
Bu bölümünde, çalışmada kullanılan yöntemler ve veri kümesine açıklık getirilmiştir.

3.1. VERİ KÜMESİ

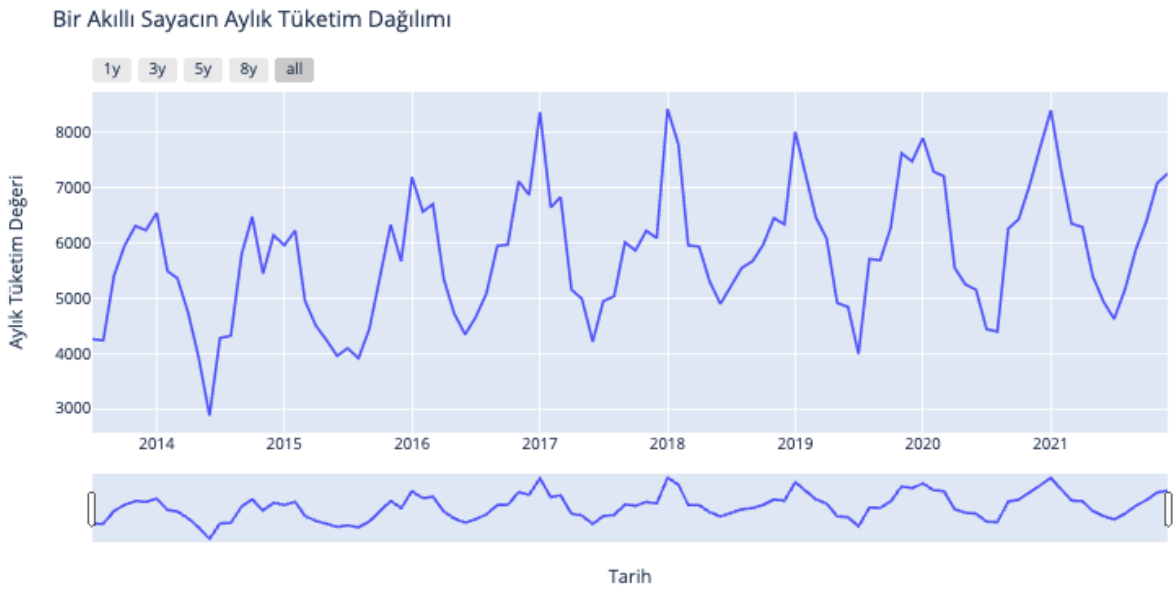
Bu çalışmanın uygulanmasında, aydınlatmada genel kullanımına yönelik cadde, sokak, alt geçit-üst geçit, meydan, park, bahçe vb. gibi alanların aydınlatılmasında kullanılan akıllı sensörlerden (sayaçlar) gelen veriler üzerinde anomali tespit çalışması ve tahminleme yapılmıştır. Veri kümesi tek değişkenli olarak sınıflandırılmıştır [28]. Bu tez çalışmasında kullanılan veri kümesi elektrik dağıtım şirketlerinin kendi sitelerinde paylaştıkları verilerden yararlanarak oluşturulmuş ve anonimleştirilmiştir [52]-[63]. Elde edilen örnek veri kümesi toplam 8 yıllık (2013 ile 2022 arası) ve aylık tüketim değerlerini içeren 25 sayaç verisinden oluşmaktadır. Bu toplanan veriler üzerinde her akıllı sayaçtaki anomalileri hızlı ve doğru bir şekilde tespit edebilmek ve sayaç üzerinde uygun tahmin yapabilmek için akıllı sayaçlardan gelen veriler üzerinde ön işlemler uygulanmıştır. Veri kümesi ile ilgili çeşitli metrikler Şekil 7.1'de veri kümesinin sweetviz_report grafiği olarak ekler bölümünde sunulmuştur.



Şekil 3. 1: Kullanılan veri kümesinin nokta dağılımı.



Şekil 3. 2: Kullanılan veri kümesinin grafiksel dağılımı.

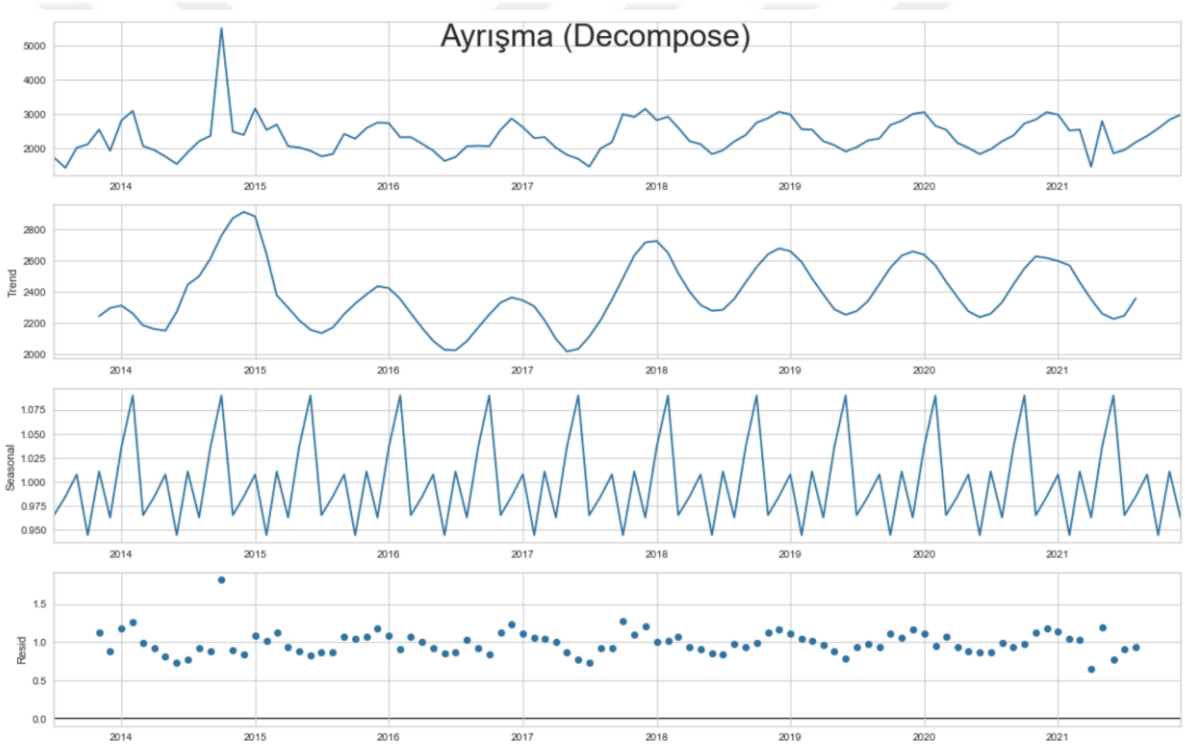


Şekil 3. 3: Bir akıllı sayacın aylık tüketim dağılımı.

Şekil 3.1, Şekil 3.2 ve Şekil 3.3'te görüldüğü gibi veri kümesi mevsimsellik içermektedir ve Tablo 3.1'deki gibi bir yapısı vardır.

Tablo 3. 1: Bir Akıllı sayaca ait aylık tüketim veri yapısı

Tarih	Sayaç Bilgisi	Aylık Tüketim Değeri
2013-07-31	Sayaç_1	1747.000
2013-07-31	Sayaç_2	1450.000
2013-08-31	Sayaç_1	4258.187
2013-08-31	Sayaç_2	4233.026



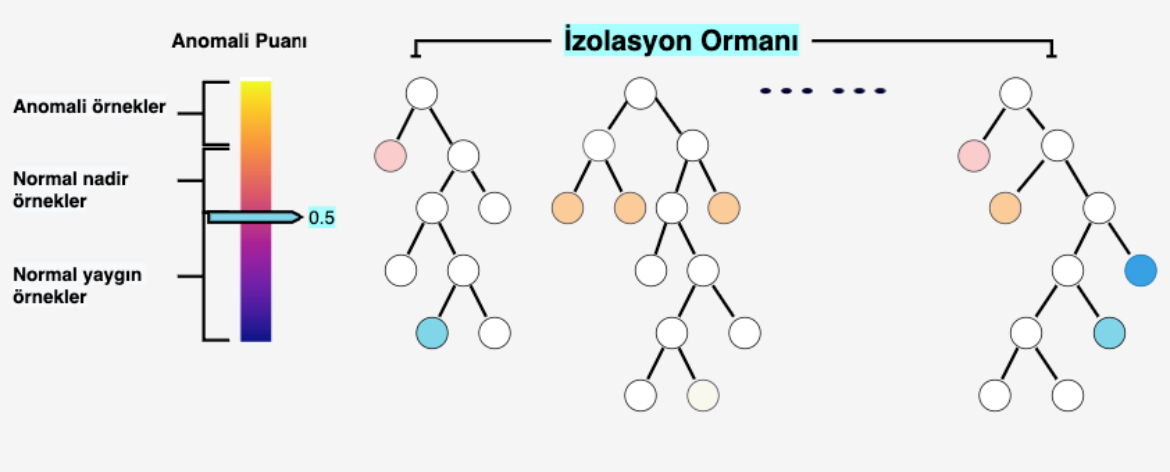
Şekil 3. 4: Bir akıllı sayaca ait aylık tüketim verisinin eğilim, mevsimsellik ve gürültü ayrışımını veren STL ayrışması dağılımı.

Şekil 3.4'te görüldüğü gibi zaman serilerini, eğilimin, mevsimsel indeksin ve artık değerlerin eklemeli (additive) veya çarpımsal (multiplicative) bir kombinasyonu olarak temel düzeyde ayrıştırabiliriz. Burada temel bir istatistiksel ayrıştırma olan kalıntı sapmasını (residue) analiz ederek anomali olabilecek noktaları gözlemleyebiliriz [64].

3.2. KULLANILAN ALGORİTMALAR

3.2.1. İzolasyon Ormanı (Isolation Forest - IF)

Anomaliler, beklenen veri noktalarından çok farklı özellik değerlerine sahip nadir olaylar olduğunu varsayarak, veri noktalarını bir izolasyon ağacının düğümleri olarak yapılandırır. Bu nedenle, anomaliler, yapraklar yerine ağacın köküne daha yakın izole edildiklerinden, beklenen veri noktalarından daha fazla izolasyona duyarlıdır. Bir veri noktasının izole edilebileceği ve daha sonra ağacın köküne olan uzaklığına göre sınıflandırılabileceği sonucu çıkar [45], [65], [66]. IF algoritması, veri kümesindeki aykırı değerleri tespit etmek için rastgele bir orman algoritması (karar ağaçları) kullanan denetimsiz bir anomali tespit algoritmasıdır. Algoritma, veri noktalarını, her bir noktayı diğerlerinden izole edilecek şekilde bölmeye çalışır. Genellikle anomaliler veri noktaları kümesinden uzaktadır, bu nedenle normal veri noktalarına kıyasla anomalileri yalıtmak daha kolaydır. Anomali puanı tüm veri noktaları için hesaplanır ve bu değer gözlemin anomali derecesini belirtir. Anomali puanı ne kadar yüksek ise o derecede aykırı bir gözlem olarak değerlendirilir. Burada uygun eşik (treshold) belirlenebildiğinde anomali puanı eşik değerinden büyük şeklinde olan noktalar anomali olarak kabul edilebilir. Şekil 3.5'te genel IF algoritmasının uygulanma biçimi ve anomali skorun tespit edilmesi gösterilmiştir.



Şekil 3. 5: İzolasyon Ormanı [67].

3.2.2. Yerel Aykırı Değer Faktörü (Local Outlier Factor - LOF)

Komşu tabanlı yöntemler (Neighbor-based methods) aykırı değerleri belirlemek için her bir veri noktasının komşuluğunu (neighborhood) inceler. LOF iyi bilinen bir mesafe tabanlı yaklaşımdır [68]. LOF bir noktanın anomali olup olmadığına karar vermek için veri

noktalarının yoğunluğunu dikkate alan başka bir anomali tespit tekniğidir. Yerel anomalileri bulmak da başarılıdır [45]. Domingues vd., yaptıkları çalışmaya göre LOF, Tek-Sınıf Destek Vektör Makineleri (One-class SVM) 'ne göre daha iyi performans göstermektedir [47]. LOF, noktanın çevresindeki komşuluğa göre ne kadar izole olduğunu ölçen, anomali puanı adı verilen bir anomali puanı hesaplar. Anomali skorunu hesaplamak için yerel yoğunluğun yanı sıra global yoğunluğu da hesaba katar [68]. LOF hesaplaması yapılmasında bazı önemli konseptler mevcuttur, bunlar aşağıda tanımlanmıştır:

- K-mesafesi : Nokta ile K komşuları arasındaki mesafe.
- Reachability Distance (Erişebilirlik Mesafesi): Maksimum mesafe (Öklid, Manhattan, vs.)
- Local Reachability Distance (Yerel Erişebilirlik Mesafesi): Noktanın, komşularından ortalama erişebilirlik mesafesinin tersi.
- Local Outlier Factor: Her bir noktanın, yerel erişebilirlik mesafesinin, k-komşularının ortalama erişebilirlik mesafesine oranı.

Scikit-learn altında bulanan LOF ile anomali tespitini, Aykırı değer (outlier) tespiti ve yenilik tespiti (novelty detection) olmak üzer iki çatı altında ele almaktadır [69]. Aykırı değer tespitinde kullanılacak veri kümesi, anomalilerden arındırılmış, temiz bir veri kümesi değildir. Bu aykırı değer tespitinde denetimsiz algoritmaları ile anomaliler bulunur. Yenilik tespitinde ise, kullanılacak veri kümesi temiz (anomalilerden temizlenmiş) olmalı ve eklenen yeni verinin anomali olup olmadığını yarı denetimli öğrenme algoritmaları ile tespit etmeye çalışır. Anomali tespitinde neighbors.LocalOutlierFactor davranışı Tablo 3.2'de özetlenmiştir.

Tablo 3. 2: neighbors.LocalOutlierFactor davranışı [69]

Metot	Aykırı Değer Tespiti	Yenilik Tespiti
fit_predict	Mevcut	Mevcut değil
predict	Mevcut değil	Sadece yeni veri üzerinde kullanılır
decision_function	Mevcut değil	Sadece yeni veri üzerinde kullanılır
score_samples	negative_outlier_factor_ kullanarak	Sadece yeni veri üzerinde kullanılır
negative_outlier_factor_	Mevcut	Mevcut

3.2.3. Ekstrem Gradyan Arttırma (eXtreme Gradient Boosting – XGBoost)

XGBoost, Gradient Boosting algoritmasının optimize edilmiş yüksek performanslı geliştirilmiş versiyonudur. Tianqi Chen vd., 2016 yılında yayınladıkları makale bu algoritma literatüre dahil olmuştur [70], [71]. XGBoost'un en önemli özellikleri yüksek tahmin gücü elde edebilmesi, aşırı öğrenmenin önüne geçebilmesi, boş verileri yönetebilmesi ve bunları hızlı yapabilmesidir. Tianqi Chen vd., XGBoost diğer popüler algoritmalardan daha hızlı çalışmakta olduğunu belirmektedirler. Literatürde kullanımı oldukça fazladır [51], [72], [73] vb.

3.2.4. FbProphet

FbProphet, Facebook ekibi tarafından geliştirilen, tek değişkenli (univariate) yüksek oranda mevsimsel zaman serisi verilerini eklemeli (additive) bir modele dayalı olarak analiz eden ve tahmin eden, trend değişiklikleri için tahmin değerleri ve güven aralıkları üreten açık kaynaklı bir kütüphanedir [74], [75]. Model için otomatik olarak iyi bir hiper-parametre küme bulma kapasitesine sahiptir. Bu kütüphane doğrusal olmayan zaman serileri verileri üzerinden dönemlik, haftalık, aylık, günlük tahminlerde bulunmayı sağlayan prosedürleri içermektedir. Zaman serilerinde kullanılan diğer tahminleme mekanizmalarına (ARIMA vb) göre eksik ve uç değerler için daha iyi sonuçlar vermekte, tahmin işlemi yaparken değişim noktaları, mevsimsellik, özel gün etkileri gibi faktörleri dikkate alma özelliğine sahiptir.

FbProphet modeli özünde dört ana bileşeni olan bir regresyon modelidir:

- Parçalı doğrusal veya lojistik büyüme eğrisi eğilimleri.
- Verilerdeki değişiklik noktalarını seçerek trendlerdeki değişiklikleri (change points) otomatik olarak algılama.
- Fourier serisi kullanılarak modellenen yıllık bir mevsimsel bileşen.
- Kukla değişkenleri kullanılarak oluşturulan haftalık bir mevsimsel bileşen.
- Önemli tatillerin kullanıcı tarafından belirtilmesine imkân verme.

3.3. ANOMALİ TESPİTİ VE TAHMİNLEME

Anomali algılama ile tahminleme yöntemleri anomali tespiti ve tahminleme için aynı çatı altında değerlendirebileceği gibi ayrı konseptler olarak da veri kümesi üzerinde farklı metotlar ile uygulanabilir. Bu çalışma özelinde anomali tespiti için, FbProthet kütüphanesinin sunduğu çözüm ile bu iki konsept beraber değerlendirilip uygulanmıştır. XGBoost yöntemi ile sadece tahminleme, Isolation Forest ve LOF ile de olası aykırı değer tespiti (anomali algılama)

yapılmış ve XGBoost ile FbProthet yöntemleri ile elde edilen tahminleme sonuçları MAPE metriklerine göre karşılaştırılarak değerlendirilmiştir.

3.3.1. Anomali Tespitinde Kullanılan Yapılar

Bu araştırmada anomali anlamında olası anomalilerin tespit edilmesi için IF, LOF ve FbProphet beraber uygulanmıştır.

3.3.1.1. FbProphet Kütüphanesinin Uygulanması

Burada uygulanan yöntem ile FbProphet kütüphanesinin tek değişkenli zaman serisi üzerinde tahminleme (forecasting) yapması sonucunda üst (upper_bound) ve alt (lower_bound) limitler belirlenmektedir. Normal şartlarda var olan değerin tahmin edilen ile uyuşmaması anomali olarak atanabilirken, FbProphet'in alt ve üst tahmin limitlerinden yararlanarak potansiyel olarak daha yüksek olan noktalara önem derecesi verilerek anomaliler belirlenebilmektedir. FbProphet ile anomalilerin belirlenmesi, modelin güven aralıklarına bağlıdır. FbProphet kütüphanesi kullanımında yapılan aksiyonlar aşağıdaki gibi belirtilmiştir:

- FbProphet iki özellikli bir veri kümesini girdi olarak almaktadır. Bu özelliklerden ilki olan ds zaman damgasıdır. Diğer özellik olan y ise tahmine konu olan nümerik ölçüm değeridir. İçerisinde ds ve y değerleri içeren veri kümesi ile eğitilen model farklı periyotlarda tahminler üretebilir.
- Model üzerinden yapılan tarih bazlı gelecek tahminleri bir değer aralığı olarak üretilir. Muhtemel tahmin değeri yhat, tahmin üst sınırı yhat_upper ve tahmin alt sınırı yhat_lower değerleri ile bir değer bandı olarak üretilir.
- Her bir ay için tüketim değerleri tahmin edilmiştir. Yani zaman serisinin bir sonraki değeri tahmin edilmiştir.
- Elde edilen tahmin değerinin alt ve üst sınırları kullanılarak güven aralıkları belirlenmiştir. Bu güven aralığı dışında kalan değerler anomali olarak tanımlanmıştır.
- Analiz sürecinde, sayaçların eş zamanlı olarak analiz edilebilmesi için Apache Spark kullanılmıştır. Spark, paralel işlem yapmaya olanak tanırken hem analiz süresini düşürmekte hem de sunduğu ara yüz (Spark-Shell) sayesinde "job" olarak tanımlanan işlemlerin görüntülenmesini ve olası bir hatanın detaylı incelenmesini sağlamaktadır.

- Çalışmada, ilgili genel aydınlatma sayacının son 8 yıllık verileri analiz edilerek her ay için enerji tüketiminde bir anomali olup olmadığı tespit edilmiştir. Anomali tespiti yapılırken ilgili aya ilişkin tüketim değerinin model tahmin ile belirlenen alt ve üst beklenti sınırlarının dışında kalıp kalmadığı kontrol edilmiştir.
- Kurulan modelin başarısını ölçmek için MAPE kullanılmıştır.
- Tespit edilen anomali notlarının önem derecesi hesaplanmıştır. Tespit edilen anomalilere ilişkin gerçek değerlerin ilgili tahmin üst ve alt sınırlarına olan mutlak uzaklıkları ölçülmüştür. Sonrasında, söz konusu uzaklıklar için min-max normalizasyonu yapılarak 0-1 arasındaki önem dereceleri elde edilmiştir.

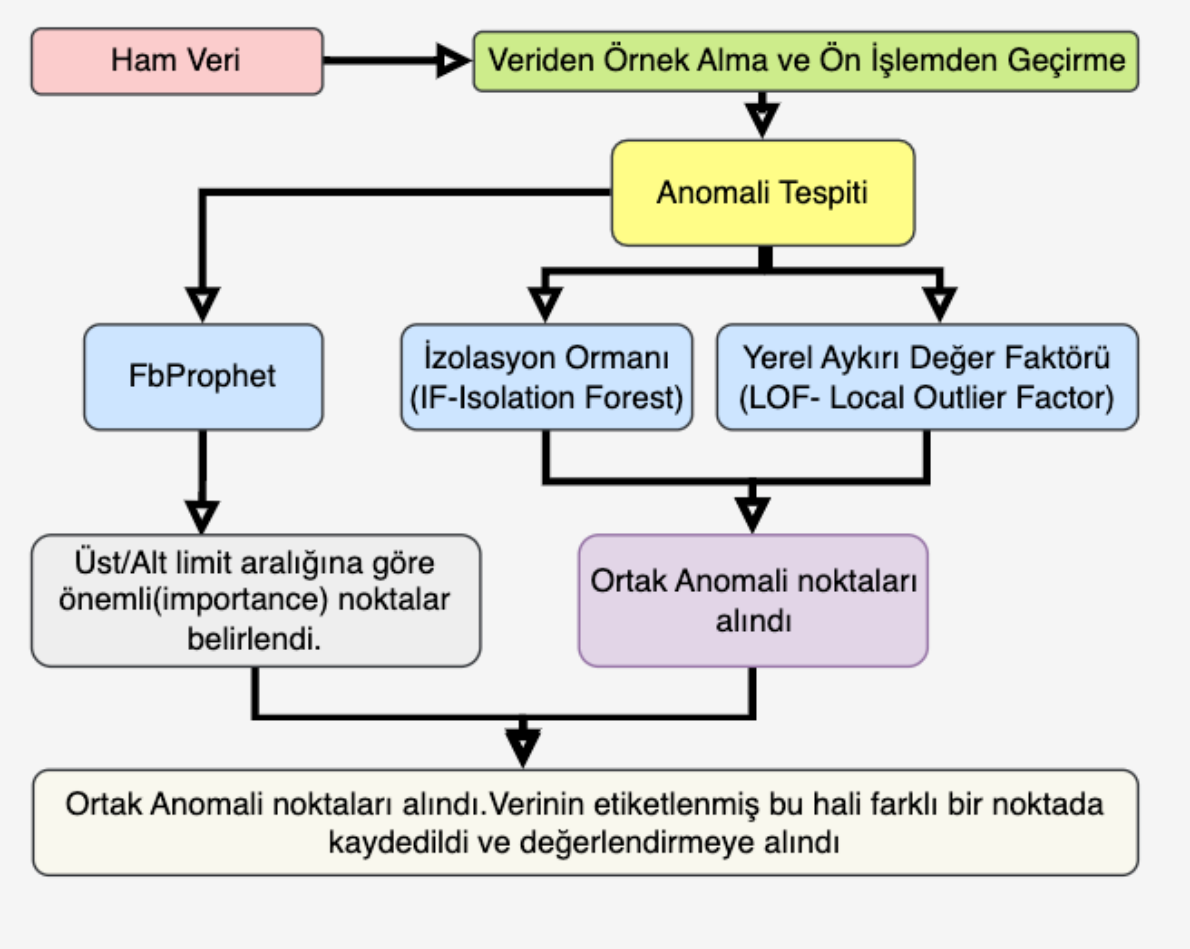
3.3.1.2. İzolasyon Ormanı (IF) ve Yerel Aykırı Değer Faktörü (LOF) Uygulanması

Bu yöntemler ayrı ayrı değerlendirilebildiği gibi hibrit olarak kullanıldığı çalışmalar da mevcuttur. Genel olarak IF algoritması yalnızca küresel aykırı değerlere karşı hassastır ve yerel aykırı değerlerle başa çıkmakta zayıftır. LOF, yerel aykırı değer tespitinde iyi performans göstermesine rağmen, yüksek zaman karmaşıklığına sahiptir. Cheng vd., IF ve LOF'un zayıflıklarının üstesinden gelmek için, aykırı değer tespiti için iki aşamalı bir kolektif yöntem önermişlerdir [76]. Bu önerilen yöntem ile düşük zaman karmaşıklığına sahip karmaşık veri kümelerindeki aykırı değerleri doğru bir şekilde tespit edileceğini göstermişlerdir. Aynı zamanda, çalışmalarında IF-LOF'u IF, LOF ve farklı bir birkaç yöntemle de karşılaştırmışlardır. Sonuçlarında mevcut yöntemlerle karşılaştırıldığında, önerilen kolektif yöntemin aykırı değer algılama oranını önemli ölçüde iyileştirebileceğini ve zaman karmaşıklığını büyük ölçüde azaltabileceğini göstermişlerdir.

Bu çalışmada olası anomali noktalarının doğruluk derecesini artırmak için IF ve LOF ve FbProphet algoritmaları beraber değerlendirilmiştir. Burada sırasıyla IF ve LOF beraber değerlendirilip sonrasında FbProphet çıktıları ile süzgeçten geçirilmiştir. Veri kümesi üzerinden IF ve LOF algoritmaları ile potansiyel anomali olabilecek noktalar belirlenmiştir. Buradaki amaç veri kümesi üzerinde veriyi şekillendirmek ve anomali noktalarını belirleyip değerlendirmektir. Bu iki algoritmanın bulduğu anomali noktaları karşılaştırılarak ortak olanları alınmıştır. Uygulanan algoritmanın bulduğu noktalar gerçek anomali olmama ihtimali taşımaktadır. Bu algoritmaların anomali olarak belirlediği noktalar için aynı zamanda anomaly_score değeri de hesaplanmaktadır. Yani, bu yöntemler ile anormali tespit modeli gözlemin anomali olup olmadığına dair kategorik bir etiket ekleme ve bir skor değeri atamaktadır. Skor bize gözlemin ne derecede anomali olduğunu da söylediğinden etikete göre

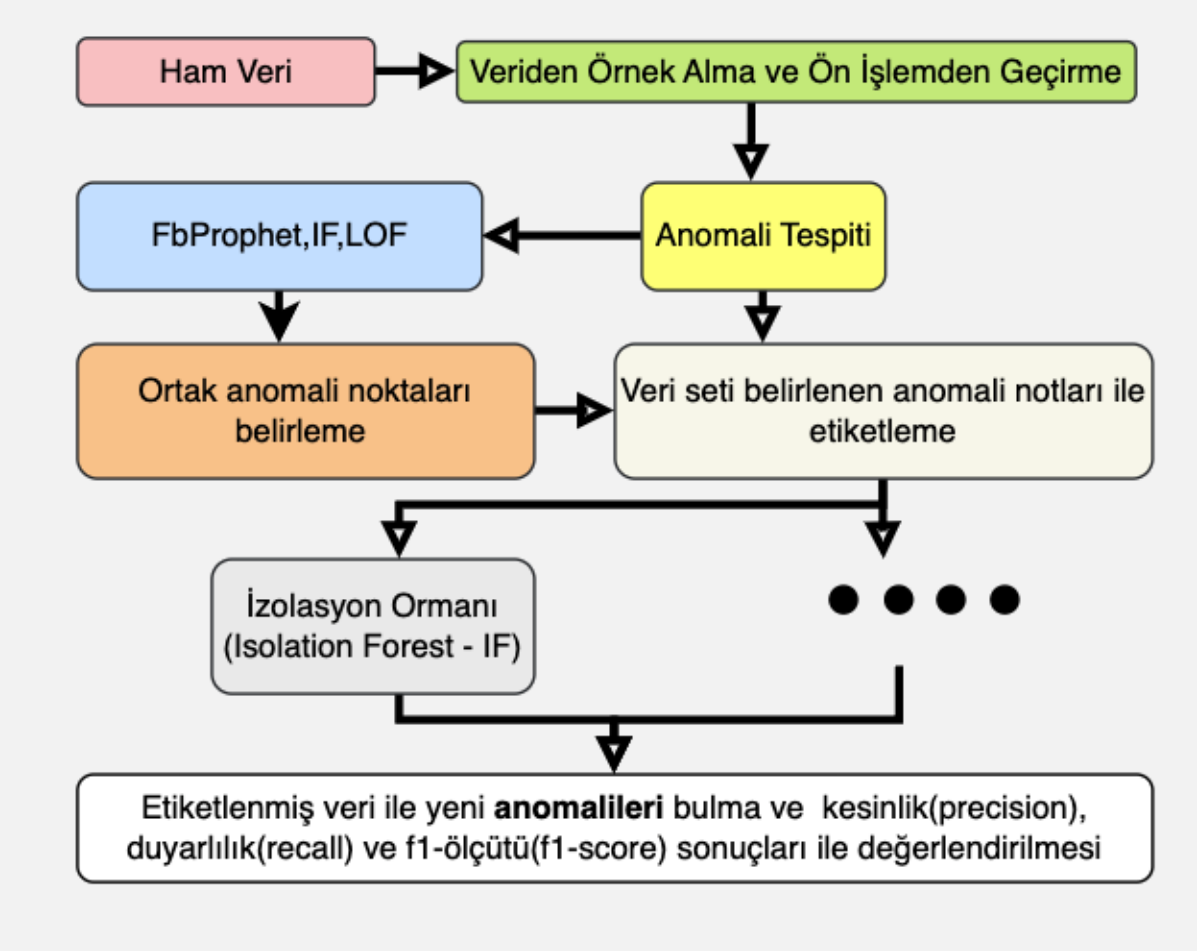
daha çok bilgi taşır. IF ve LOF ortak çıktıları ile FbProphet harmanlanıp anomalilerin belli olduğu veri kümemizin son halini elde edilmiştir. Burada veri kümesi üzerinde anomali için bir eşik değer (threshold) belirleme mümkün değil çünkü başvurabileceğimiz bir doğruluk verisi içermemektedir. Veri kümesi alarm ve değerlendirmeler ile zenginleştirilmesi için rapor sistemi oluşturulabilir. Sonuç olarak üç algoritmanın yapısından faydalanarak bulunan ortak anomali noktaları birleştirilerek, veriler farklı bir noktada kayıt altına alınıp, uygulama kapsamında değerlendirme için alarmlar da oluşturulabilir. Bu anomali alarmlarının önemli olanları kritik dereceye göre sınıflandırılabilir.

IF ve LOF algoritmaları önemli bir kontaminasyon (contamination) parametresine sahiptirler. Bu parametre, zaman serisi verilerimizdeki anomalilerin sayısını belirtir. Verilerimizdeki noktaların yüzdesini anormal olacak şekilde ayarlar. Bu değer veri yapımızda referans alınacak bir yer olmadığından “auto” olarak belirlenmiştir. Veri kümesinin anomalilerle nasıl "kirlendiğini"nin belirlenmesi gerekmektedir. Buradaki yaklaşımımız, sistemdeki tüm anomali noktalarını yakalamaktır. Bu yüzden, normal olabilecek birkaç noktayı anormal (yanlış pozitif) olarak belirlemek, bir anomaliyi (gerçek negatif) yakalamayı kaçırmaktan daha iyidir. Bu çalışmadaki anomali tespiti akışı Şekil 3.6’da özetlenmiştir.



Şekil 3. 6: Tez çalışmasındaki anomali tespiti akışı.

Genel olarak, sınıflandırma tekniği kullanılarak verilerdeki anomalilerin tespit edilebilmesi için verilerin anomali olarak etiketlenmesi gerekir. Yani veri kümemizde daha önce anomali olarak tanımlanmış ve anomali olarak etiketlenmiş veriler bulunmalıdır. Veri kümesinde bu şekilde etiketlenmiş veriler varsa tüm sınıflandırma algoritmaları bu amaçla kullanılabilir. Burada elde ettiğimiz yeni veri kümesinin sınıflandırma gibi ileride yapılacak çalışmalarda Şekil 3.7'dekine benzer bir akış ile detaylı işlenmesi planlanmaktadır.



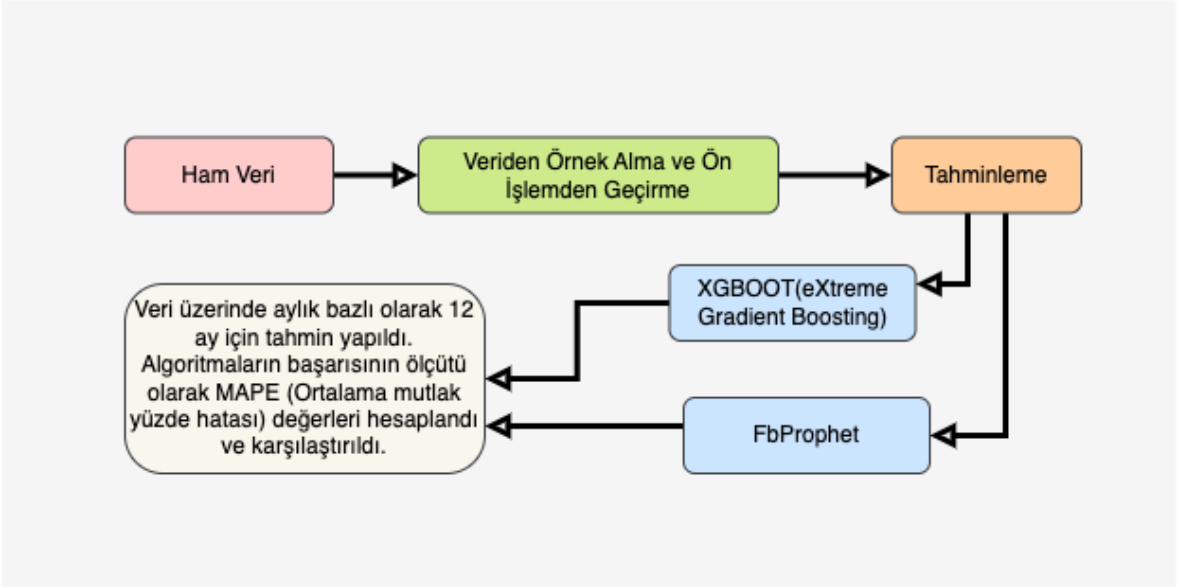
Şekil 3. 7: Bu çalışmanın çıktısı üzerinde ileri ki zamanlarda yapılabilecek çalışma örneğinin akışı.

3.3.2. Tahmilemede (forecasting) Kullanılan Yapılar

Çalışma kapsamında yapılan tahminlemenin ileriye dönük olması hedeflenmiştir. Her sayaç üzerinden ayrı ayrı modelleme yapılmıştır. FbProphet ile geçmişe yönelik tahminlemeler yapmak mümkün olduğundan anomali tespitinde değerlendirmenin yanında, zaman serisinin mevsimsellik kalıplarını başarılı bir şekilde tespit edebildiğinden ileriye yönelik tahmin yapabilmektedir. Bu çalışmada her sayacın ileriye yönelik bir aylık değeri tahmin edilmiştir. Buradaki başarı geçmişe yönelik yapılan tahmin değerlerinin başarısına göre yorumlanmıştır. En iyi modelin seçilmesinde ortalama tahmin hata (MAPE) çıktısına göre hiper parametreler belirlenmiştir.

XGBoost algoritması ile tahminleme için belirli bir 102 aylı veri kümesinden 12 aylık kısmı test için kullanılmıştır. Her sayaç üzerinden ayrı ayrı modelleme yapılarak MAPE çıktısına göre hiper parametreler belirlenmiştir. Bu algoritma ile FbProhpet'a benzer ileriye dönük her sayaç için bir aylık değer tahmin edilmiştir. Bu iki algoritmanın başarım

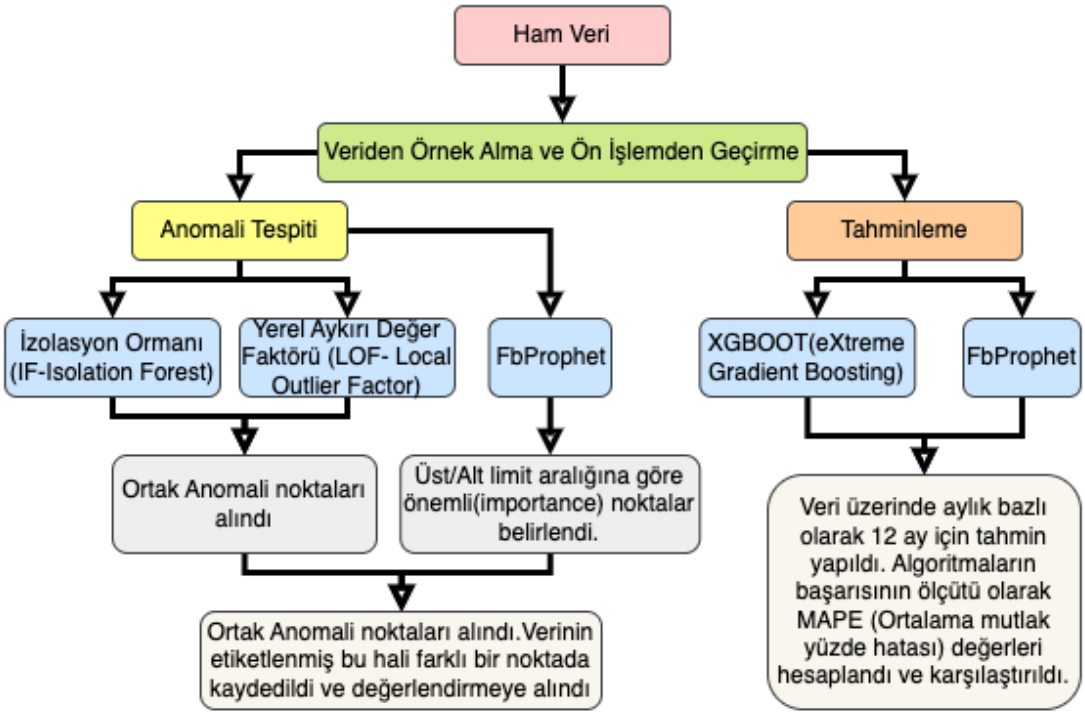
karşılaştırılması MAPE değerleri üzerinden yorumlanmıştır. Bu yapıların uygulanmasının akışı şekil 3.8'de özetlenmiştir.



Şekil 3. 8: Bu çalışmada kullanılan veri üzerinde tahminleme akışı.

4. BULGULAR

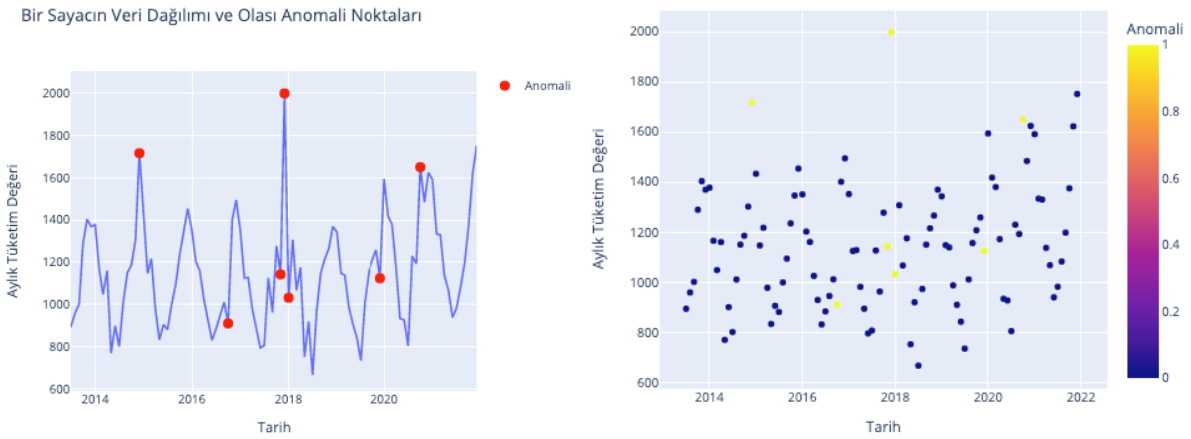
Bu bölümde, çalışma kapsamında önerilen yöntemin uygulanmasından elde edilen sonuçlara yer verilmiştir. Burada, FbProphet, XGBoost, LOF ve IF algoritmaları ile yapılan deneylerin sonuçları ele alınmıştır. Şekil 4.1'de çalışmanın genel akış şeması özetlenmiştir.



Şekil 4. 1: Çalışmanın akış şeması.

4.1. ANOMALİ TESPİTİ İLE İLGİLİ BULGULAR

Bu çalışmada her sayaç için FbProphet, LOF ve IF ile ayrı ayrı anomali tespiti çalışması yapılmıştır. Genel bir sayaca ait verilerin dağılımı ve olası anomali noktalarını belirten grafik Şekil 4.2'de gösterilmiştir.



Şekil 4. 2: Bir sayaca ait veri dağılımı ve olası anomali noktaları.

Tablo 4.1'de anomali tespiti için bu çalışmada kullanılan algoritmaların hiper parametrelerin listesi gösterilmiştir.

Tablo 4. 1: Anomali için Kullanılan Algoritmaların Hiper Parametre Listesi.

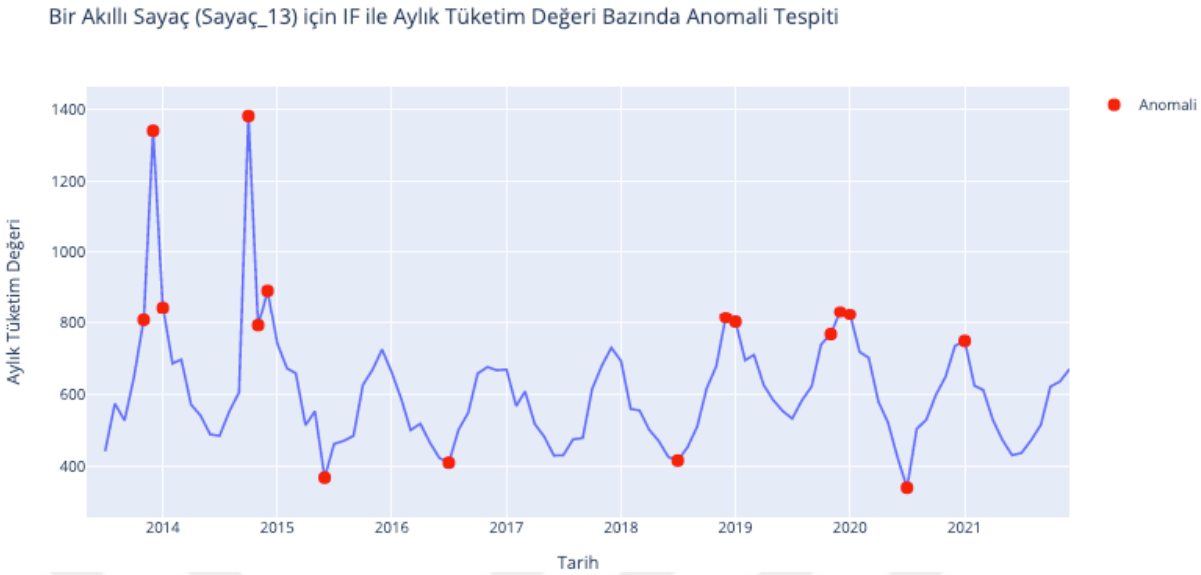
Algoritma	**Parametre**	**Atanan Değer**
LOF	k	12
	distance function	Minkowski distance
	Kontaminasyon (Contamination) değeri	Find automatically
	novelty	False
IF	iTrees Sayısı	12
	Kontaminasyon (Contamination) değeri	Find automatically
	n_estimators	12
FbProphet	interval_width	0.90
	yearly_seasonality	True
	weekly_seasonality	False
	daily_seasonality	False
	seasonality_prior_scale	0.1
	freq	MS (for month start)
	periods	1

Tablo 4.2'de görüldüğü gibi her bir yöntem, farklı sayıda anomali noktası belirlemiştir. LOF ve IF algoritma yaklaşımlarına göre anomali olarak belirlenen noktalara skor(score) yapısını da sunmaktadır.

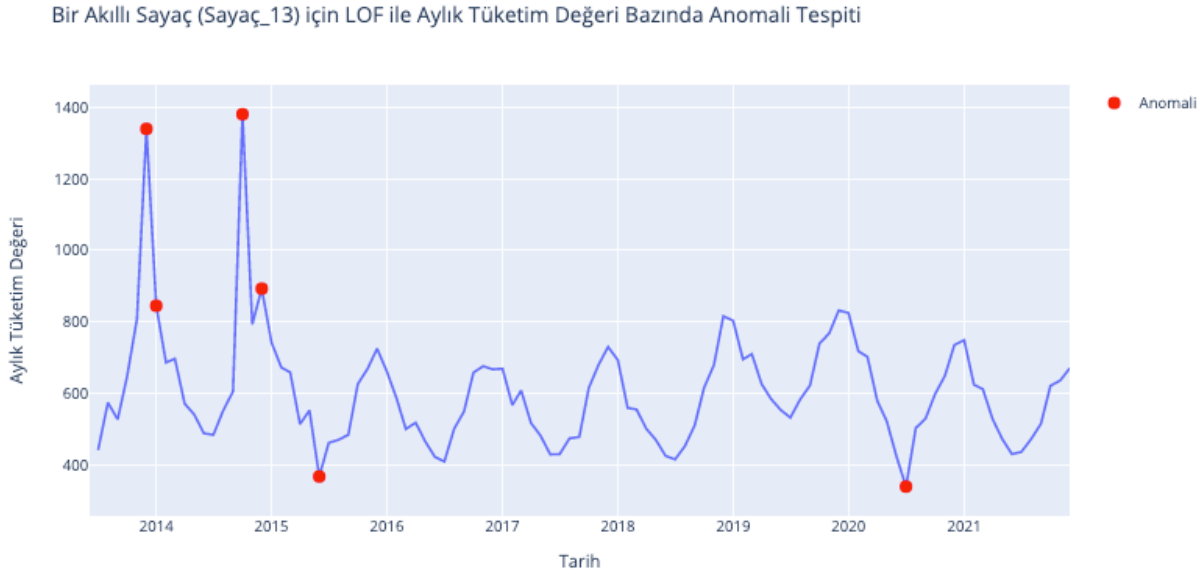
Tablo 4. 2: Kullanılan yaklaşımlarının anomali çıktıları.

Sayaç	Toplam Örnek Sayısı	FbProphet - Anomali nokta sayısı	LOF- Anomali nokta sayısı	IF -Anomali nokta sayısı
Sayaç_1	102	6	5	11
Sayaç_2	102	8	4	25
Sayaç_3	102	4	8	18
Sayaç_4	102	7	8	10
Sayaç_5	102	8	9	18
Sayaç_6	102	6	7	12
Sayaç_7	102	2	6	14
Sayaç_8	102	5	5	12
Sayaç_9	102	14	12	23
Sayaç_10	102	8	7	17
Sayaç_11	102	6	7	10
Sayaç_12	102	6	5	16
Sayaç_13	102	3	6	16
Sayaç_14	102	5	10	14
Sayaç_15	102	7	9	23
Sayaç_16	102	5	5	18
Sayaç_17	102	6	13	22
Sayaç_18	102	11	7	20
Sayaç_19	102	8	5	17
Sayaç_20	102	12	13	25
Sayaç_21	102	8	9	26
Sayaç_22	102	7	7	20
Sayaç_23	102	2	8	10
Sayaç_24	102	4	5	21
Sayaç_25	102	4	7	14
Toplam	2550	162	187	432
Ortalama Anomali Sayısı	2550	%6.3	%7.3	%16.9

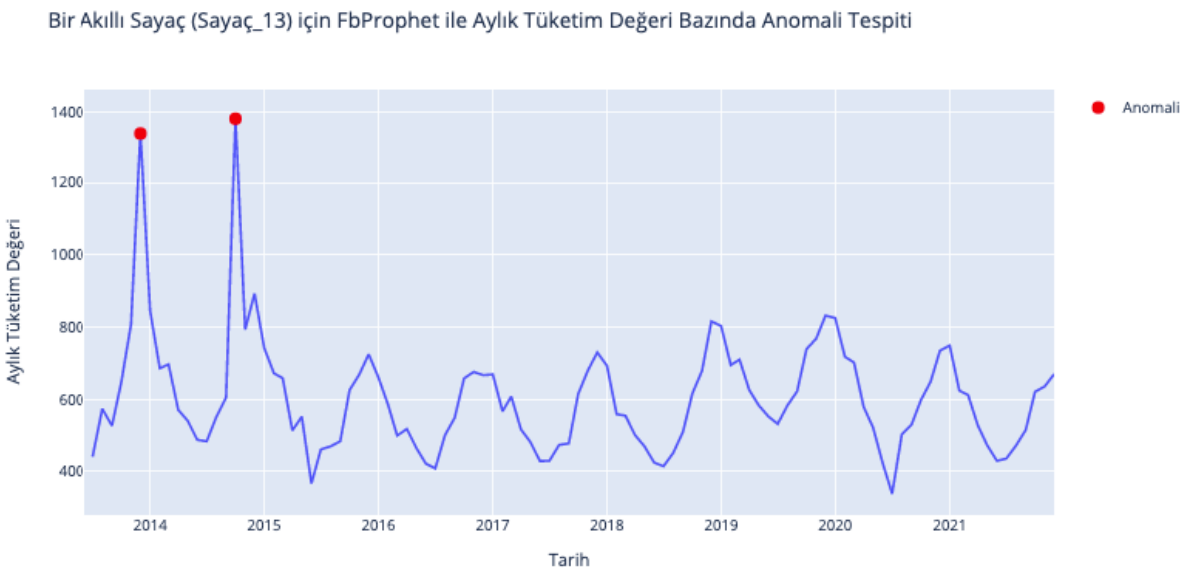
Şekil 4.3, Şekil 4.4 ve Şekil 4.5'te görüldüğü üzere LOF, Isolation Forest ve FbProphet algoritmalarıyla örnek aynı sayaç numarası (Sayaç_13) ile bakıldığında benzer şekilde olası üç anomali noktalarının tespit edilebildikleri gözlemlenmiştir. LOF ve IF çıktılarını skor sonuçları ile değerlendirilebileceği gibi FbProphet yöntemi ile de tespit edilen noktaların önem derecesi Şekil 4.6'teki gibi belirlenebilir.



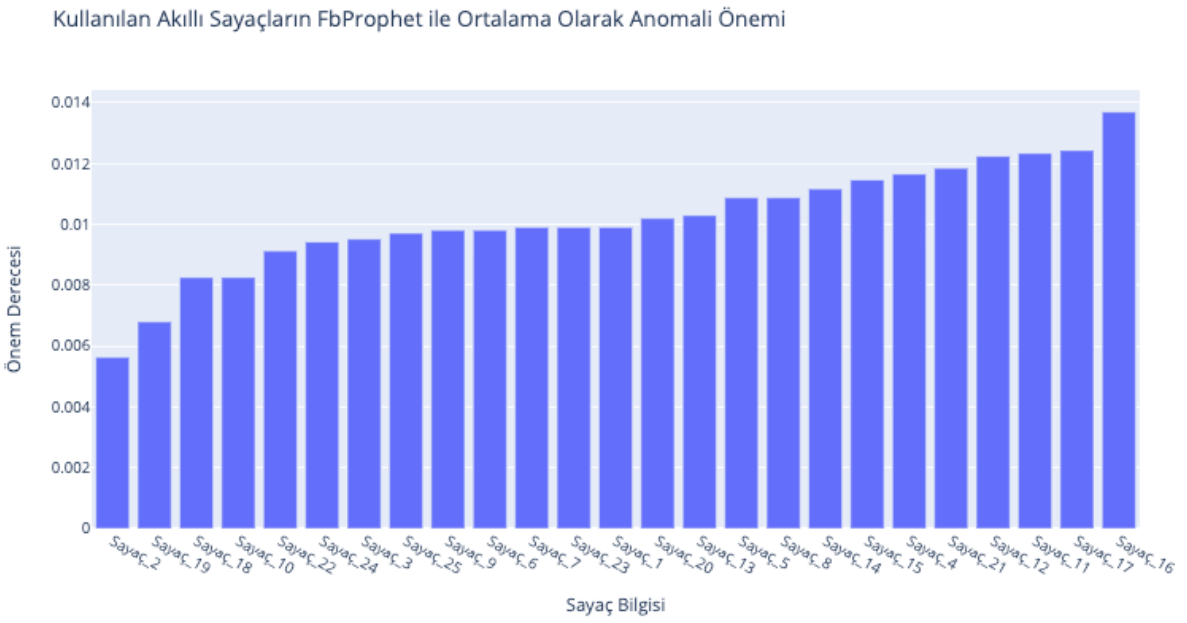
Şekil 4. 3: Bir Akıllı Sayaç (Sayaç_13) için IF ile Aylık Tüketim Değeri Bazında Anomali Tespiti.



Şekil 4. 4: Bir Akıllı Sayaç (Sayaç_13) için LOF ile Aylık Tüketim Değeri Bazında Anomali Tespiti.



Şekil 4. 5: Bir Akıllı Sayaç (Sayaç_13) için FbProphet ile Aylık Tüketim Değeri Bazında Anomali Tespiti.



Şekil 4. 6: Kullanılan Akıllı Sayaçların FbProphet ile Ortalama Olarak Anomali Önemi.

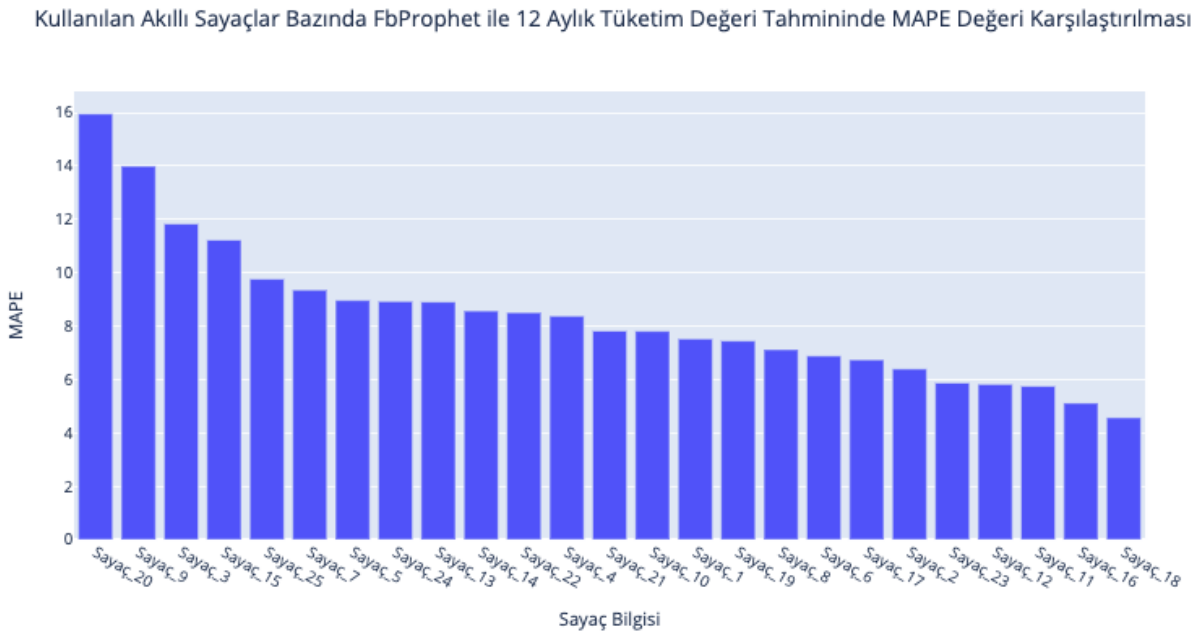
4.2. TAHMİNLEME İLE İLGİLİ BULGULAR

FbProphet ile XGBoost algoritmaların uygulaması sonucundu hesaplanan MAPE değerleri ve uygulanan sayaç verisinin bilgisi Tablo 4.3'te belirtilmiştir.

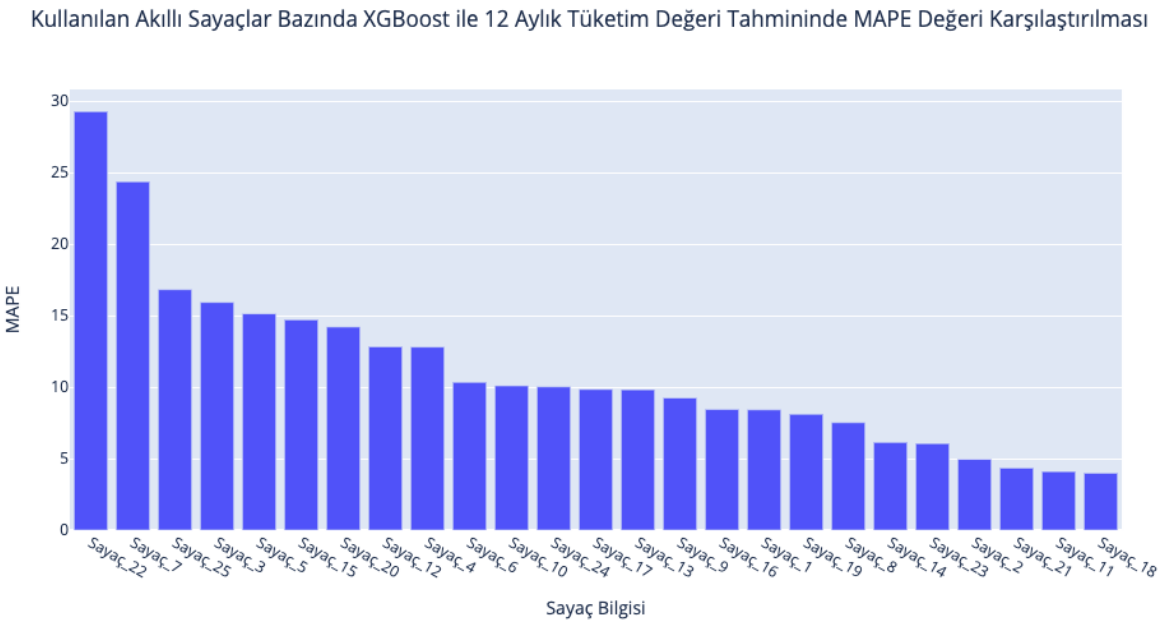
Tablo 4. 3: FbProphet ve XGBoost'nin MAPE değeri ile karşılaştırması.

Algoritma	**Veri Kümesi Kapsamı (sayaç sayısı 25)**	**MAPE (tüm sayaçların ortalaması)**
FbProphet	Her sayacın 102 aylık girdisi var	%8.37
XGBoost	Her sayacın 102 aylık girdisi var	%11.10

Şekil 4.7 ve Şekil 4.8'de gösterilen hata metriği MAPE değeri 25 sayaç için ayrı ayrı hesaplanmıştır. MAPE değerlerinin gösterimi sayaçların toplam 102 aylık değerlerinin işleme alınması ile elde edilmiştir. Görüldüğü gibi FbProphet, XGBoost algoritmasına göre önemli derecede daha başarılı tahminler yapmaktadır.

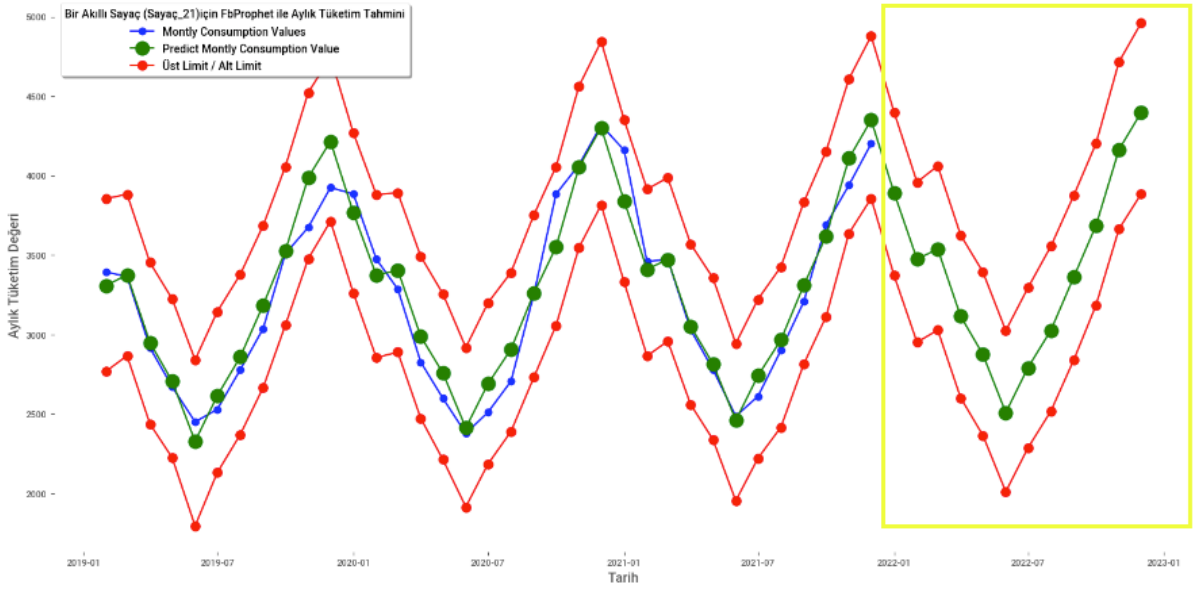


Şekil 4. 7: Kullanılan Akıllı Sayaçlar Bazında FbProphet ile 12 Aylık Tüketim Değeri Tahmininde MAPE Değeri Karşılaştırılması.

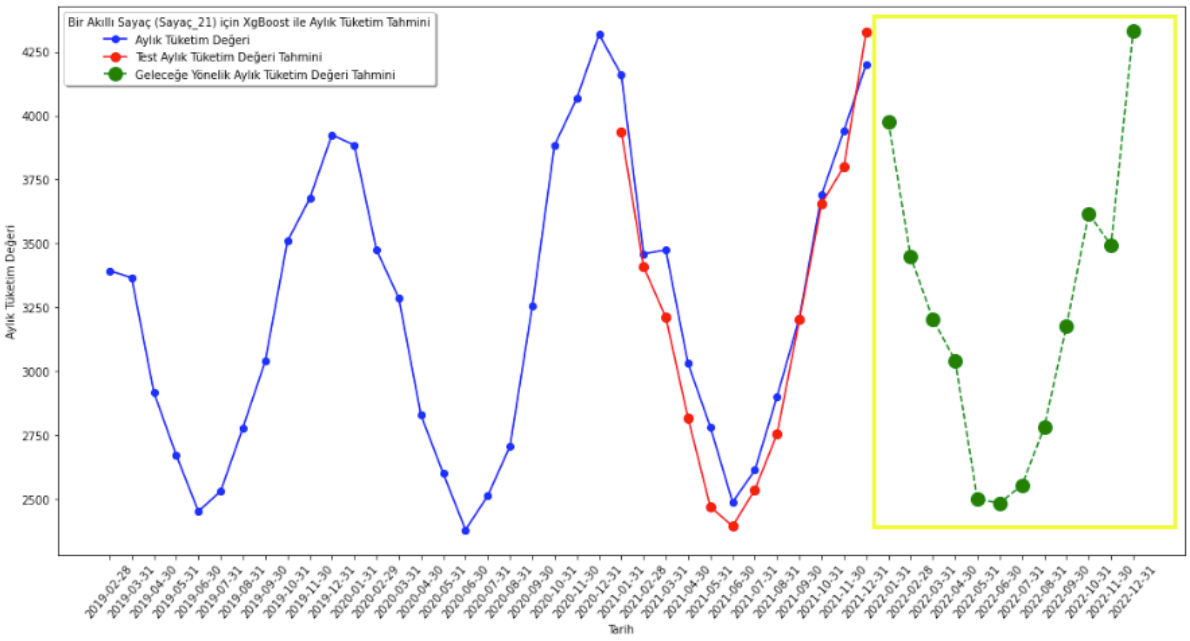


Şekil 4. 8: Kullanılan Akıllı Sayaçlar Bazında XGBoost ile 12 Aylık Tüketim Değeri Tahmininde MAPE Değeri Karşılaştırılması.

Şekil 4.9 ve Şekil 4.10'de FbProphet ile XGBoost algoritmasının bir sayaç için tüketim değerinin tahminleri görselleştirilmiştir. Sarı şeklinde belirtilen değerler her bir algoritmanın ileriye dönük 12 ay için tahmin ettiği tüketim değerini belirtmektedir.



Şekil 4. 9: Bir Akıllı Sayaç için FbProphet ile Aylık Tüketim Tahmini.



Şekil 4. 10: Bir Akıllı Sayaç için XGBoost ile Aylık Tüketim Tahmini.

5. TARTIŞMA

Zaman serileri üzerinde yapılan analizler literatürde önemli bir yer tutmaktadır. Bu analizler, verinin yapısına göre uygun metodun belirlenmesi ile belirli süreçlerden geçerek çeşitli problemlere çözüm getirmektedir. Enerji sektöründe de bu tarz zaman serisi oluşturan veri yapıları mevcuttur. Bu alanda ortaya çıkan elektrik hırsızlığı, dolandırıcılık, kesintiler, sistem arızaları, sayaçların yanlış ölçümü gibi sorunlar dağıtım şirketlerinden son kullanıcılara kadar birçok paydaşı etkilemektedir. Bu sorunların bazıları anomali tespiti problemi olarak ele alınıp uygun çözümler üretilmeye çalışılmıştır. Zaman serisi verilerindeki anomali tespit teknikleri, girdi verilerinin tek değişkenli mi yoksa çok değişkenli mi olduğuna, anomali türünün nokta, altdizi veya zaman serisi olmasına ve uygulanan yöntemin yapısına bağlı olarak değişmektedir. Bu alanda akıllı sayaçların doğru, güvenilir ve ayrıntılı veri üretmesi modellerin performansına büyük kazanımlar getirmektedir. Akıllı sayaçlar ile kullanıcı hareketlerini ve sayaç durumlarını gerçek zamanlı olarak izlemek, tahmin etmek ve daha verimli enerji yönetimi için kararlar almak mümkündür. Literatürde genel olarak bu alanlar özelinde önemli çalışmaların bulunmasına ek olarak bu tez kapsamında getirdiğimiz çözüm önemli bir değer yaratmaktadır.

Bu çalışmada, akıllı sayaçlarla ölçülen aylık tüketim değerlerine dayalı olarak enerji tüketimindeki olası anomalilerin tespit edilmesi ve farklı yöntemler kullanılarak tüketimin tahmin edilmesine odaklanılmıştır. Bu tez çalışması, akıllı sayaç verileri üzerinde çeşitli makine öğrenmesi yöntemlerinin bir analizini içerdiğinden, enerji sektöründe faaliyet gösteren dağıtım şirketlerinin olası finansal kayıpları en aza indirmesine, enerji kayıplarını önlemesine ve diğer olumsuz etkileri azaltmasına yardımcı olmaktadır. Yapılan literatür incelmesinde zaman serisi verileri üzerinde yapılan tahminleme (forecasting) çalışmalarında sıklıkla ARIMA, SARIMA, LSTM, Yapay Sinir Ağları gibi algoritmalar kullanılmıştır. Lakin, burada uygulanacak metodun seçilmesinde probleme çözüm getirme hızı, verinin yapısı gibi önemli birtakım faktörler mevcuttur. Örneğin, ARIMA ile zaman serileri verileri üzerinde iyi sonuç almak için verinin genellikle durağan olması gerekmektedir.

Bu çalışmada zaman serisi içeren sayaç verisinde anomali tespiti için FbProphet kütüphanesi ve scikit-learn kütüphanesinin IF ve LOF algoritmaları uygulanmıştır. Tahminleme için popüler olan yöntemlerden XGBoost ve FbProphet algoritmaları kullanılmıştır. Şekil 4.1'de tüm sürecin akışı özetlenmiştir.

6. SONUÇ VE ÖNERİLER

Bu çalışmada zaman serilerini içeren sayaç verilerinde anomali noktalarının tespit edilmesi için önerdiğimiz FbProphet, IF ve LOF olmak üzere 3 farklı yöntem uygulanmıştır. Bu yöntemlerin her biri uç noktaları başarıyla bulabilmektedir. Anomali tespitinde temel amaç anormal noktaların normal olarak etiketlenmemesidir. Bu süreçde bazı normal noktaların anomali olarak etiketlenmesi sorun teşkil etmemektedir. Bu gibi noktaları bir eşik (threshold) değeri ile söz konusu metotlar kullanılarak filtrelemek mümkündür. IF algoritması başarılı sonuçlar vermesine rağmen hesaplama karmaşıklığı yüksektir ve bu nedenle büyük veri kümeleri için çok önerilmemektedir. Tespit edilen anomalilerle ilgili aksiyon almayı kolaylaştırmak ve doğru önceliklendirme yapabilmek adına her bir anomali için bir önem derecesi ve skor değeri hesaplanmıştır. Önceliklendirmeye göre bir alarm sistemi ve raporlama sistemi kurmak da mümkündür. Burada ileriki çalışmalarda veriyi çevresel faktörlerin etkisi ile zenginleştirip değerlendirmek ya da elde edilen yeni etiketli veri ile Şekil 3.7'deki gibi gelecekteki anomalileri başarılı bir şekilde bulabilmek için sınıflandırma yöntemleri gibi uygun metotları bu çalışmanın çıktısı üzerinde uygulamak mümkün olacaktır.

Bu tez çalışmasında anomalinin yanında ileriye yönelik tahminler de yapılmıştır. Burada popüler olan iki yöntem uygulanmıştır. Kurulan metotların başarısını ölçmek için, literatürde sıklıkla kullanılan hata metriklerinden biri olan MAPE kullanılarak model doğruluğu skorları değerlendirilmiştir. Tüm sayaçların ortalaması olarak FbProphet kütüphanesi MAPE değerini %8 ve XGBoost algoritması da %11 olarak hesaplamıştır. Burada enerji sektöründe genel aydınlatma sayaç verisi olan zaman serileri üzerinde FbProphet yönteminin ortalama olarak en iyi sonucu verdiği tespit edilmiştir.

Sonuç olarak, FbProphet kütüphanesi ve scikit-learn kütüphanesinin IF ve LOF algoritmalarının olası anomali noktalarını başarılı bir şekilde tespit ettiği ve tahminleme için FbProphet'in başarı olarak XGBoost algoritmasından daha iyi sonuçlar verdiği görülmüştür.

KAYNAKLAR

1- Nakayama,K. And Sharma, R., 2017, Energy Management Systems With İntelligent Anomaly Detection And Prediction , Published İn: 2017 Resilience Week (Rws) ,Conference Location: Wilmington, De, Usa ,18-22 September 2017, Publisher: Ieee, Doi: 10.1109/Rweek.2017.8088643

2- Bansal,A., Rompikuntla,S.K., Gopinadhan, J., Kaur, A., Kazi, Z.A., 2015, Energy Consumption Forecasting For Smart Meters , Baı Conference 2015 At Iım Bangalore, India, Arxiv.Org/Abs/1512.05979v1

3- Zhang, J.E., Wu, D., Boulet, B., 2021, Time Series Anomaly Detection For Smart Grids: A Survey,Ieee Canadian Electrical Power And Energy Conference (Epec2021), https://arxiv.org/pdf/2107.08835.pdf

4- Cook, A., Mısırlı, G., And Fan, Z., 2020 , Anomaly Detection For Iot Time-Series Data: A Survey, Published İn: Ieee Internet Of Things Journal (Volume: 7, Issue: 7, July 2020),Issn: 2327-4662,Page(S): 6481 – 6494, Doi: 10.1109/Jıot.2019.2958185

5- Chandel, P., Thakur, T., 2019, Smart Meter Data Analysis For Electricity Theft Detection Using Neural Networks , Adv. Sci. Technol. Eng. Syst. J. 4(4), 161-168 (2019); doi: 10.25046/Aj040420, https://dx.doi.org/10.25046/aj040420

6- Xu, X., Zhao, H., Liu, H., Sun, H., 2020 , Lstm-Gan-Xgboost Based Anomaly Detection Algorithm For Time Series Data , Published İn: 2020 11th International Conference On Prognostics And System Health Management (Phm-2020 Jinan) ,Ieee,23-25 October 2020 Jinan, China ,Isbn:978-1-7281-5181-6, Doi: 10.1109/Phm-Jinan48558.2020.00066

7- Chiosa, R., Piscitelli, M.S. And Capozzoli, A., 2021, A Data Analytics-Based Energy Information System (Eıs) Tool To Perform Meter-Level Anomaly Detection And Diagnosis İn Buildings , Department Of Energy "Galileo Ferraris", Tebe Research Group, Baeda Lab, Politecnico Di Torino, Corso Duca Degli Abruzzi 24, 10129 Turin, Italy,Energies 2021, 14(1), 237; https://doi.org/10.3390/en14010237

8- Lone, H.R., Singh, P., 2018, Detecting Anomalous Energy Consumption In Buildings Using Smart Meter Data, Department Of Computer Science And Engineering Indraprastha Institute Of Information Technology - Delhi New Delhi - 110 020, India December, 2018 , https://repository.iiitd.edu.ın/xmlui/handle/123456789/706

9- Vafeiadis, T., Alexiadis, A., Dimaridou, V., Krinidis, S., Kitsikoudis, K., Makris, L., Davidović, D., Ioannidis, D., Tzovaras, D., 2019, Anomaly Detectıon In Smart Meters- 14. Konferenca Slovenskıh Elektroenergetıkov – Laško 2019 14 Th Conference Of Slovenian Electrical Power Engineers Cıred Šk 6-08

10- Khan, Z.A., 2019, Smart Meter Based Profılıng For Load Forecastıng And Demand Sıde Management In Smart Grıds , University Of Birmingham. Ph.D., School Of Engineering, Department Of Electronic, Electrical And Systems Engineering, Url:http://etheses.bham.ac.uk/id/eprint/9770

11- Shaukat, K., Alam, T.M., Luo, S., Shabbir, S., Hameed, I.A. , Li ,J., Abbas, S.K., Javed, U., 2021, Advances In Intelligent Systems And Computing Book Series (Aisc,Volume 1363),Chapter A Review Of Time-Series Anomaly Detection Techniques: A Step To Future Perspectives. In: Arai, K. (Eds) Advances In Information And Communication. Ficc 2021. Advances In Intelligent Systems And Computing, Vol 1363. Springer, Cham. https://doi.org/10.1007/978-3-030-73100-7_60: 13 Pages

12- Al-Ghaili, M.A., Ibrahim, Z.A., Hairi, S.A.S.S., Rahim, F.A., Baskaran, H., Ariffin, N.A.M., And Kasim, H., 2021, A Review Of Anomaly Detection Techniques In Advanced Metering Infrastructure, Bulletin Of Electrical Engineering And Informaticsvol.10, No.1, February 2021, Pp. 266~273ıssn: 2302-9285, Doi: 10.11591/Eei.V10i1.2026
13- Shah, M.B., 2019, Anomaly Detection In Electricity Demand Time Series Data, Kth, School Of Electrical Engineering And Computer Science (Eecs), P. 41 ,Stockholm, Sweden 2019, Supervisors:Herman And Pawel, Diva, Id: Diva2:1415892
14- Steinborn, F., 2021, Time Series Analysis And Prediction Of Electricity Load Demand With Machine Learning Techniques, University Of Regensburg, https://epub.uni-regensburg.de/51504/1/steinbornf_dissertation_fınal-publıshed.pdf
15- Kardi, M., Alskaif, T., Tekinerdogan B., Catalão J.P. S., 2021, Anomaly Detection İn Electricity Consumption Data Using Deep Learning, 2021 Ieee International Conference On Environment And Electrical Engineering And 2021 Ieee Industrial And Commercial Power Systems Europe (Eeeıc / I&Cps Europe),Isbn:978-1-6654-3613-7,07-10 September 2021 Bari, Italy, ,Doi: 10.1109/Eeeıc/Icpseurope51590.2021.9584650
16- İşyapar, M.T., 2013, Classification Of Electricity Customers Based On Real Consumption Values Using Data Mining And Machine Learning Techniques And Its Corresponding Applications, Mıddle East Technıcal Unıversıty, M.Sc., Department Of Computer Engineering, Supervisor: Prof. Dr. Ferda Nur Alpaslan, http://etd.lib.metu.edu.tr/upload/12616261/ındex.pdf
17- Tajraq, F. G., 2020, Lectricity Consumption Forecasting Of Turkey Usingrecurrent Neural Networks, Supervisor: Dr. Öğr. Üyesi Sefer Baday, Msc, Istanbul Technical University / Informatics Institute / Department Of Informatics Applications / Department Of Information And Communication Engineering, https://tez.yok.gov.tr/ulusaltezmerkezi/tezgoster?key=fl0kw4p1rmmdotykrdyv1orow-ıjacj-7dle38ewdkzrvwaqs6o0nqaxqbx2oxhh
18- Dağ, H., Alamin, M., 2016, Power Consumption Estimation Using In-Memory Database Computation, Ieee,2016 Honet-Ict,Isbn:978-1-5090-3784-1,13-14 October 2016 Nicosia, Cyprus, Doi: 10.1109/Honet.2016.7753443
19- Aksoy, F., 2018, Short-Term Forecasting Of Wind Power Production Using Machine Learning And Deep Learning Methods , Msc, Istanbul Technical University / Institute Of Science And Technology / Department Of Electrical Engineering / Department Of Electrical Engineering, Thesis Advisor: Prof. Dr. Ömer Usta, Https://Tez.Yok.Gov.Tr/Ulusaltezmerkezi/Tezdetay.Jsp?Id=U_Lh_Zgkjtxxypep_N3vaa&No=Qvmyyqjoc7kv_Wvq5naivw
20- Mwansa, M.N., 2020, Modelling And Predicting Energy Usage From Smart Meter Data And Consumer Behaviours İn Residential Houses ,Doctoral Thesis, Liverpool John Moores University. Doi:10.24377/Ljmu.T.00016411,Supervisors:Hurst, W And Yuanyuan, S Urı:https://researchonline.ljmu.ac.uk/id/eprint/16411
21- de Mattos Neto, P.S.G., De Oliveira, J.F.L., Barbosa, L., Alves, E.P., Bassetto, P., Siqueira, H.V., Marinho, M.H.N., Rissi, G.F., Li, F.,2021 Energy Consumption Forecasting For Smart Meters Using Extreme Learning Machine Ensemble. Sensors 2021, 21, 8096. https://doi.org/10.3390/s21238096
22- Resulaj R., 2019, Smart Meter Based Load Forecasting For Residential Customers Using Machine Learning Algorithms, Master's Thesıs ,University Of Stavanger, Faculty Of Science And Technology, Supervisor:Mina Farmanbar, Co-Supervisors:Antorweep Chakravorty, Aida Mehdipourpirbazari, Publisher:University Of Stavanger, ,Series:Masteroppgave/Uıs-Tn-Ide/2019, Uri:http://hdl.handle.net/11250/2620507

23- Maatug,F.,2021,Anomaly Detection Of Smart Meter Data, Master's Thesıs ,University Of Stavanger, Faculty Of Science And Technology Department Of Electrical Engineering And Computer Science, Publisher:Uis, Supervisors: Reggie Davidrajuh Rituka Jaiswal, https://hdl.handle.net/11250/2825119
24- Wen, L., Zhou, K., Yang ,S., Li, L., 2018, Compression Of Smart Meter Big Data: A Survey. Renewable And Sustainable Energy Reviews, 91(), 59–69. doi:10.1016/j.rser.2018.03.088
25- Smart Electricity Meters - Global Market Trajectory & Analytics, Report 342 Pages September 2022 Region: Global Global Industry Analysts, Inc Id: 3301148, https://www.researchandmarkets.com/r/dzpcvw ,[Ziyaret Tarihi: 20.09.2022]
26- The Global Smart Meter Market Size Was Usd 24.53 Billion İn 2018 And İs Projected To Reach Usd 44.18 Billion By 2026, Exhibiting A Cagr Of 7.7% During The Forecast, 2019-2026 , Published On Apr,2020, Based Year:2018, [Ziyaret Tarihi: 20.09.2022] https://www.fortunebusinessinsights.com/smart-meter-market-102666
27- Braei, M., 2019, Anomaly Detection Of Timeseries: A Comparison Of Statistical Vs Classical Machine Learning Vs Deep Learning Approaches , Master Thesis İn The Department Of Computer Science,1. Review: Prof. Max Mühlhäuser,2. Review: Dr. Sebastian Kauschke,Technical University Of Darmstadt, Doi: 10.13140/Rg.2.2.17687.80801
28- Braei M., Wagner S., 2020, Anomaly Detection In Univariate Time-Series: A Survey On The State-Of-The-Art,Cornell University, Https://Doi.Org/10.48550/Arxiv.2004.00433
29- Chandola, V., Banerjee, A., And Kumar, V., 2009, Anomaly Detection: A Survey , Research-Artıcle, Acm Computing Surveys Volume 41 Issue 3 July 2009 Article No.: 15pp 1–58 , https://doi.org/10.1145/1541880.1541882
30- Vitale, F., 2021, Run-Time Anomaly Detection With Process Mining: Methodology And Railway System Compliance Case-Study , Master Thesis, Linnaeus University, Faculty Of Technology, Department Of Computer Science And Media Technology (Cm),Author: Francesco Vitale Supervisor: Francesco Flammini Examiner: Morgan Ericsson.
31- Borges, H., Akbarinia, R., Masseglia, F., 2021, Anomaly Detection İn Time Series. Transactions On Large-Scale Data- And Knowledge-Centered Systems L, Lncs. Tldks - 12930, , Pp.46-62, 2021, Lecture Notes İn Computer Science. Transactions On Large-Scale Data- And Knowledge-Centered Systems, 978-3-662-64553-6. ⟨10.1007/978-3-662-64553-6_3⟩. ⟨Lirmm-03359500⟩
32- Aggarwal, C.C., 2017, Outlier Analysis ,Publisher: Springer Cham,Ebook Isbn: 978-3-319-47578-3, Edition Number:2, Doi https://doi.org/10.1007/978-3-319-47578-3
33- Liu, X., Per Nielsen, S., 2018, Scalable Prediction-Based Online Anomaly Detection For Smart Meter Data, Information Systems Volume 77, September 2018, Pages 34-47, https://doi.org/10.1016/j.is.2018.05.007
34- Liu, X., Iftikhar, N., Nielsen, P.S., Heller, A. ,2016, Online Anomaly Energy Consumption Detection Using Lambda Architecture. In: Madria, S., Hara, T. (Eds) Big Data Analytics And Knowledge Discovery. Dawak 2016. Lecture Notes In Computer Science(), Vol 9829. Springer, Cham. https://doi.org/10.1007/978-3-319-43946-4_13
35- Chou, J.S. And Telaga, A.S., 2014, Real-Time Detection Of Anomalous Power Consumption, Renewable And Sustainable Energy Reviews, Renewable And Sustainable Energy Reviews Volume 33, May 2014, Pages 400-411,Department Of Civil And Construction Engineering, National Taiwan University Of Science And Technology, 43, Sec. 4, Keelung Rd., Taipei, 106, Taiwan, Https://Doi.Org/10.1016/J.Rser.2014.01.088

36- Zhang, Y., Chen, W., And Black, J., 2011, Anomaly Detection In Premise Energy Consumption Data , 2011 Ieee Power And Energy Society General Meeting, Electronic Isbn:978-1-4577-1002-5,24-28 July 2011 Detroit, Mi, Usa, Doi:10.1109/Pes.2011.6039858
37- Jaiswal, R., Chakravorty, A. , Rong, C., 2020, Distributed Fog Computing Architecture For Real-Time Anomaly Detection İn Smart Meter Data, 2020 Ieee Sixth International Conference On Big Data Computing Service And Applications (Bigdataservice),Electronic Isbn:978-1-7281-7022-0,03-06,Oxford,Uk, Doi: 10.1109/bigdataservice49289.2020.00009
38- Saad, A., And Sisworahardjo, N., 2017, Data Analytics-Based Anomaly Detection İn Smart Distribution Network, 2017 International Conference On High Voltage Engineering And Power Systems (Ichveps),Electronic Isbn:978-1-5386-0945-3,02-05 October 2017 Denpasar, Indonesia, Doi: 10.1109/ıchveps.2017.8225855
39- Cardenas, A.A., Amin, S., Schwartz, G., Dong, R., Sastry, S., 2012, A Game Theory Model For Electricity Theft Detection And Privacy-Aware Control İn Amı Systems, Publisher: Ieee,2012 50th Annual Allerton Conference On Communication, Control, And Computing (Allerton),Electronic Isbn:978-1-4673-4539-2,01-05 October 2012monticello, Il, Usa, http://dx.doi.org/10.1109/allerton.2012.6483444
40- Andrysiak,T., Saganowski, L., And Kiedrowski, P., 2017, Anomaly Detection İn Smart Metering Infrastructure With The Use Of Time Series Analysis, Institute Of Telecommunications, Faculty Of Telecommunications And Electrical Engineering, University Of Technology And Life Sciences (Utp), Ul. Kaliskiego 7, 85-789 Bydgoszcz, Poland,Volume 2017 , Article Id 8782131, Https://Doi.Org/10.1155/2017/8782131
41- Zhu, Z., 2022, Anomaly Detection Over Time Series Data, Preprints, 2022070407, doi: 10.20944/preprints202207.0407.v1
42- Goldstein ,M. And Uchida ,S., 2016, A Comparative Evaluation Of Unsupervised Anomaly Detection Algorithms For Multivariate Data. Plos One 11(4): E0152173. Doi:10.1371/Journal. Pone.0152173 Editor: Dongxiao Zhu, Wayne State University, United States,
Https://Journals.Plos.Org/Plosone/Article?Id=10.1371/Journal.Pone.0152173
43- Huang,C.,2018, Featured Anomaly Detection Methods And Applications, Llege Of Engineering, Mathematics And Physical Sciences University Of Exeter,Degree Of Doctor Of Philosophy İn Computer Science,
https://ore.exeter.ac.uk/repository/bitstream/handle/10871/34351/huangc.pdf
44- https://weatherspark.com/y/103262/average-weather-in-ı%c4%9fd%c4%b1r-turkey-year-round#figures-temperature - 2022
45- Falcão,F., Santos, A., Zoppi, T., Fonseca, B., Bondavalli, A., Silva ,C.B.V., Ceccarelli,A., 2019, Quantitative Comparison Of Unsupervised Anomaly Detection Algorithms For Intrusion Detection , Sac '19: Proceedings Of The 34th Acm/Sıgapp Symposium On Applied Computing April 2019 Pages 318–327, 8 April 2019 New York, Ny, United States, Isbn:978-1-4503-5933-7 , https://doi.org/10.1145/3297280.3297314
46- Schmid, S., Wenig, P., Papenbrock, T., 2022, Anomaly Detection In Time Series: A Comprehensive Evaluation, Proceedings Of The Vldb Endowment ,Volume 15, Issue 9, May 2022 ,Pp 1779–1797, https://doi.org/10.14778/3538598.3538602
47- Domingues, R., Filippone, M., Michiardi, P., Zouaoui, J., 2017, A Comparative Evaluation Of Outlier Detection Algorithms: Experiments And Analyses, Department Of Data Science, Eurecom, Sophia Antipolis, France, Pattern Recognition Volume 74, February 2018, Pages 406-421, https://doi.org/10.1016/j.patcog.2017.09.037,

48- Kim, G., Lim,S., and Euom, L., 2022, A Study On Performance Metrics For Anomaly Detection Based On Industrial Control System Operation Data - https://www.mdpi.com/2079-9292/11/8/1213
49- Hwang W.S.; Yun, J.H.; Kim, J.; Kim, H.C. ,2019, Time-Series Aware Precision And Recall For Anomaly Detection: Considering Variety Of Detection Result And Addressing Ambiguous Labeling. In Proceedings Of The 28th Acm International Conference On Information And Knowledge Management, Cikm '19, Beijing, China, 3–7 November 2019; Pp. 2241–2244.
50- Alfares,H.K. And Nazeeruddin, M., 2010, Electric Load Forecasting: Literature Survey And Classification Of Methods. International Journal Of Systems Science , Int. J. Syst. Sci. , Vol. 33, Pp. 23–34, Nov. 2010, Doi: 10.1080/00207720110067421
51- Fang, Z., Yang, S., Lv, C., An, S., Wu, W., 2022, Application Of A Data-Driven Xgboost Model For The Prediction Of Covıd-19 İn The Usa: A Time-Series Studybmj Open 2022;12:E056685. Doi: 10.1136/Bmjopen-2021-056685
52- https://www.toroslaredas.com.tr/yasal-bildirim/aydinlatma-tuketim-bilgileri, [Ziyaret Tarihi: 15.01.2022]
53- https://www.sedas.com/tr-tr/DagitimHizmetleri/Pages/Genel-Aydinlatma.aspx, [Ziyaret Tarihi: 15.01.2022]
54- https://www.firatedas.com.tr/BilgiDanisma/GenelAydinlatmaTutarlari?id=1010, [Ziyaret Tarihi: 15.01.2022]
55- https://www.baskentedas.com.tr/yasal-bildirim/aydinlatma-tuketim-bilgileri, [Ziyaret Tarihi: 15.01.2022]
56- https://www.kcetas.com.tr/genel-aydinlatma/, [Ziyaret Tarihi: 15.01.2022]
57- https://www.cedas.com.tr/tr/sayfalar/206/aydinlatma-tutarlari-ve-komisyon-kararlari, [Ziyaret Tarihi: 15.01.2022]
58- https://www.ayedas.com.tr/yasal-bildirim/aydinlatma-tuketim-bilgileri#,[Ziyaret Tarihi: 15.01.2022]
59- https://www.tedas.gov.tr/#!dagitim_srkt, [Ziyaret Tarihi: 15.01.2022]
60- http://www.tredas.com.tr/icerik/genel-aydinlatma-bilgileri-165 , [Ziyaret Tarihi: 15.01.2022]
61- https://www.ayedas.com.tr/yasal-bildirim/ticari-kalite , [Ziyaret Tarihi: 15.01.2022]
62- https://www.meramedas.com.tr/tr/2022-13.html#detay, [Ziyaret Tarihi: 15.01.2022]
63- https://www.tedas.gov.tr/#!tedas_tarifeler_1 , [Ziyaret Tarihi: 15.01.2022]
64- Dudek, G., 2022 , Std: A Seasonal-Trend-Dispersion Decomposition Of Time Series, Journal Of Latex Class Files, Vol. 14, No. 8, August 2015, https://doi.org/10.48550/arxiv.2204.10398
65- Liu, F.T., Ting, K.M., And Zhou, Z., 2008, Isolation Forest, 2008 Eighth Ieee International Conference On Data Mining,Electronic Issn: 2374-8486,15-19 December 2008 Pisa, Italy, Doi: 10.1109/Icdm.2008.17.
66- Liu, F.T., Ting, K.M. And Zhou, Z., 2012, Isolation-Based Anomaly Detection, Acm Transactions On Knowledge Discovery From Data,Volume 6, Issue 1, March 2012 Article No.: 3, Pp 1–39,Https://Doi.Org/10.1145/2133360.2133363
67- Qin ,Y., And Lou, Y., 2019, Hydrological Time Series Anomaly Pattern Detection Based On Isolation Forest , 2019 Ieee 3rd Information Technology, Networking, Electronic And Automation Control Conference (Itnec),Electronic Isbn:978-1-5386-6243-4,15-17 March 2019 Chengdu, China, Doi: 10.1109/Itnec.2019.8729405
68- Breunig, M.M., Kriegel, H., Ng, R.T., And Sander, J., 2000, Lof: Identifying Density-Based Local Outliers , Acm Sıgmod Record,Volume 29,Issue 2,June 2000, Pp 93–104,https://doi.org/10.1145/335191.335388

69- https://scikit-learn.org/stable/modules/outlier_detection.html -2022
70- Chen, T., Guestrin C., 2016, Xgboost: A Scalable Tree Boosting System, https://doi.org/10.48550/arxiv.1603.02754
71- https://xgboost.readthedocs.ıo/en/stable/ındex.html# , [Ziyaret Tarihi: 20.12.2022]
72- Zhang, L., Bian, W., Qu, W., Tuo, L. And Wang, Y., 2021, Time Series Forecast Of Sales Volume Based On Xgboost, J. Phys.: Conf. Ser. 1873 012067, Doi: 10.1088/1742-6596/1873/1/012067
73- Yu,E., Wei, H., Han, Y., Hu ,P., And Xu, G., 2021, ,Application Of Time Series Prediction Techniques For Coastal Bridge Engineering. Aben 2, https://doi.org/10.1186/s43251-020-00025-4
74- Jha ,B.K., And Pande,S., 2021, Time Series Forecasting Model For Supermarket Sales Using Fb-Prophet , 2021 5th International Conference On Computing Methodologies And Communication (Iccmc),Electronic Isbn:978-1-6654-0360-3 , 08-10 April 2021 Erode, India,, Doi:10.1109/Iccmc51019.2021.9418033
75- Open Source Facebook Fbprophet, 2021, [Ziyaret Tarihi: 20.12.2022], https://facebook.github.io/prophet/docs/quick_start.html#python-api
76- Cheng, Z., Zou, C., & Dong, J. ,2019, Outlier detection using isolation forest and local outlier factor. In Proceedings of the Conference on Research in Adaptive and Convergent Systems, September 2019 ,Pages 161,168, https://doi.org/10.1145/3338840.3355641

EKLER

Ek-1.Yöntemlerin Modellemesindeki Önemli Kod Blokları

Ek-1.1. IF

```
#model=IsolationForest(n_estimators=12,max_samples='auto' ,random_state=random_state
,contamination='auto' )

#model.fit(meter[['Montly_Consumption_Value']])

#meter['Anomaly_Score'] = model.decision_function(meter[['Montly_Consumption_Value']])

#meter['Anomaly'] = model.predict(meter[['Montly_Consumption_Value']])
```

Ek-1.2. LOF

```
#model = LocalOutlierFactor(algorithm='auto', contamination='auto', leaf_size=30,
metric='minkowski', metric_params=None, n_jobs=-1, n_neighbors=12, novelty=False, p=2)

#meter['Anomaly']= model.fit_predict(meter_lof)

#meter['Anomaly_Score']  =  model.negative_outlier_factor_
```

Ek-1.3. FbProphet

```
#model = Prophet(seasonality_prior_scale=0.1,interval_width=0.90,
yearly_seasonality=True,)

#results_pd.loc[results_pd['y'] > results_pd['yhat_upper'], 'anomaly'] = 1

#results_pd.loc[results_pd['y'] < results_pd['yhat_lower'], 'anomaly'] = -1

#results_pd.loc[results_pd['anomaly'] ==1, 'importance'] = \

    ( (results_pd['y'] - results_pd['yhat_upper'])/interval_range)

 #results_pd.loc[results_pd['anomaly'] ==-1, 'importance'] = \

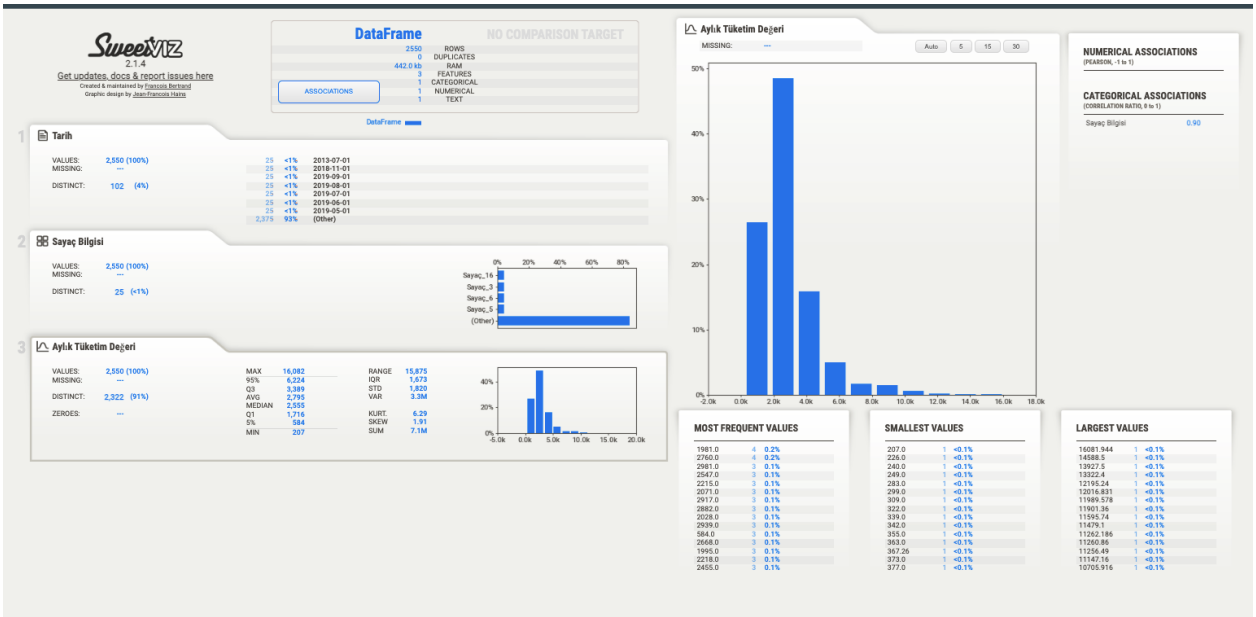
    ( (results_pd['yhat_lower'] - results_pd['y'])/interval_range)
```

Ek-1.4. XGBoost

```
  # model = XGBRegressor(nthread=-1, objective='reg:squarederror',    n_estimators=1000,

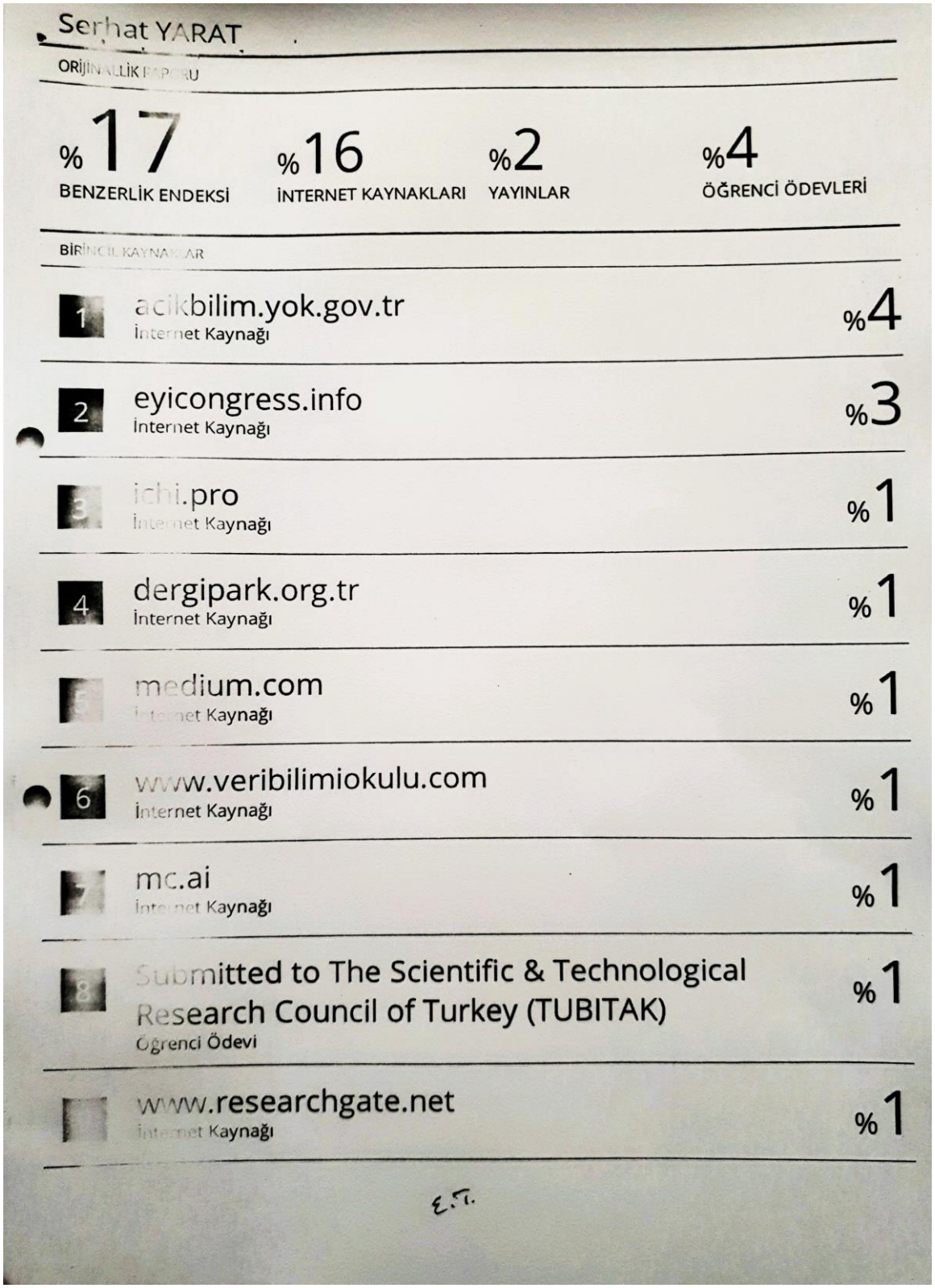
    learning_rate=0.01)
```

Ek-2. Veri Kümesi ile İlgili Metrikler



Ek-2. 1: Veri kümesinin sweetviz_report ile gösterimi.

İNTİHAL RAPORU İLK SAYFASI



Serhat YARAT

ORİJİNALLİK RAPORU

%17	%16	%2	%4
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	acikbilim.yok.gov.tr İnternet Kaynağı	%4
2	eyicongress.info İnternet Kaynağı	%3
3	ichi.pro İnternet Kaynağı	%1
4	dergipark.org.tr İnternet Kaynağı	%1
5	medium.com İnternet Kaynağı	%1
6	www.veribilimiokulu.com İnternet Kaynağı	%1
7	mc.ai İnternet Kaynağı	%1
8	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Öğrenci Ödevi	%1
9	www.researchgate.net İnternet Kaynağı	%1

E.T.