



Araştırma Makalesi

Elektrik Güç Dağıtımında Akıllı Sayaç Verileri için Anomali Tespiti ve Tahminleme

Serhat Yarat¹, Zeynep Orman^{1*}¹*İstanbul Üniversitesi-Cerrahpaşa, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye*

Anahtar Kelimeler:

Anomali Tespiti
Tahminleme
Zaman Serileri
Elektrik Güç Dağıtımı
Akıllı Sayaç

ÖZ

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 gerçekleştirilen 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 elektrik sayaçları ile ölçülen aylık tüketim değerlerine dayalı olarak enerji tüketimindeki olası anomalilerin tespit edilmesi ve farklı makine öğrenmesi yöntemleri kullanılarak gelecek tüketimin tahmin edilmesi amaçlanmıştır. 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.

Anomaly Detection and Prediction for Smart Meter Data in Electrical Power Distribution

Keywords:

Anomaly Detection
Forecasting
Time Series
Electrical Power
Distribution
Smart Meters

ABSTRACT

With rapidly increasing energy demand as a result of increasing population density and economic growth, energy and electricity grids are now facing more challenges. In the face of this demand, 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 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 users. Identifying such anomalies in the energy sector, data analytics studies, and big data technologies play an important role to produce solutions. 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. As a result; the experiments and observations on general lighting meters showed that Isolation Forest (IF), 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 time series, FbProphet outperforms XGBoost.

*Sorumlu Yazar

*(ormanz@iuc.edu.tr) ORCID ID 0000-0002-0205-4198
(serhat.yarat@gtech.com.tr,yarat.serhat@gmail.com) ORCID ID 0000 - 0003 - 1531 - 3790

e-ISSN: 2717-8579

Geliş Tarihi: 24/12/2022; Kabul Tarihi: 24/02/2023

Bilgisayar Bilimleri ve Teknolojileri Dergisi

1. GİRİŞ

Enerjiye bağımlılığın artmasıyla birlikte zaman içerisinde elektrik üretim ve dağıtım sistemleri yüksek teknolojiye 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) şeklinde çözümler geliştirilebilir (Nakayama ve Sharma,2017).

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 yöntemlere bağlı olarak verilerde çeşitli faktörlerin herhangi birinden veya birkaçından etkilenen birçok model oluşabilmektedir. İnsan yaşamının 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 (Bansal vd, 2015).

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, sistemlerinin arızalara ve siber saldırılara karşı daha duyarlı olması gibi bazı dezavantajları da vardır (Zhang vd, 2021).

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. Nesnelere interneti (Internet of things - IoT) dünyasında bu kavramın önem kazanmasıyla birlikte birçok anomali tespit yönteminin 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 (Cook vd, 2020). 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 (Shaukat, 2021).

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 (Vafeiadis vd, 2019). 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. 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 (Al-Ghaili vd, 2021). Ayrıca, mevcut çalışmalar tarafından kullanılan çeşitli ölçüm yöntemleri ve yaklaşımları ayrıntılı inceleyerek çalışmaları değerlendirmişlerdir.

Enerji alanında sınıflandırma, anomali tespiti ve tahmini için Türkiye’de yapılan bazı çalışmalar da bulunmaktadır. İşyapar, çalışmasında Türkiye’de faaliyet gösteren bir enerji dağıtım şirketinin abonelerine ait verilerin makine öğrenmesi teknikleri kullanılarak sınıflandırılmasını ele almıştır (İşyapar,2013). Tajraq, Özyinelemeli Sinir Ağı (Recurrent Neural Network - RNN), uzun kısa süreli bellek (Long Short-Term Memory - LSTM) ve Kapı Özyinelemeli Geçitler (Gated Recurrent Units - GRU) yöntemlerini kullanarak Türkiye’deki enerji tüketim verileri üzerinde tahminler yaparak bu modellerin performanslarını karşılaştırmıştır (Tajraq, 2020). 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.

Resulaj vd., Norveç’teki elli beş hanede kısa vadeli yük tahmini için makine öğrenimi algoritmalarını kullanmışlardır (Resulaj, 2019). 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.

Maatug, bazı bina ve evlere ait olan akıllı sayaç verilerinin güç tüketimindeki anomalileri FbProphet ile tespit etmiştir (Maatug, 2021). 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 araştırma, akıllı sayaç zaman serisi verilerinde anomali tespiti ve tahminleme için popülerliği artan yöntemlere odaklanarak enerji tüketiminde önlemler alabilmemize olanak sağlayacak çıktılar üretmektedir. Bu amaçla, genel aydınlatma sayaçlarıyla ölçülen aylık elektrik tüketim değerleri üzerinde olası anomali tespiti ve tüketim tahmini için İzolasyon Ormanı (Isolation Forest -IF), Aykırı Değer Faktörü (Local Outlier Factor-LOF), FbProphet ve Ekstrem Gradyan Arttırma (Extreme Gradient Boosting - XGBOOST) yöntemlerinin uygulanmasında elde edilen sonuçların karşılaştırılması ve hangi yöntemin daha başarılı olduğu gibi sorulara cevap arayarak probleme çözüm getirmeyi amaçlamaktadır.

2. YÖNTEM

2.1. Zaman Serileri

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.

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 (Braei, 2019), (Wagner ve Braei, 2020).

2.2. Anomaliler ve Anomali Tespiti

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 (Wagner ve Braei, 2020), (Chandola , 2009), (Vitale, 2021). Genel olarak anomalilerle ilgili iki ana özellik öne çıkmaktadır:

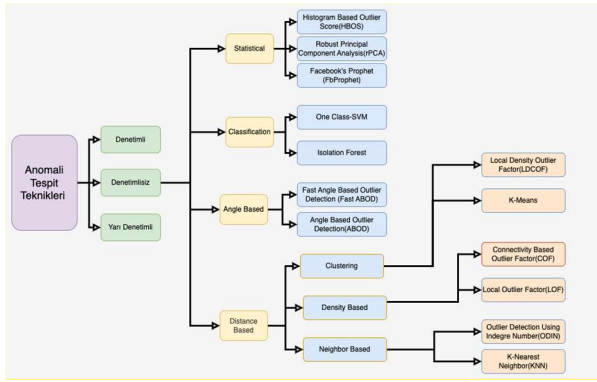
- Anomalilerin dağılımı, verilerin genel dağılımından önemli ölçüde sapmaktadır.
- Veri kümesinin büyük çoğunluğu normal veri noktalarından oluşmaktadır. Anomaliler, veri kümesinin yalnızca çok küçük bir bölümünü oluşturur.

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 kaçırmamanın 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 (Braei ,2019), (Wagner ve Braei,2020), (Borges,2021):

1. *Nokta Anomalileri (Point Anomalies)*: Bir nokta verisi içinde bulunduğu veri kümesinden önemli ölçüde sapsa, nokta anomalisi olarak kabul edilir. Örneğin, diğer işlemlerden farklı olarak büyük bir kredi işlemi nokta anomalisidir. Bu nedenle, değeri $[X_{t-k}, X_{t+k}]$, $k \in \mathbb{R}$ 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: Örneğin, 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.

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.

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.



Şekil 1. Bazı anomali tespiti algoritmaları (Goldstein ve Uchida 2016), (Falcão vd, 2019), (URL-15)

Şekil 1'de bazı anomali tespiti yöntemlerinin bir sınıflandırması yapılmıştır. Bu sınıflandırmayı farklı metrikler için de yapmak mümkündür. Bazı algoritmaların sınıflandırılması birden fazla yöntemle gerçekleştirilebilir. Örneğin LOF algoritmasının denetimli yapısı da vardır. Schmid vd., zaman serilerindeki anomalileri tespit eden teknikler için 158 kaynaktaki kullanımı inceleyerek sınıflandırmışlardır (Schmid vd, 2022).

Domingues vd., anomali tespitinde denetimsiz makine öğrenimi algoritmalarının kullanımına yönelik bir özet çalışma yapmışlardır (Domingues vd, 2017). Anomali tespiti bağlamında, sentetik ve gerçek veri kümeleri üzerinde 14 algoritmanın ortalama kesinlik, sağlamlık, hesaplama süresi ve bellek kullanımı karşılaştırmışlardır.

Bu çalışmada anomali tespiti için tahmine dayalı algoritmalar olan FbProphet, iyi bilinen ve yaygın kullanılan anomali tespiti algoritmaları olan LOF ve IF kullanılmıştır.

2.3. Zaman Serilerinde Tahmin Yaklaşımları

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 aşamalarda 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 (Al-Ghaili vd, 2021), (Resulaj, 2019), (Alfares ve Nazeeruddin, 2010):

- **Kısa vadeli tahmin**, bir hafta öncesine kadar bir gün tahmin ufkuна 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 1 gün ile 1 yıl arası tahminler yapar.
- **Uzun vadeli tahmin**, gelecekte birkaç yıllık bir tahmin ufkuна 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.

2.3.1. 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 (Fang vd, 2022):

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

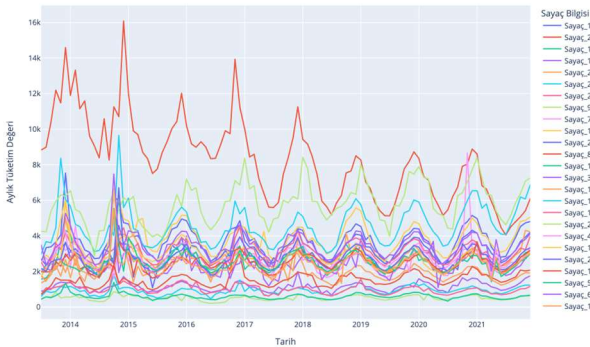
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

(1)-(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.

2.4. Veri Kümesi

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 (Braei ve Wagner 2020). Bu çalışmada kullanılan veri kümesi elektrik dağıtım şirketlerin kendi sitelerinde paylaştıkları verilerden yararlanarak oluşturulmuş ve anonimleştirilmiştir (URL-1) - (URL-12). Elde edilen örnek veri kümesi toplam 8 yıllık 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.



Şekil 2. Kullanılan veri kümesinin grafiksel dağılımı

Şekil 2 de görüldüğü gibi veri kümesi mevsimsellik içermektedir ve Tablo 1'deki gibi bir yapısı vardır.

Tablo 1. Bir Akıllı sayaca ait aylık tüketim veri yapısı

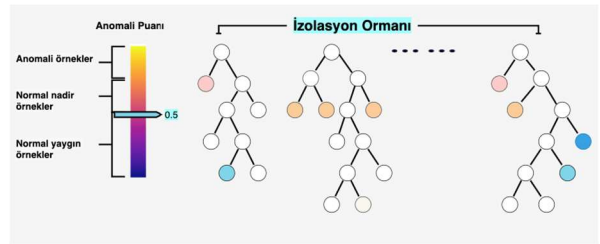
Tarih	Sayaç Bilgisi	Aylık Tüketim Değeri
2013-07-31	Sayaç_1	1748.000

2013-07-31	Sayaç_2	1430.000
2013-08-31	Sayaç_1	4258.180
2013-08-31	Sayaç_2	4232.020

2.5. Kullanılan Algoritmalar

2.5.1. 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 (Falcão vd, 2019), (Liu vd, 2008),(Liu vd, 2012). 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 edecek şekilde bölmeye çalışır. Genellikle anomaliler veri noktaları kümesinden uzaktadır, bu nedenle normal veri noktalarına kıyasla anomalileri yalıtım 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'te genel IF algoritmasının uygulanma biçimi ve anomali skorunun tespiti gösterilmiştir.



Şekil 3. İzolasyon ormanı (Qin ve Lou, 2019)

2.5.2. 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 (Breunig vd, 2000). LOF bir noktanın anomali olup olmadığına karar vermek için veri noktalarının yoğunluğunu dikkate alan başka bir anomali tespit tekniğidir. Yerel anomalileri bulmak da başarılıdır (Falcão vd, 2019). 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 (Breunig vd, 2000). 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 bulunan LOF ile anomali tespitini, Aykırı değer (outlier) tespiti ve yenilik tespiti (novelty detection) olmak üzere iki çatı altında ele almaktadır (URL-13). Outlier 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. LOF ile yapılan yenilik tespiti (novelty detection) bu çalışmada kapsam dışı bırakılmıştır. Anomali tespitinde neighbors.LocalOutlierFactor davranışı Tablo 2'de özetlenmiştir.

Tablo 2. neighbors.LocalOutlierFactor davranışı (URL-13).

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

2.5.3. 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 (Chen ve Guestrin, 2016), (URL-14). 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 algoritmalarından daha hızlı çalışmakta olduğunu belirtmektedirler. Literatürde kullanımı oldukça fazladır (Fang vd, 2022) , (Yu vd, 2021) vb.

2.5.4. 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 (Jha ve Pande, 2021), (URL-15). 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 (Oto regresif Hareketli Ortalama (Autoregressive Moving Average Model - 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.

2.6. Anomali Tespiti ve Tahminleme

Anomali algılama ile tahminleme yöntemleri anomali tespiti ve tahminleme için aynı çatı altında değerlendirilebileceği gibi ayrı konseptler olarak da veri kümesi üzerinde farklı metotlar ile uygulanabilir. Bu çalışma özelinde anomali tespiti için, FbProphet 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 FbProphet yöntemleri ile elde edilen tahminleme sonuçları MAPE metriklerine göre karşılaştırılarak değerlendirilmiştir.

2.6.1. Anomali tespitinde kullanılan yapılar

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

2.6.1.1. 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 değer 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 kullanımını ile 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. 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.

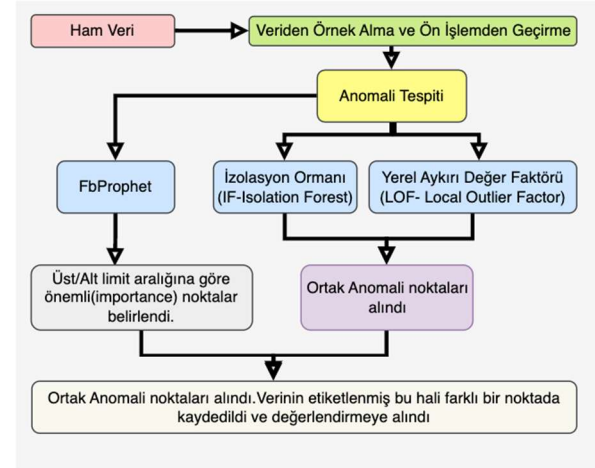
2.6.1.2. İzolasyon ormanı (IF) ve yerel aykırı değer faktörü (LOF) uygulanması

Bu yöntemler ayrı ayrı değerlendirilebileceği 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 (Cheng vd ,2019). 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 bulunduğu anomali noktaları karşılaştırılarak ortak olanları alınmıştır. Uygulanan algoritmanın bulunduğu 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 değeri (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 4'te özetlenmiştir.



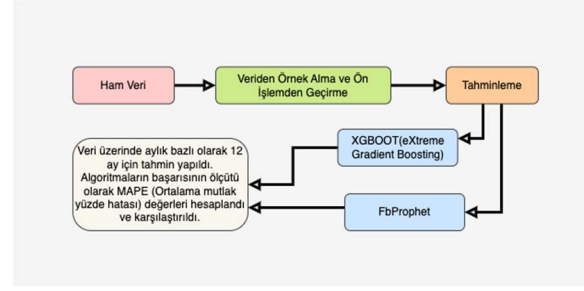
Şekil 4. Çalışmadaki anomali tespiti akışı

2.6.2. Tahmiledede (forecasting) kullanılan yapılar

Çalışma kapsamında yapılan tahminlemenin ileriye dönük olması hedeflenmiştir. Her sayac 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 aylık veri kümesinden 12 aylık kısmı test için kullanılmıştır. Her sayaç üzerinden ayrı ayrı modelleme yapılarak MAPE çıktısına göre hiper parametreler belirlenmiştir. Bu algoritma ile FbProphet'a benzer ileriye dönük her sayaç için bir aylık değer tahmin edilmiştir. Bu iki algoritmanın başarımlarını karşılaştırılması MAPE değerleri üzerinden yorumlanmıştır. Bu yapıların uygulanmasının akışı şekil 5'te özetlenmiştir.

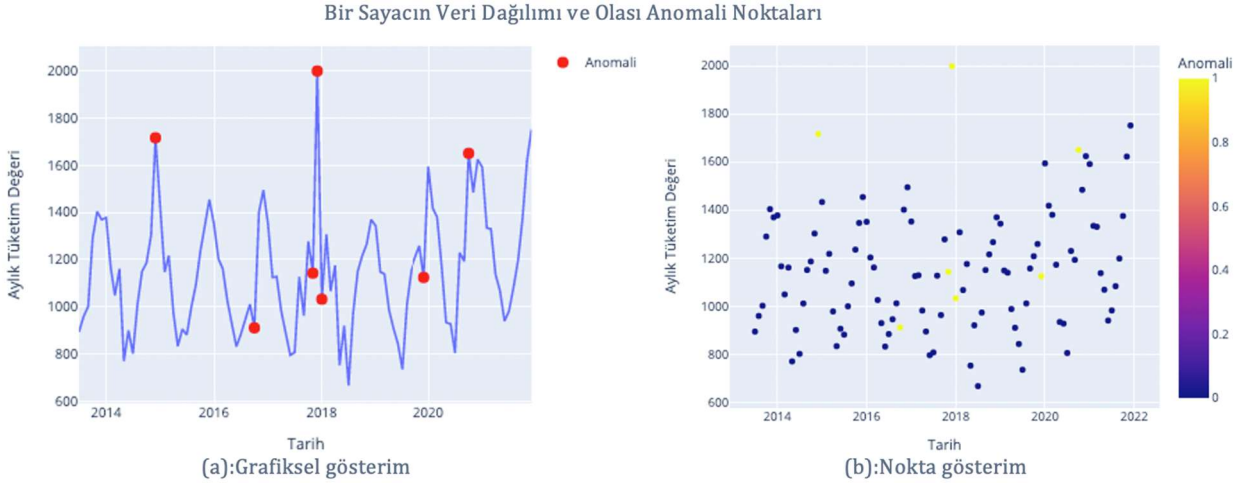


Şekil 5. Bu çalışmada kullanılan veri üzerinde tahminleme akışı

3. BULGULAR

3.1. Anomali Tespiti ile İlgili Bulgular

Bu çalışmada her sayaç için FbProphet, LOF ve IF ile ayrı ayrı anomali tespiti çalışması yapılmıştır. Genel bir sayaca ait *Scikit-learn* altında bulunan *LOF* verilerin dağılımı ve olası anomali noktalarını belirten grafik Şekil 6'da gösterilmiştir.



Şekil 6. Bir sayaca ait veri dağılımı ve olası anomali noktaları

Tablo 3'te anomali tespiti için bu çalışmada kullanılan algoritmaların hiper parametrelerin listesi gösterilmiştir.

Tablo 3. Anomali için kullanılan algoritmaların hiper parametre listesi

Algoritma	Parametre	Atanan Değer
LOF	k	12
	distance function	minkowski
	Kontaminasyon (Contamination) değeri	auto
	novelty	False
IF	iTrees Sayısı	12
	Kontaminasyon (Contamination) değeri	auto
	n_estimators	12
FbProphet	interval_width	0.90
	yearly_seasonality	True
	weekly_seasonality	False
	daily_seasonality	False
	seasonality_prior_scale	0.1
	freq	MS
	periods	1

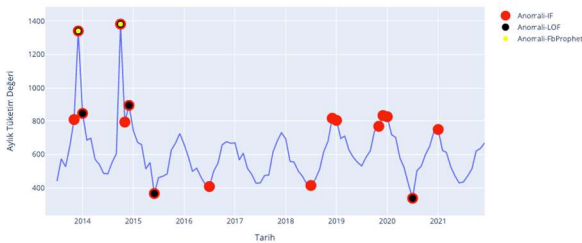
Tablo 4'te görüldüğü gibi her bir yöntem, farklı sayıda anomali noktası belirlemiştir. LOF ve IF algoritma yaklaşımlarına göre anomali olarak belirlenen noktalara skor(score) yapısını da sunmaktadır.

Tablo 4. Kullanılan yaklaşımların 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

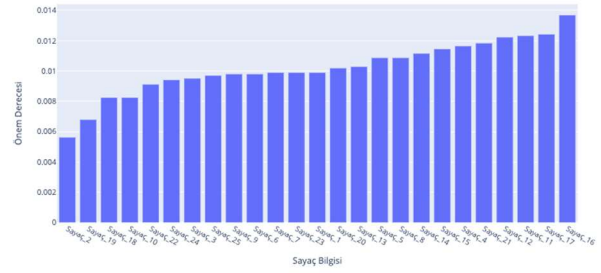
Şekil 7'de görüldüğü üzere LOF, Isolation Forest ve FbProphet algoritmalarıyla örnek aynı sayaç numarası (Sayaç_13) ile bakıldığında benzer şekilde olası üç anomali noktalarının tespit edilebildikleri gözlemlenmiştir. LOF ve IF çıktıları skor sonuçları ile değerlendirilebileceği gibi FbProphet yöntemi ile de tespit edilen noktaların önem derecesi Şekil 8'deki gibi belirlenebilir.

Bir akıllı sayaç (sayaç_13) için FbProphet/LOF/IF ile aylık tüketim değeri bazında anomali tespiti



Şekil 7. Bir akıllı sayaç (sayaç_13) için FbProphet/LOF/IF ile aylık tüketim değeri bazında anomali tespiti

Kullanılan Akıllı Sayaçların FbProphet ile Ortalama Olarak Anomali Önemi



Şekil 8. Kullanılan akıllı sayaçların FbProphet ile ortalama olarak anomali önemi

3.2. Tahminleme ile ilgili Bulgular

FbProphet ile XGBoost algoritmalarının uygulaması sonucunda hesaplanan MAPE değerleri ve uygulanan sayaç verisinin bilgisi Tablo 5'te belirtilmiştir.

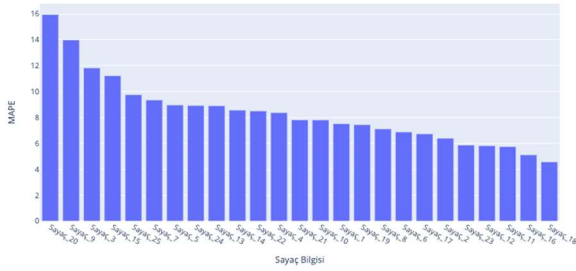
Tablo 5. FbProphet ve XGBoost'nin MAPE değeri ile karşılaştırması

Algoritma	Veri Kümesi Kapsamı (sayaç sayısı 25)	MAPE (Tüm sayaçların ortalaması)
FbProphet	Her sayaçın 102 aylık girdisi var	%8.37

XGBoost	Her sayacın 102 aylık girdisi var	%11.10
---------	-----------------------------------	--------

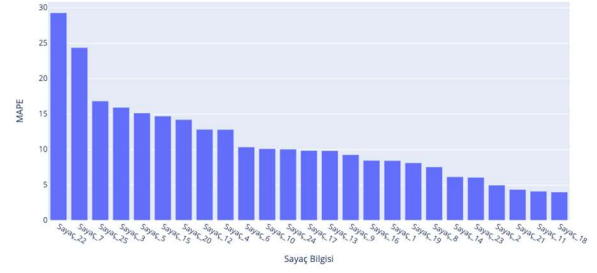
Şekil 9 ve Şekil 10'da gösterilen hata metriği MAPE değeri 25 sayaç için ayrı ayrı hesaplanmıştır. MAPE değerlerinin gösterimi sayaçların toplam 102 aylık değerlerinin işleme alınması ile elde edilmiştir. Görüldüğü gibi FbProphet, XGBoost algoritmasına göre önemli derecede daha başarılı tahminler yapmaktadır.

Kullanılan Akıllı Sayaçlar Bazında FbProphet ile 12 Aylık Tüketim Değeri Tahmininde MAPE Değeri Karşılaştırılması



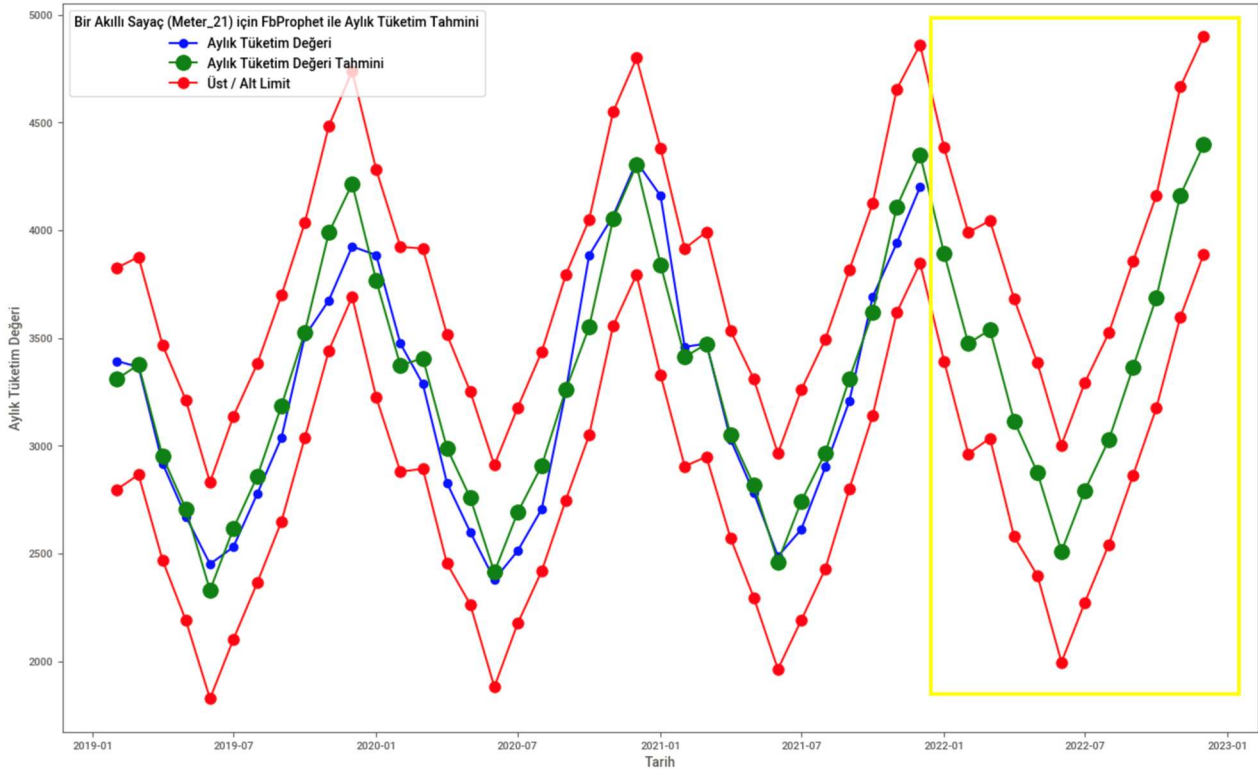
Şekil 9. Kullanılan akıllı sayaçlar bazında FbPorphet ile 12 aylık tüketim değeri tahmininde MAPE değeri karşılaştırılması

Kullanılan Akıllı Sayaçlar Bazında XGBoost ile 12 Aylık Tüketim Değeri Tahmininde MAPE Değeri Karşılaştırılması

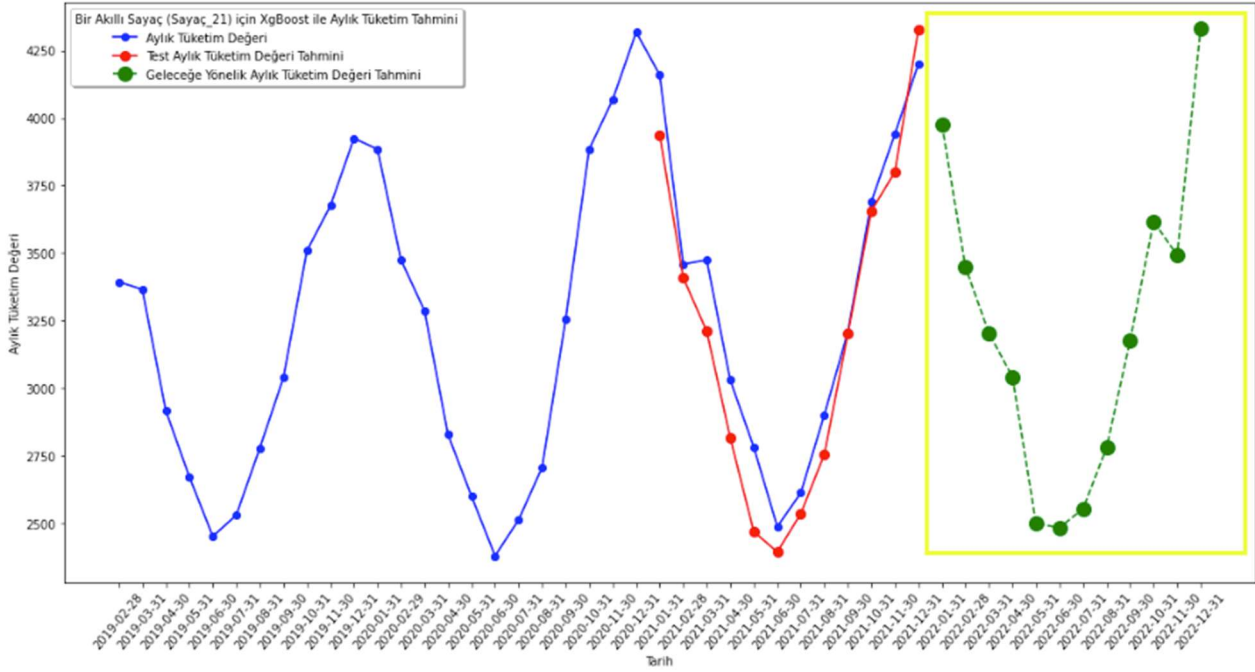


Şekil 10. Kullanılan akıllı sayaçlar bazında XGBoost ile 12 aylık tüketim değeri tahmininde MAPE değeri karşılaştırılması

Şekil 11 ve Şekil 12'de FbProphet ile XGBoost algoritmasının bir sayaç için tüketim değerinin tahminleri görselleştirilmiştir. Sarı şekilde belirtilen değerler her bir algoritmanın ileriye dönük 12 ay için tahmin ettiği tüketim değerini belirtmektedir. Her iki metot ile elde edilen tahmin değerlerin karşılaştırılması Tablo 6'da yapılmıştır.



Şekil 11. Bir akıllı sayaç için fbprophet ile aylık tüketim tahmini



Şekil 12. Bir akıllı sayaç için xgboost ile aylık tüketim tahmini

Tablo 6. FbProphet ve XGBoost ile elde edilen tahmin değerleri karşılaştırılması

Tarih	FbProphet ile Aylık Tüketim Değeri Tahmini	XGBoost ile Aylık Tüketim Değeri Tahmini
2022-01-31	3891.9	3976.0
2022-02-31	3474.5	3448.0
2022-03-31	3538.3	3203.0
2022-04-31	3115.3	3041.0
2022-05-31	2875.8	2498.0
2022-06-31	2509.0	2483.0
2022-07-31	2792.0	2553.0
2022-08-31	3027.3	2779.0
2022-09-31	3362.7	3175.0
2022-10-31	3686.7	3617.0
2022-11-31	4163.3	3492.0
2022-12-31	4397.4	4333.0

4. TARTIŞMA VE SONUÇLAR

Bu çalışmada, akıllı sayaçlarla ölçülen aylık tüketim değerlerine dayalı olarak enerji tüketimindeki olası anomalilerin tespit edilmesi ve farklı makine öğrenmesi yöntemleri kullanılarak tüketimin tahmin edilmesine odaklanılmıştır. Çalışmada önerilen model ile enerji sektöründe

faaliyet gösteren dağıtım şirketlerinin olası finansal kayıpları en aza indirmesine, enerji kayıplarını önlemesine ve diğer olumsuz etkileri azaltmasına yardımcı olunması amaçlanmıştır.

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 süreçte 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 kurulması da mümkündür. Burada ileriki çalışmalarda veriyi çevresel faktörlerin etkisi ile zenginleştirip değerlendirmek ya da elde edilen yeni etiketli veri ile gelecekteki anomalileri başarılı bir şekilde bulabilmek için sınıflandırma yöntemleri gibi uygun metotları bu çalışmanın çıktısı üzerinde uygulamak mümkün olacaktır.

Çalışmada 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.

BİLGİLENDİRME/TEŞEKKÜR

Bu çalışma süreci boyunca gösterdikleri yardım ve desteklerden dolayı GTech (G Teknoloji Bilişim San. Tic. A.Ş.) şirketindeki Büyük Veri ve Analitik ekibinin değerli desteklerine teşekkür ederiz.

KAYNAKÇA

- Alfares, H.K. ve Nazeeruddin, M. (2010). Electric Load Forecasting: Literature Survey and Classification of Methods, *International Journal of Systems Science*, 33, 23-34. Doi: 10.1080/00207720110067421
- Al-Ghaili, M.A., Ibrahim, Z.A., Hairi, S.A.S.S., Rahim, F.A., Baskaran, H., Ariffin, N.A.M., ve Kasim, H. (2021). A Review of Anomaly Detection Techniques in Advanced Metering Infrastructure, *Bulletin of Electrical Engineering and Informatics*, 10(1), 266-273. Doi: 10.11591/eei.v10i1.2026
- Breunig, M.M., Kriegel, H., Ng, R.T., ve Sander, J. (2000). Lof: Identifying Density-Based Local Outliers, *Acm Sigmod Record*, 29(2), 93-104. <https://doi.org/10.1145/335191.335388>
- Bansal, A., Rompikuntla, S.K., Gopinadhan, J., Kaur, A., ve Kazi, Z.A. (2015). Energy Consumption Forecasting For Smart Meters, ArXiv abs/1512.05979, Cornell university. Elsevier.
- Braei, M. (2019). Anomaly Detection of Time series: A Comparison of Statistical vs Classical Machine Learning vs Deep Learning Approaches, *Master Thesis in The Department of Computer Science, Kauschke, Technical University Of Darmstadt*. Doi: 10.13140/rg.2.2.17687.80801
- Braei M., ve 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>
- Borges, H., Akbarinia, R., ve Massegia, F. (2021). Anomaly Detection in Time Series, *Transactions On Large-Scale Data-and Knowledge-Centered Systems L, Lecture Notes in Computer Science*, 12930, 46-62.
- Cook, A., Mısırlı, G., ve Fan, Z. (2020). Anomaly Detection For Iot Time-Series Data: A Survey, *IEEE Internet Of Things Journal*, 7 (7), 6481 - 6494. Doi: 10.1109/jiot.2019.2958185
- Chandola, V., Banerjee, A., ve Kumar, V. (2009). Anomaly Detection: A Survey, *Acm Computing Surveys*, 41(3), 1-58. <https://doi.org/10.1145/1541880.1541882>
- Chen, T., ve Guestrin C. (2016). Xgboost: A Scalable Tree Boosting System, *arXiv:1603.02754*.
- Cheng, Z., Zou, C., ve Dong, J. (2019). Outlier detection using isolation forest and local outlier factor. In *Proceedings of the Conference on Research in Adaptive and Convergent Systems*, 161-168. <https://doi.org/10.1145/3338840.3355641>
- Domingues, R., Filippone, M., Michiardi, P., ve Zouaoui, J. (2017). A Comparative Evaluation Of Outlier Detection Algorithms: Experiments And Analyses, *Pattern Recognition*, 74, 406-421. <https://doi.org/10.1016/j.patcog.2017.09.037>
- Fang, Z., Yang, S., Lv, C., An, S., ve Wu, W. (2022). Application of a Data-Driven Xgboost Model For The Prediction of Covid-19 in The Usa: A Time-Series Study, *BMJ Open*, 12:e056685. Doi: 10.1136/Bmjopen-2021-056685
- Falcão, F., Santos, A., Zoppi, T., Fonseca, B., Bondavalli, A., Silva ,C.B.V., ve Ceccarelli,A. (2019). Quantitative Comparison of Unsupervised Anomaly Detection Algorithms For Intrusion Detection, *Sac '19: Proceedings Of The 34th Acm/Sigapp Symposium On Applied Computing*, 318-327.
- Goldstein, M. ve Uchida, S. (2016). A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data, *Plos One*, 11(4), E0152173.

- İşyapar, M.T. (2013). Classification of Electricity Customers Based On Real Consumption Values Using Data Mining And Machine Learning Techniques And Its Corresponding Applications, *M.Sc. thesis*, Middle East Technical University, M.Sc., Department Of Computer Engineering.
<http://etd.lib.metu.edu.tr/upload/12616261/index.pdf>.
- Jha, B.K., ve Pande, S. (2021). Time Series Forecasting Model For Supermarket Sales Using Fb-Prophet, *5th International Conference On Computing Methodologies And Communication*, Isbn:978-1-6654-0360-3.
- Liu, F.T., Ting, K.M., ve Zhou, Z. (2008). Isolation Forest, *IEEE International Conference On Data Mining*, Electronic Issn: 2374-8486,15-19. doi: 10.1109/icdm.2008.17
- Liu, F.T., Ting, K.M. ve Zhou, Z. (2012). Isolation-Based Anomaly Detection, *Acm Transactions On Knowledge Discovery From Data*, 6(1), 1–39. <https://doi.org/10.1145/2133360.2133363>
- Maatug, F. (2021). Anomaly Detection Of Smart Meter Data, *Master's Thesis, University of Stavanger, Faculty of Science And Technology Department of Electrical Engineering and Computer Science*.
- Nakayama, K. ve Sharma, R. (2017). Energy Management Systems with Intelligent Anomaly Detection and Prediction, *Published in: 2017 Resilience Week (Rws)*, 24-29.
- Resulaj R. (2019). Smart Meter Based Load Forecasting For Residential Customers Using Machine Learning Algorithms, *Master's Thesis*, University of Stavanger, Faculty of Science and Technology.
- Schmid, S., Wenig, P., and Papenbrock, T. (2022). Anomaly Detection in Time Series: A Comprehensive Evaluation, *Proceedings of the Vldb Endowment*, 15(9), 1779–1797. <https://doi.org/10.14778/3538598.3538602>
- Shaukat, K., Alam, T.M., Luo, S., Shabbir, S., Hameed, I.A., Li, J., Abbas, S.K., and Javed, U. (2021). A Review of Time-Series Anomaly Detection Techniques: A Step To Future Perspectives. In: *Arai, K. (Eds) Advances In Information And Communication*. 1363, Springer, Cham. 13 pages.
- https://doi.org/10.1007/978-3-030-73100-7_60
- Tajraq, F.G. (2020). Electricity Consumption Forecasting Of Turkey Using recurrent Neural Networks, *Master Thesis*, Istanbul Technical University, Informatics Institute, Department of Informatics Applications.
- 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.
- Vafeiadis, T., Alexiadis, A., Dimaridou, V., Krinidis, S., Kitsikoudis, K., Makris, L., Davidović, D., Ioannidis, D., ve Tzovaras, D. (2019). Anomaly Detection In Smart Meters, *14 Th Conference Of Slovenian Electrical Power Engineers*, 6-08.
- Zhang, J.E., Wu, D., ve Boulet, B. (2021). Time Series Anomaly Detection For Smart Grids: A Survey, *Ieee Canadian Electrical Power And Energy Conference (Epec2021)*.
- Zhu, Z. (2022). Anomaly Detection over Time Series Data, Preprints, 2022070407, Doi: 10.20944/preprints202207.0407.v1.
- Yu,E., Wei, H., Han, Y., Hu ,P., ve Xu, G. (2021). Application Of Time Series Prediction Techniques For Coastal Bridge Engineering, *Advances in Bridge Engineering*. 6. <https://doi.org/10.1186/s43251-020-00025-4>
- Qin ,Y., ve Lou, Y. (2019). Hydrological Time Series Anomaly Pattern Detection Based on Isolation Forest, *IEEE 3rd Information Technology, Networking, Electronic And Automation Control Conference*, 15-17.
- URL-1: <https://www.toroslaredas.com.tr/yasal-bildirim/aydinlatma-tuketim-bilgileri>
[Erişim Tarihi: 15.01.2022]
- URL-2: <https://www.sedas.com/tr-tr/DagitimHizmetleri/Pages/Genel-Aydinlatma.aspx>
[Erişim Tarihi: 15.01.2022]
- URL-3: <https://www.firatedas.com.tr/BilgiDanisma/GenelAydinlatmaTutarlari?id=1010>
[Erişim Tarihi: 15.01.2022]

- URL-4: <https://www.baskentedas.com.tr/yasal-bildirim/aydinlatma-tuketim-bilgileri>
[Erişim Tarihi: 15.01.2022]
- URL-5: <https://www.kcetas.com.tr/genel-aydinlatma/>
[Erişim Tarihi: 15.01.2022]
- URL-6: <https://www.cedas.com.tr/tr/sayfalar/206/aydinlatma-tutarlari-ve-komisyon-kararlari>
[Erişim Tarihi: 15.01.2022]
- URL-7: <https://www.ayedas.com.tr/yasal-bildirim/aydinlatma-tuketim-bilgileri#>
[Erişim Tarihi: 15.01.2022]
- URL-8: https://www.tedas.gov.tr/#!dagitim_srkt
[Erişim Tarihi: 15.01.2022]
- URL-9: <http://www.tredas.com.tr/icerik/genel-aydinlatma-bilgileri-165>
[Erişim Tarihi: 15.01.2022]
- URL-10: <https://www.ayedas.com.tr/yasal-bildirim/ticari-kalite>
[Erişim Tarihi: 15.01.2022]
- URL-11: <https://www.meramedas.com.tr/tr/2022-13.html#detay>
[Erişim Tarihi: 15.01.2022]
- URL-12: https://www.tedas.gov.tr/#!tedas_tarifeler_1
[Erişim Tarihi: 15.01.2022]
- URL-13: https://scikit-learn.org/stable/modules/outlier_detection.html
[Erişim Tarihi: 12.20.2022]
- URL-14: <https://xgboost.readthedocs.io/en/stable/index.html#>
[Erişim Tarihi: 12.20.2022]
- URL-15: https://facebook.github.io/prophet/docs/quick_start.html#python-api, OpenSource
Facebook Fbprophet, 2021
[Erişim Tarihi: 12.20.2022]