Çukurova Üniversitesi
Mühendislik Fakültesi Dergisi
Çukurova University
Journal of the Faculty of Engineering
ISSN: 2757-9255 CİLT/VOLUME: 40 SAYI/ISSUE: 3 EYLÜL/SEPTEMBER 2025

Akıllı Şebekelerde Elektrik Hırsızlığı Tespiti İçin Gelişmiş Makine Öğrenmesi Yöntemlerinin Uygulanması

Ömer Can TOLUN [1,a], **Kasım ZOR** [1,b], **Önder TUTSOY** [1,c]

[1]*Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Adana, Türkiye*

[a]***ORCID**: 0000-0002-1956-4303; [b]**ORCID**: 0000-0001-6443-114X; [c]**ORCID**: 0000-0001-6385-3025*

Makale Bilgileri

Geliş : 25.08.2025
Kabul : 10.09.2025
DOI: 10.21605/cukurovaumfd.1772073

Sorumlu Yazar

Ömer Can TOLUN
octolun@atu.edu.tr

Anahtar Kelimeler

Bayes optimizasyon
Boyut indirgeme
Elektrik hırsızlığı tespiti
LightGBM

***Atıf şekli:** TOLUN, Ö.C., ZOR, K., TUTSOY, Ö., (2025). Akıllı Şebekelerde Elektrik Hırsızlığı Tespiti İçin Gelişmiş Makine Öğrenmesi Yöntemlerinin Uygulanması. Çukurova Üniversitesi, Mühendislik Fakültesi Dergisi, 40(3), 627-641.*

ÖZ

Elektrik hırsızlığı, dağıtım sistemlerinde hem ekonomik kayıplara hem de şebeke güvenilirliğinin azalmasına yol açan en kritik sorunlardan biridir. Bu çalışmada, elektrik hırsızlığı tespiti (EHT) için veri dengesizliği ve model iyileştirme problemlerini aynı anda ele alan gelişmiş bir makine öğrenmesi algoritması sunulmaktadır. Öncelikle, meskenlerden elde edilen gerçek elektrik tüketim verilerine ön işleme adımları uygulandıktan sonra dengesizliği azaltmak için K-ortalamalar tabanlı bir dengeleme yöntemi uygulanmıştır. Boyut indirgeme aşamasında, modelin yalnızca en anlamlı girdilerle eğitilmesi amacıyla karşılıklı bilgi temelli en iyilerin seçilmesi yöntemi tercih edilmiştir. Sınıflandırma aşamasında hafif gradyan artırmalı makine (LightGBM) algoritması kullanılmış ve tahmin performansı Bayes optimizasyon yöntemi ile iyileştirilmiştir. Test verisi üzerinde yapılan değerlendirmelerde, önerilen yöntemin sırasıyla %89,1 doğruluk, %97,2 kesinlik, %80,6 duyarlılık ve %88,1 F1 skoru ile güçlü sonuçlar verdiği görülmüştür.

Application of Advanced Machine Learning Methods for Electricity Theft Detection in Smart Grids

Article Info

Received : 25.08.2025
Accepted : 10.09.2025
DOI: 10.21605/cukurovaumfd.1772073

Corresponding Author

Ömer Can TOLUN
octolun@atu.edu.tr

Keywords

Bayesian optimisation
Dimension reduction
Electricity theft detection
LightGBM

***How to cite:** TOLUN, Ö.C., ZOR, K., TUTSOY, Ö., (2025). Application of Advanced Machine Learning Methods for Electricity Theft Detection in Smart Grids. Cukurova University, Journal of the Faculty of Engineering, 40(3), 627-641.*

ABSTRACT

Electricity theft is one of the most critical issues in distribution systems causing both economic losses and reduced network reliability. In this study, an advanced machine learning framework is proposed for electricity theft detection along with addressing data imbalance and model improvement problems simultaneously. Firstly, after applying preprocessing steps to the actual electricity consumption data obtained from households, a K-means-based balancing method was employed to reduce data imbalance. In the dimension reduction stage, the SelectKBest method based on mutual information was preferred to ensure that the model was trained with only the most significant inputs. In the classification phase, the light gradient boosting machine (LightGBM) algorithm was employed and the prediction performance was enhanced using the Bayesian optimisation method. Evaluations on the test dataset demonstrated that the proposed method achieved strong results with 89.1% accuracy, 97.2% precision, 80.6% recall, and 88.1% F1-score, respectively.

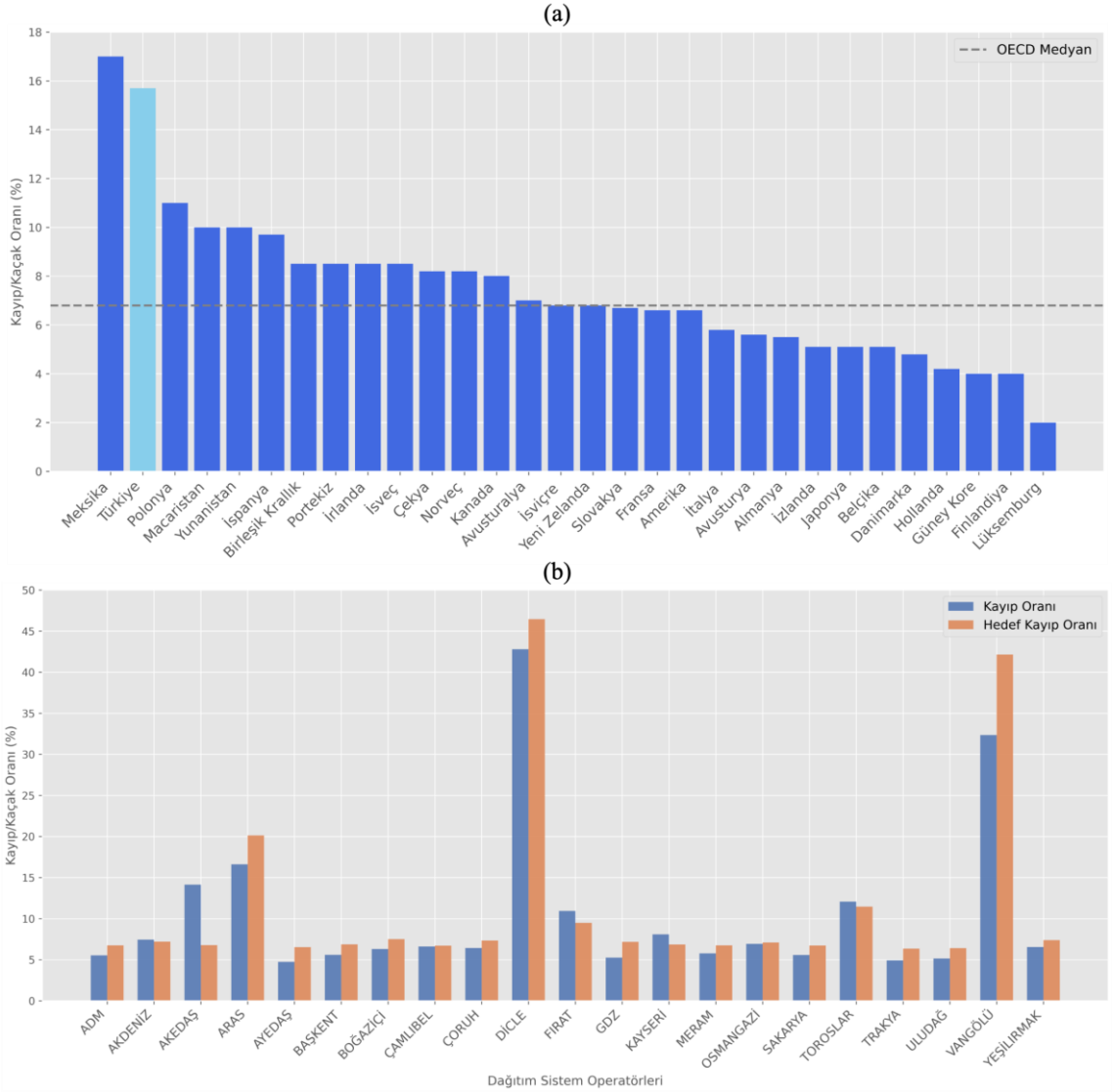
1. GİRİŞ

Akıllı şebekelerin gelişimi, enerji ağlarının işletimini ve yönetimini köklü bir şekilde değiştirmiştir. Geleneksel güç sistemlerinin dijitalleşmesi, veriye dayalı enerji yönetimini kolaylaştırmakta ve enerji arz güvenliği ile sistem istikrarı gibi temel bileşenlerin etkin şekilde sağlanmasını geliştirmektedir. Ancak bu dönüşüm süreci, akıllı sayaçlardan ve dijital enerji yönetim sistemlerinden elde edilen verilerin güvenliğini tehlikeye atan yeni sorunlar da doğurmuştur. Bu geçişte öne çıkan önemli bir gelişme, elektrik üretim, iletim, dağıtım ve tüketim süreçlerini dönüştüren ileri ölçüm altyapısının (AMI) yaygın olarak uygulanmasıdır. AMI'nin temel bir bileşeni olan akıllı sayaçlar, kurcalamaya ve müdahaleye karşı oldukça savunmasız olan geleneksel elektromekanik sayaçların yerini giderek daha fazla almaktadır [1]. Akıllı şebekelerin hızlı benimsenmesine ve önümüzdeki on yıl içinde 50 milyar doları aşması beklenen bu pazarın büyümesine rağmen, elektrik sistemleri hâlâ önemli operasyonel kayıplar yaşamaktadır.

Bu kayıplar genellikle iki ana grupta sınıflandırılır: teknik kayıplar ve teknik olmayan kayıplar. Teknik kayıplar, elektrik iletimi ve dağıtımı sırasında doğal olarak ortaya çıkar; iletkenlerde ve trafolar gibi ekipmanlarda enerji kaybı şeklinde görülür. Buna karşılık teknik olmayan kayıplar büyük ölçüde elektrik hırsızlığından kaynaklanır. Bu durum, bölgesel, altyapısal ve teknolojik faktörlere bağlı olarak farklı şekillerde ortaya çıkan, küresel ölçekte yaygın bir sorundur. Akıllı şebeke uygulamaları hızla artarken, özellikle sayaç altyapısına yetkisiz erişim veya kötü niyetli müdahalelerden kaynaklanan teknik olmayan kayıpların azaltılması hem dağıtım şirketleri hem de düzenleyici kurumlar için kritik bir öncelik haline gelmiştir [2]. Elektrik hırsızlığı, dağıtım sistemlerinin ekonomik sürdürülebilirliğini zora sokmakta, dağıtım sistemi operatörlerine (DSO) ve tüketicilere ciddi mali kayıplar yaşatmaktadır.

Elektrik hırsızlığının temel nedenleri arasında sayaç arızaları, sayaca müdahale ile tüketimin eksik veya yanlış ölçülmesi, dağıtım hattına doğrudan bağlantılar ve tüketim-faturalama uyumsuzlukları yer almaktadır [3]. Elektrik hırsızlığının görülme sıklığı ülkelere ve bölgelere göre değişiklik göstermekte, bu durum hem DSO'lara hem de ulusal ekonomilere zarar vermektedir. Northeast Group'un 2021 tarihli araştırmasına göre, hırsızlık ve kayıpların küresel maliyeti yılda 101,2 milyar doları aşmaktadır [4]. Türkiye'nin üyesi olduğu Ekonomik İş Birliği ve Kalkınma Örgütü (OECD) içerisinde kayıp/kaçak tahminlerindeki sıralaması Şekil 1a'da gösterildiği gibi 38 üye ülkenin ortalamasından daha yüksektir [5]. Bununla birlikte, 2004 yılında gerçekleştirilen ilk özelleştirmeler öncesinde, elektriğin dağıtımı tek bir çatıdan gerçekleştirilmekteydi. 2013 yılına kadar devam eden bu süreç sonucunda Türkiye Elektrik Dağıtım Anonim Şirketi'nin (TEDAŞ) özelleştirilmesi ile Türkiye 21 farklı dağıtım bölgesine ayrılmıştır. Türkiye elektrik güç sistemi incelendiğinde, 2023 yılı Enerji Piyasası Düzenleme Kurumu (EPDK) raporuna göre DSO'ların fiili ve hedef kayıp oranları Şekil 1b'de gösterilmektedir. Şekil 1b'ye göre en büyük üç kayıp oranının %42,79 ile Dicle bölgesinde, %32,35 ile Vangölü bölgesinde ve %16,61 ile Aras bölgesinde olduğu görülmektedir [6].

Elektrik hırsızlığı sadece finansal kayıplara yol açmakla kalmamakta, aynı zamanda şebekenin güvenilirliğini ve istikrarını da olumsuz etkilemektedir. Literatürde EHT için önerilen yöntemler; sosyal ve ekonomik çalışmalara dayalı teorik yaklaşımlar, maliyetli ve entegrasyonu zor donanım tabanlı yöntemler ve ek donanım gerektirmeyen, mevcut altyapıya kolayca uygulanabilen veri odaklı yöntemler olmak üzere üç grupta ele alınmaktadır [7]. Bu veri odaklı yöntemler genellikle makine öğrenmesi tekniklerini içerir ve gelişmiş örüntü tanıma ile tahmin yetenekleri sayesinde yenilikçi çözümler sunar. Enerji internetinin gelişimi, DSO'lara ait akıllı sayaçlardan, kablosuz sensör ağlarından ve diğer izleme cihazlarından elde edilen büyük veri kümelerinin makine öğrenmesi teknikleri ile analiz edilmesini mümkün kılmıştır [8]. Böylece tüketim profillerindeki karmaşık anormallikler tespit edilebilmekte ve müşteriler dürüst veya dürüst olmayan şeklinde sınıflandırılabilmektedir.



Şekil 1. (a) OECD üyesi ülkelerdeki kayıp/kaçak oranları, (b) DSO'ların 2023 yılında gerçekleşen ve hedeflenen kayıp oranları.

Veri tabanlı yöntemler, sürekli öğrenme ve uyum sağlama özellikleri sayesinde doğruluklarını zamanla artırabilir ve farklı dağıtım bölgelerindeki kayıpların tespitine uygun hale getirilebilmektedir [9]. Son dönem çalışmalarda, denetimli, denetimsiz ve pekiştirmeli, derin, topluluk ve melez yöntemler elektrik hırsızlığının tespitinde kullanıldığı görülmüştür. Ancak mevcut araştırmaların büyük bir kısmı, elektrik hırsızlığına dair gerçek veri setlerinin kullanımının eksikliği nedeniyle sınırlıdır. Literatürde, gerçek hırsızlık vakalarının nadirliği, dengesiz veri dağılımı ve etiketleme zorlukları, veri odaklı modellerin başarısını olumsuz etkilemektedir [10]. Bu çalışmanın özgün katkıları aşağıdaki şekilde özetlenebilir:

- Veri dengesizliği problemini çözmek amacıyla K-ortalamalar tabanlı bir dengeleme yöntemi önerilmiştir.
- Modelin yalnızca en anlamlı değişkenlerle eğitilmesi için karşılıklı bilgi temelli en iyilerin seçilmesi yöntemi uygulanmış, böylece hem öğrenme kapasitesi hem de yorumlanabilirliği geliştirilmiştir.
- Sınıflandırma aşamasında LightGBM algoritması kullanılmış ve hiperparametreler Bayes optimizasyon ile ayarlanarak yüksek ve kararlı performans elde edilmiştir.

Giriş bölümünde literatüre yönelik kısa bir değerlendirmeye yer verilmesinin ardından makale şu şekilde düzenlenmiştir: Literatür araştırması bölümünde, EHT'ye ilişkin mevcut çalışmaların güncel durumu

ortaya konulmakta ve kuramsal çerçeve sunulmaktadır. Materyal ve metot bölümünde, kullanılan veri seti tanıtılmakta ve temel özellikleri açıklanmakta; ayrıca önerilen yaklaşımın metodolojisini gösteren bir akış diyagramı sunulmakta ve kullanılan yöntemler ayrıntılı biçimde açıklanmaktadır. Bulgular ve tartışma bölümünde, uygulanan yöntemlerden elde edilen sonuçlar analiz edilmekte ve literatürle karşılaştırmalı olarak tartışılmaktadır. Son olarak, sonuçlar bölümünde çalışmanın temel çıktıları özetlenmekte ve gelecek araştırmalara yönelik katkılar ortaya konulmaktadır.

2. LİTERATÜR ARAŞTIRMASI

Bu bölümde, literatürdeki dağıtım sistemlerinde EHT çalışmaları sunulmaktır. Bu çalışmalar öğrenme yöntemlerine göre gözetimli, gözetimsiz, pekiştirmeli, derin, topluluk ve melez olarak ayrı alt başlıklar altında incelenmiştir. Aşağıdaki alt bölümlerde, EHT'de uygulanan her bir öğrenme tekniğine ilişkin ayrıntılı bir literatür incelemesi sunulmakta; bu tekniklerin metodolojileri, etkinlikleri ve pratik uygulamaları analiz edilmektedir. Bu çalışma özellikle son üç yılda gerçekleştirilen araştırmalara odaklanmakta ve makine öğrenmesi tabanlı alanındaki güncel gelişmeleri ve ortaya çıkan eğilimleri vurgulamaktadır. Ayrıca, Çizelge 1 incelemeye dâhil edilen çalışmalara ilişkin kapsamlı bir özet sunmaktadır.

Çizelge 1. Güncel denetimli, denetimsiz, pekiştirmeli ve derin öğrenme, topluluk, melez tabanlı EHT yaklaşımlarına genel bakış

Referans	Teknikler	Alt kategori	Güçlü yönler	Sınırlılıklar
[11]	ADASYN-SGWO-PRF	Gözetimli	Sınıf dengesizliğini melez bir aşırı örnekleme–az örnekleme yaklaşımıyla etkili bir şekilde yönetir; PRF hattı ile yüksek doğruluk sağlar.	Aşırı örnekleme nedeniyle gürültü ve sınıf örtüşmesi oluşabilir; model performansı ön işleme aşamalarının dikkatli ayarlanmasına bağlıdır.
[12]	DGRGNN	Gözetimli	Dinamik artık grafik ağları kullanarak karmaşık zamansal ve yapısal ilişkileri yakalar; gerçek zamanlı topoloji güncellemelerini entegre eder.	KNN grafik yapımında k-değerine duyarlıdır; yük eğrilerinden derin özellikler çıkarmakta zorlanır.
[13]	CLOF	Gözetimsiz	LOF ile kümeleri birleştirerek örtüşen aykırı değerleri tespit eder; geniş bir sahtecilik davranışı yelpazesini belirlemede etkilidir; dengesiz veri kümelerine karşı dayanıklıdır.	LOF örtüşen aykırı değerlerle zorlanır; gerçek tüketime çok benzeyen küçük saldırılarda performans düşebilir; küme yapılandırmasına bağlıdır.
[14]	DWT-FCM	Gözetimsiz	Güvenli gözlemci sayaç verilerini ve dalgacık tabanlı özellik çıkarımını entegre eder; farklı kullanıcı türleri ve saldırı senaryoları arasında üstün performans ve düşük yanlış alarm sağlar.	Performans, gözlemci sayaçların kalitesi ve yapılandırmasından etkilenebilir; gerçek kurcalanmış veri üzerinde test edilmemiştir; hesaplama verimliliğine dair sınırlı bilgi vardır.
[15]	DQN	Pekiştirmeli	Ön işleme olmadan veri dengesizliğini giderir; özellik-maliyet verimliliğini entegre eder.	Gerçek hırsızlık verisi eksikliği nedeniyle sentetik verilere dayanır; optimum özellik seçimi ve yakınsama için kapsamlı benzetim gerektirir.
[16]	DQN	Pekiştirmeli	Tüketim kalıplarındaki değişikliklere ve yeni saldırılara uyum sağlar; keşif ve sömürü mekanizmaları sayesinde tam yeniden eğitmeden kaçınır.	Model karmaşıklığı nedeniyle hesaplama yükü fazladır; yüksek dinamik ortamlarda aşırı öğrenmeye yatkındır, büyük veri kümeleri gerektirir.
[17]	CNN	Derin	Özellik mühendisliği ile yorumlanabilirliği artırır; dikkat mekanizmasına dayalı çok ölçekli CNN kullanarak anomali tespiti ve performansı geliştirir.	Orantılı veya ters manipülasyonları tespit etmekte zorlanabilir; özellik çıkarım kalitesine duyarlıdır.
[18]	D-FFNN	Derin	Kötü niyetli ve iyi niyetli anomalileri ayırt eder; gerçek zamanlı öz öğrenme ve hava durumu gibi dış faktörleri entegre eder.	Yüksek hesaplama gereksinimleri ve yeni şebeke teknolojilerine sınırlı uyum; yanlış sınıflandırma riski.
[19]	ChebGCN	Derin	Çoklu enerji yükleri arasındaki bağımlılıkları yakalar; elektrik, ısıtma, soğutma ve su sistemleri arasındaki korelasyonları modeller; ChebGCN kullanır.	Sınırlı kapsam; konut dışı ortamlar için genelleme zayıf olabilir; meteorolojik veya voltaj verileri gibi ek girdiler dahil edilmemiştir.

Çizelge 1. *Devamı*

Referans	Teknikler	Alt kategori	Güçlü yönler	Sınırlılıklar
[20]	XGBoost	Topluluk	SCADA verilerini XGBoost ile entegre ederek birden fazla sentetik saldırı tipini yüksek doğrulukla tespit eder.	Değerlendirme, simülasyon tabanlı senaryolarla sınırlıdır; gerçek zamanlı uygulama tartışması eksiktir.
[21]	XGBoost	Topluluk	Sağlam özellik çıkarımı ve otomatik optimizasyon için AutoXGB ile OS-CNN önermektedir.	Karmaşık mimari ve aşırı uyum riskinden dolayı zarar görür; hiperparametre ayarında potansiyel aşırıya kaçma yaşanabilir.
[22]	XGBoost	Topluluk	Düşük örnekleme oranına sahip geleneksel ölçüm verilerine melez makine öğrenmesi uygular; az araştırılmış bir bağlamı ele alır.	Veri seti kısıtlamaları ve düşük frekanslı okumalar üzerine odaklanma nedeniyle genelleştirilebilirliği sınırlıdır.
[23]	CNN-LSTM	Melez	Dengesiz ve anormal verilerde performansı artırmak için topluluk öğrenmesi ve prototip öğrenmeyi birleştirir.	Prototip temsiliyetine bağlıdır; yüksek veri gürültüsü veya kaymalarında daha az etkili olabilir.
[24]	CNN-Adaboost	Melez	Melez CNN–AdaBoost modeli önerir ve ADGM'yi tanıtarak tabular verilerle EHT modeli dayanıklılığını kaçış saldırılarına karşı etkili bir şekilde değerlendirir.	Melez model nedeniyle yüksek hesaplama maliyeti içerir; dürüst olmayan veri tek bir veri seti kullanılarak üretilmiştir, bu da model genelleştirilebilirliğinin değerlendirilmesini sınırlar.
[25]	LSTM-Bi-LSTM-GRU-Bi-GRU SVM-RF-DT-AdaBoost	Melez	Dengesiz verilerde performansı artırmak için sağlam yeniden örnekleme teknikleriyle birlikte yığılmış topluluk öğrenmesini kullanır.	Modelin karmaşıklığı ve birden fazla temel/meta-sınıflandırıcıya olan bağımlılığı, artan hesaplama maliyetine yol açabilir.

2.1. Gözetimli Öğrenme

Literatürde EHT kapsamında farklı denetimli öğrenme yaklaşımlarının kullanıldığı görülmektedir. Tripathi ve arkadaşları, aşırı örnekleme, özellik seçimi ve sınıflandırma adımlarını bir arada ele alan uyarlanabilir sentetik örnekleme yaklaşımı (ADASYN), alt küme gri kurt optimizasyonu (SGWO) ve boru hattı rastgele orman sınıflandırıcısı (PRF) tabanlı bir makine öğrenmesi yöntemi geliştirmiştir [11]. Zhuang ve arkadaşları, düğüm bilgisi toplama doğruluğunu artırmak amacıyla dinamik artık grafik sinir ağı (DGRGNN) modelini önermiştir [12].

2.2. Gözetimsiz Öğrenme

Literatürde veri odaklı yöntemlerin bir alt kümesi olan denetimsiz öğrenme teknikleri de EHT'de kullanılmaktadır. Bu teknikler, etiketli verilere ihtiyaç duymadan veri kümesini benzerlik ve farklılıklara göre gruplara ayırmaktadır. Bu bağlamda tüketiciler, yük profilleri üzerinden dürüst, kötü niyetli veya hileli olarak sınıflandırılabilmekte ve anormal tüketim davranışları tespit edilebilmektedir. Peng ve arkadaşları, k-ortalama ve yerel aykırı faktör (LOF) yöntemlerini birleştiren korelasyon tabanlı yerel aykırı faktör (CLOF) yaklaşımını önermiştir [13]. Bunun yanı sıra Qi ve arkadaşları, tüketici yük profilleri ile gözlemci sayaç profillerini analiz ederek hileli kullanıcıları tespit etmeyi amaçlayan, ayrık dalgacık dönüşümü (DWT) ve bulanık c-ortalama kümeleme (FCM) tabanlı yeni bir denetimsiz öğrenme yöntemini geliştirmiştir [14].

2.3. Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme, bir ajanın çevresiyle etkileşime girerek ödül ve ceza mekanizması üzerinden karar verme stratejileri öğrenmesine dayalıdır. Elektrik hırsızlığı tespitinde ise tüketim anomalilerinin uyarlanabilir biçimde izlenmesi ve tespit stratejilerinin geliştirilmesi amacıyla kullanılmaktadır. Literatürde, Lee ve arkadaşları güç şebekelerinde teknik olmayan kayıpların tespiti için derin Q-öğrenme tabanlı bir yöntem geliştirmiş ve bu sayede veri dengesizliği problemini aşarak veri çoğaltmaya gerek kalmadan hırsızlık tespitini mümkün kılmıştır [15]. El-Thoukhy ve arkadaşları ise dinamik tüketim kalıpları ve siber saldırıların tespitine yönelik derin pekiştirmeli öğrenme (DRL) tabanlı bir yaklaşım önermiştir. Bu yöntem, yeni tüketim alışkanlıklarını ve saldırı türlerini öğrenebilme kapasitesiyle öne çıkmaktadır [16].

2.4. Derin Öğrenme

Derin öğrenme, makine öğrenmesinin bir alt dalı olup çok katmanlı yapay sinir ağlarını kullanarak büyük ve karmaşık veri setlerinde hiyerarşik temsil öğrenmeyi amaçlamaktadır. İnsan beyninin işleyişinden esinlenen bu yöntem, özellikle çok boyutlu ve yoğun verilerde yüksek başarı göstermektedir. Literatürde, Zhang ve Dai, özellik mühendisliği tabanlı birçok ölçekli anomali tespit modeli önermiştir. Bu modelde özellik çıkarımı için özellik mühendisliği, özellik seçimi için aşırı gradyan artırma (XGBoost) algoritması ve anomali tespiti için dikkat mekanizmasıyla desteklenen çok ölçekli evrişimsel sinir ağı (CNN) yapısı kullanılmıştır [17]. Althobaiti ve arkadaşları ise çift katmanlı derin ileri beslemeli sinir ağı (D-FFNN) tabanlı uyarlanabilir bir sistem geliştirmiş ve bu sistem, enerji hırsızlığı ile yanlış yapılandırmaları ayırt etmede yüksek doğruluk sağlamıştır [18]. Liao ve arkadaşları ise çoklu enerji yüklerinden gizli özellikleri ve korelasyonları yakalamak amacıyla Chebyshev grafik evrişimsel ağ (ChebGCN) tabanlı bir yöntem önermiştir [19].

2.5. Topluluk Öğrenmesi

Topluluk öğrenmesi, birden fazla bireysel modeli bir araya getirerek tahmin doğruluğu ve sağlamlığı artırmayı hedefleyen bir makine öğrenmesi yaklaşımıdır. Bu yöntemin temel mantığı, farklı modellerin çıktılarının birleştirilmesiyle tek bir modelin tek başına ulaşabileceğinden daha iyi bir performans elde etmek; aşırı uyum, varyans ve yanlılık risklerini azaltmaktır. Literatürde, Sharma ve Tiwari, denetim kontrol ve veri toplama (SCADA) sistemiyle entegre edilmiş optimize edilmiş bir XGBoost sınıflandırıcı önermiş ve akıllı şebekelerde anomali tespitinde daha yüksek doğruluk elde etmiştir [20]. Zhu ve arkadaşları, sentetik azınlık aşırı örnekleme tekniği ile düzeltilmiş en yakın komşular (SMOTEENN) tabanlı melez bir örnekleme yöntemi ile birlikte min–maks ölçekleme ve sıfır–skor standartlaştırma gibi ön işleme tekniklerini kullanarak çok ölçekli CNN ve Otomatik XGBoost (AutoXGB) modellerini entegre ederek elektrik hırsızlığı tespiti için bir yaklaşım geliştirmiştir [21]. Norouzi ve arkadaşları ise geleneksel sayaçların kullanıldığı bölgelerde, düşük örnekleme hızına ve eşzamanlı olmayan veriye ilişkin zorlukları gidermek amacıyla aylık veri dönüşümü, sentetik veri dengeleme yöntemleri ve XGBoost sınıflandırıcı içeren melez bir çerçeve önermiştir [22].

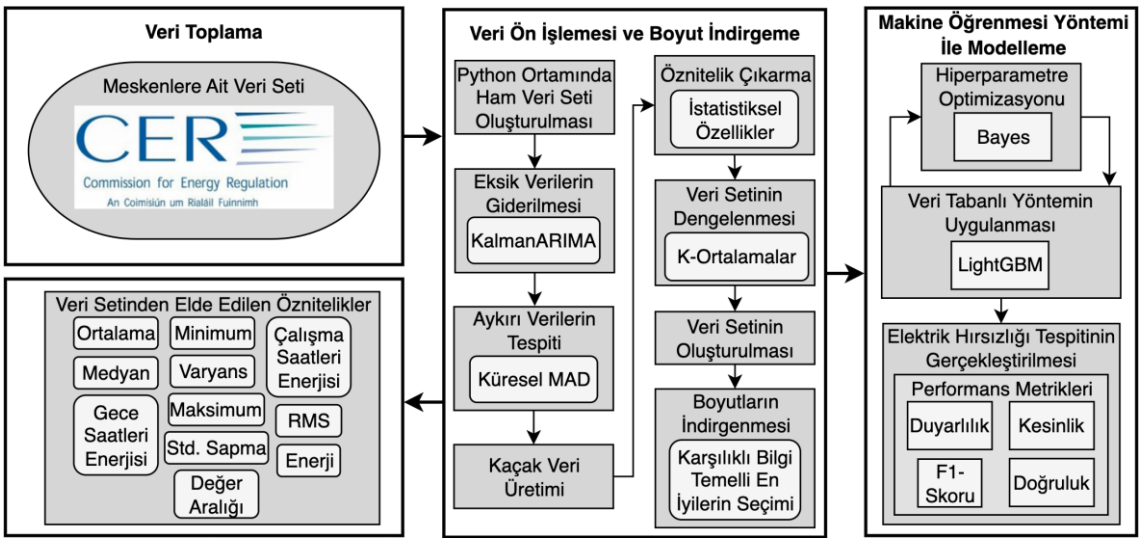
2.6. Melez Öğrenme

Melez Öğrenme, farklı öğrenme paradigmalarını, modellerini veya tekniklerini bir araya getirerek her birinin güçlü yönlerinden faydalanmayı ve genel performansı artırmayı hedefleyen bir makine öğrenmesi yaklaşımıdır. Bu yöntem, bireysel modellerin sınırlılıklarını aşarak daha iyi genelleme, verimlilik ve uyarlanabilirlik sağlamayı amaçlamaktadır. Literatürde, Sun ve arkadaşları, CNN ve uzun kısa süreli bellek ağı (LSTM) tabanlı özellik çıkarımı ile prototip tabanlı sınıflandırmayı birleştiren melez bir EHT çerçevesi önermiştir [23]. Nirmal ve arkadaşları, CNN ile uyarlamalı artırma (AdaBoost) tabanlı bir çerçeve geliştirmiş ve modeli hızlı gradyan işareti yöntemi (FGSM) ve düşman veri üretim yöntemi (ADGM) kullanarak düşman saldırıları altında test etmiştir [24]. Ullah ve arkadaşları ise birden fazla makine öğrenmesi modelini birleştirerek yığılmış genelleştirme (stacked generalisation) yaklaşımına dayalı melez bir çerçeve sunmuştur [25].

3. MATERYAL VE METOT

Elektrik dağıtım sistemlerinde EHT için önerilen metodoloji, Şekil 2'de gösterildiği üzere çok aşamalı ve organize bir süreçten oluşmaktadır. Süreç, İrlanda Enerji Düzenleme Komisyonu (CER) veri seti kullanılarak Python ortamında ham veri setinin oluşturulması ile başlamaktadır. Bu aşamayı takiben, veri setini zenginleştirmek amacıyla dürüst olmayan müşterilere ait veri üretimi gerçekleştirilmekte, ardından veri ön işleme adımlarında eksik verilerin giderilmesi için zaman serisi durum-uzay modellerine dayalı Kalman ARIMA filtreleme (KA), aykırı verilerin tespiti için modifiye Z-skoru tabanlı küresel medyan mutlak sapma (Küresel MAD) işlemleri gerçekleştirilmektedir. Bu aşamadan sonra, veri setinin dengelenmesi K-ortalamalar yöntemi kullanılarak gerçekleştirilmekte, böylece sınıf dağılımları iyileştirilmekte ve nihai veri seti oluşturulmaktadır. Veri setinden elde edilen öznitelikler, ortalama, medyan, varyans, standart sapma, minimum, maksimum ve değer aralığı gibi istatistiksel ölçüler ile çalışma

saatleri enerjisi, gece saatleri enerjisi, toplam enerji ve etkin değer (RMS) gibi enerji temelli özelliklerden oluşmaktadır [26]. Öznitelik çıkarımı sonrası, boyutların indirgenmesi amacıyla karşılıklı bilgi temelli en iyilerin seçimi gerçekleştirilmekte, ardından Bayes yöntem kullanılarak hiperparametre optimizasyonu uygulanmaktadır. Son aşamada ise, LightGBM tabanlı yöntem ile elektrik hırsızlığı tespiti yapılmakta ve elde edilen sonuçlar duyarlılık, kesinlik, F1-skoru ve doğruluk gibi performans metrikleri ile değerlendirilmektedir.



Şekil 2. EHT için önerilen metodoloji.

3.1. Materyal

Bu bölümde, literatürden elde edilen dürüst veri seti ve dürüst veriden elde edilen dürüst olmayan veriler sunulmaktadır.

3.1.1. Dürüst Veri Seti

Bu makalede, CER kapsamında elde edilen enerji tüketim kayıtları kullanılmıştır. Bu veri setinde yarım saatlik aralıklarla 5.000'den fazla tüketiciye ait veri noktası bulunmaktadır. Kullanıcı sayısının fazlalığı nedeniyle bu çalışma, araştırmalar için önemli bir kaynak oluşturmaktadır [27]. Veri seti 535 günlük bir zaman periyodunu kapsamaktadır ve örnekleme aralığı 30 dakikadır. Basitlik adına, tüm 2009 ve 2010 yıllarına ait 1.000 müşteri dikkate alınmıştır.

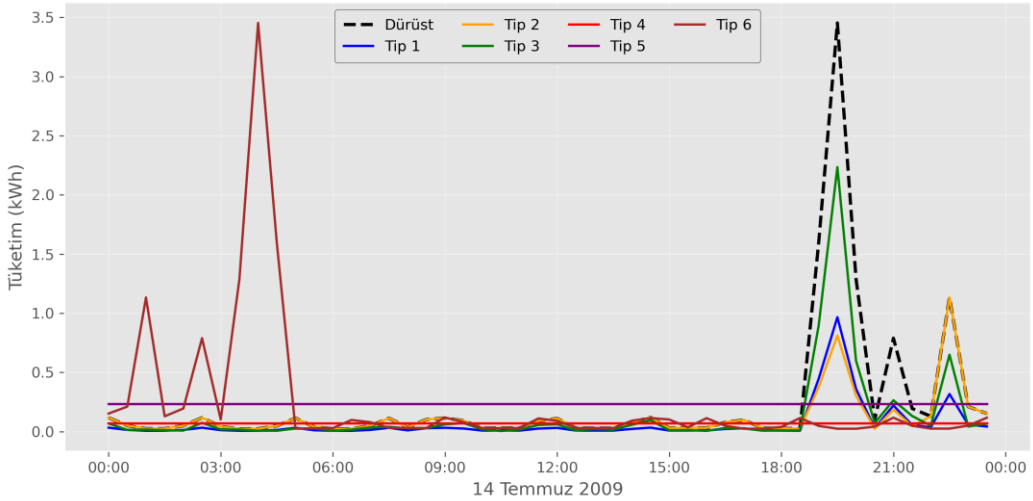
3.1.2. Dürüst Olmayan Veri Seti

Bu araştırma boyunca, dürüst olmayan müşteri davranışlarına ilişkin veri toplama sürecinde sürekli bir zorluk ortaya çıkmaktadır. Çizelge 2'de gösterildiği gibi altı farklı saldırı fonksiyonu kullanılarak, yapay şekilde yeni örnekler üretilmiştir. Tip 1 ve Tip 3 siber saldırı fonksiyonları aynı zamanda kısmi azaltma saldırıları olarak da adlandırılmaktadır [28]. Fonksiyon $y_1(x(t))$ için, enerji tüketim değeri rastgele seçilen bir kesir (α) ile azaltılmaktadır; burada α, 0.1 ile 0.8 aralığından rastgele seçilmektedir. Bu işlem, bildirilen elektrik tüketim değerini düşürmek amacıyla yapılmaktadır. Fonksiyon $y_3(x(t))$'de, α zamanla değişkenlik göstermektedir. Tip 2, enerji hırsızlığı ve siber saldırının en yaygın yöntemlerinden biridir. Belirli bir zaman aralığı için fonksiyon $y_2(x(t))$ devreye girer ve geri kalan değerler orijinal şekilde yansıtılır. Bu saldırı genellikle seçici bypass saldırısı olarak adlandırılır. Tip 5'te, fonksiyon $y_5(x(t))$ verilerin ortalamasını alarak enerji tüketim profilini sabit tutar. $y_4(x(t))$ fonksiyonunda ise kullanıcılar, ortalama alındıktan sonra verileri α azaltma katsayısı ile çarpar; böylece enerji profilinde belirgin bir azalma görülebilir. Son olarak $y_6(x(t))$ fonksiyonunda, enerji tüketim profili tersine çevrilerek maksimum kâr

elde edilmektedir [29]. Zaman serisi üzerinde altı farklı saldırı türüne ilişkin örnekler Şekil 3'te gösterilmiştir.

Çizelge 2. Elektrik hırsızlığı tiplerinin matematiksel gösterimi.

Hırsızlık tipleri	Matematiksel gösterimi
Tip 1	$y_1(x(t)) = \alpha x(t),\ \alpha = \text{rastgele}(0.1, 0.8)$
Tip 2	$y_2(x(t)) = \beta x(t),\ \beta = \begin{cases} 0 & T_{başlangıç} < t < T_{bitiş} \\ 1 & \text{aksi taktirde} \end{cases},\ T_{başlangıç} = \text{rastgele}(0, 47 - \text{zaman})$ $süre = \text{rastgele}(\text{zaman}, 48),\ T_{bitiş} = T_{başlangıç} + \text{süre}$
Tip 3	$y_3(x(t)) = \lambda(t) x(t),\ \lambda(t) = \text{rastgele}(0.1, 0.8)$
Tip 4	$y_4(x(t)) = \lambda(t)\,\text{ortalama}(x(t)),\ \lambda(t) = \text{rastgele}(0.1, 0.8)$
Tip 5	$y_5(x(t)) = \text{ortalama}(x_t)$
Tip 6	$y_6(x(t)) = x_{48-t}$



Şekil 3. Farklı tip hırsızlık senaryoları için elektrik tüketim profillerinin karşılaştırılması

3.2. Metot

Bu bölümde, EHT için önerilen metodolojideki yöntemler sunulmaktadır.

3.2.1. KalmanARIMA

ARIMA sınıfı modeller, uygun bir dönüşümle durum-uzay biçimine yazılabildiğinden parametre kestirimi, filtreleme, yumuşatma ve öngörü adımları Kalman filtresi ile birleştirilebilir. Bu yaklaşım literatürde sıklıkla KA olarak anılır ve özellikle eksik gözlemler, düzensiz örnekleme aralıkları veya gürültülü ölçümler içeren uygulamalarda esneklik sağlar. Matematiksel olarak, Kalman filtreleri aşağıdaki denklemlerle verilen durum-uzay modellerine dayalı olarak iki aşamada uygulanır [30].

$$x_t = F_t x_{t-1} + \varepsilon_t \tag{1}$$

$$y(t) = H_t x_t + \omega_t \tag{2}$$

Burada x_t, t anındaki sistemin durum vektörünü, y_t ise aynı anda elde edilen ölçüm vektörünü temsil etmektedir. F_t, sistemin durum geçiş parametresi, ε_t rastgele durum gürültü terimi, H_t ölçüm parametresi ve w_t ölçüm hatası terimidir.

3.2.2. Küresel MAD

Modifiye Z-skoru tabanlı küresel medyan mutlak sapma yöntemi, aykırı değer tespitinde medyana dayalı dayanıklı bir ölçek kestiricisi kullanır ve tüm zaman serisi/özellik boyunca tek bir küresel eşik ile çalışır. Serinin medyanı m ve medyan mutlak sapması

$$MAD = median\left(\left|x_t - m\right|\right) \tag{3}$$

ile hesaplanır. Normal dağılım altında $MAD \approx 0.6745\sigma$ olduğundan, modifiye Z-skoru

$$M_t = 0.6745\frac{x_t - m}{MAD} \tag{4}$$

şeklinde tanımlanır [31]. Gözlem $x_t, \left|M_t\right| > \tau$ ise aykırı sayılır, literatürde $\tau = 3.5$ sık kullanılan bir eşiktir. Uygulama adımları ele alındığında, (i) Her değişken için m ve MAD küresel olarak hesaplanır. (ii) M_t değerleri bulunur ve $\left|M_t\right| > \tau$ olanlar küresel aykırı olarak etiketlenir. (iii) Zaman serilerinde trend/mevsimsellik baskınsa yöntem, ham veri yerine farklar ya da ayrıştırma artıklarına uygulanır. (iv) $MAD = 0$ durumunda tespit devre dışı bırakılır veya küçük bir ε ile düzenleme yapılır. Çok değişkenli senaryolarda tespit değişken bazında yürütülür ve bayrakların birleşimi alınır [32].

3.2.3. K-Ortalamalar

Sınıf dengesizliğini giderirken örnek uzayının yapısını korumak amacıyla özellik uzayı k kümeye bölünerek her sınıf içinde küme-tabanlı örnekleme yapılır. K-ortalamalar, amaç fonksiyonunu en aza indirerek her küme için bir merkez μ_c üretir [33].

$$\min_{\{\mu_c\},\{S_c\}} \sum_{c=1}^{k} \sum_{x_i \in S_c} \left\|x_i - \mu_c\right\|^2 \tag{5}$$

Her sınıf $y \in \{0,1\}$ için küme boyları $n_c^{(y)}$ ve toplamı $N^{(y)}$ olmak üzere, sınıf içi dağılımı korumak adına hedef küme payları

$$\pi_c^{(y)} = \frac{n_c^{(y)}}{N^{(y)}} \tag{6}$$

tanımlanır. Sınıf başına hedef örnek sayısı N_* belirlendiğinde, küme bazlı hedefler

$$t_c^{(y)} = \left\lceil \pi_c^{(y)} N_* \right\rceil \tag{7}$$

olarak atanır. Böylece yeniden örnekleme her sınıfın küme içi morfolojisini korur.

Prosedür açıklanmak istenirse (i) sürekli değişkenler standartlaştırılır, k-ortalamalar yalnızca eğitim bölümü üzerinde uygulanır. (ii) Her sınıf için $n_c^{(y)}$ ile $t_c^{(y)}$ karşılaştırılır. Aşırı temsil durumunda her kümeden olasılıkça orantılı alt-örnekleme yapılır. Yetersiz temsil durumunda her kümede $t_c^{(y)} - n_c^{(y)}$ adet sentetik örnek üretilir. Sentetik üretim, küme-içi iki komşu $x_a, x_b \in S_c^{(y)}$ üzerinden

$$\tilde{x} = (1-\lambda)x_a + \lambda x_b, \quad \lambda \sim \Pi(0,1) \tag{8}$$

ile gerçekleştirilir, kategorik öznitelikler için moda-korumalı çoğullama uygulanır. (iii) Üretilen/seçilen örnekler birleştirilir, doğrulama/test bölümleri dokunulmadan bırakılır. Uygulama kısmına geçildiğinde ilk

değinilmesi gereken nokta k değerinin dirsek veya silüet katsayısı ile adaylar arasından seçilmesi ve küme kararlılığının çapraz doğrulama boyunca izlenmesidir. Aykırı değerlerin etkisini azaltmak için kümelemeye Küresel MAD temizliği sonrası girilir.

3.2.4. Karşılıklı Bilgi Tabanlı k-En İyi Özellik Seçimi

Bu metotta her aday öznitelik X_j ile hedef değişken Y arasındaki bağımlılık karşılıklı bilgi (KB) ile ölçülür ve KB puanlarına göre en iyi bilgi taşıyan k öznitelik seçilir. KB doğrusal ilişki varsayımı gerektirmeksiniz ortak dağılımdaki bilgi kazanımını niceler [34]:

$$I(X;Y)=\sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)\,p(y)} \tag{9}$$

Sürekli değişkenler için $I(X;Y)$ kestirimi pratikte k-en yakın komşu veya parçalı-ayrıklaştırma temelli pratik olmayan tahmincilerle gerçekleştirilir. Sınıflandırmada $I(X;Y)$ sınıf bilgisini, regresyonda ise doğrusal olmayan ilişkileri de içerecek biçimde bağımlılığı yakalar.

3.2.5. Bayes Yöntemiyle Hiperparametre Optimizasyonu

Modelin doğrulama kaybı $f(\lambda)$ kapalı formda bilinmeyen ve değerlendirmesi maliyetli bir kara kutu olarak ele alınmıştır. Bayes optimizasyon, bu amacı yaklaşım modeli ile istatiksel olarak modelleyip yeni denemeleri bir edinim fonksiyonu ile seçerek bulur. Gauss süreci yaklaşımına odaklanıldığında, $f(\lambda)\sim GP\left(m(\lambda),k(\lambda,\lambda')\right)$ varsayımı altında, t. iterasyona dek toplanan $D_t=\{(\lambda_i,y_i)\}_{i=1}^{t}$ ile artakalan gürültüsü $\in\sim N(?)\left(0,\sigma_\in^2\right)$ kabul edilirse, yeni bir λ noktası için artık dağılım

$$\mu_t(\lambda)=k(\lambda)^T\left(K+\sigma_\in^2 I\right)^{-1}y,\ \sigma_t^2(\lambda)=k(\lambda,\lambda)-k(\lambda)^T\left(K+\sigma_\in^2 I\right)^{-1}k(\lambda) \tag{10}$$

şeklinde elde edilir. Burada K kovaryans matrisi, $k(\lambda)$ ise eğitim noktalarıyla kovaryans vektörüdür. Edinim fonksiyonları incelendiğinde ise yeni deneme λ_{t+1}, edinimi maksimize eden noktadır. Beklenen iyileşme (EI)

$$EI_t(\lambda)=\left(f^*-\mu_t(\lambda)\right)\varphi(z)+\sigma_t(\lambda)\phi(z),\ \text{where } z=\frac{f^*-\mu_t(\lambda)}{\sigma_t(\lambda)} \tag{11}$$

ile hesaplanır ve f^* mevcut en iyi gözlemdir, φ,ϕ ise standart normal kümülatif dağılım fonksiyonu/yoğunluk fonksiyonu dur. Üst güven sınırı (UCB) ise aşağıdaki formülasyonla hesaplanır.

$$UCB_t(\lambda)=\mu_t(\lambda)-K\sigma_t(\lambda) \tag{12}$$

Burada küçük değerler daha iyi sonuç vermekle birlikte K keşif-sömürü dengesini ayarlar. İsteğe bağlı olarak iyileşme olasılığı da kullanılabilir [35].

3.2.6. LightGBM

LightGBM gradyan artırımlı karar ağaçlarının yaprak-odaklı büyüme stratejisini ve histogram tabanlı bölme algoritmasını kullanan, hafif ve yüksek hızlı bir öğrencidir. Amaç, yinelemeli olarak zayıf öğrenciler $f_m(x)$ ekleyerek modeli

$$F_m(x)=F_{m-1}(x)+\nu f_m(x) \tag{13}$$

şeklinde güncellemektir [36]. Her yinelemede sınıflandırma kaybının ikinci mertebe Taylor açılımı kullanılır, gözlem i için gradyan ve Hessian

$$g_i = \frac{\partial \ell\left(y_i, \hat{y}_i\right)}{\partial \hat{y}_i}, \quad h_i = \frac{\partial^2 \ell\left(y_i, \hat{y}_i\right)}{\partial \hat{y}_i^{\,2}} \tag{14}$$

ile hesaplanır. İkili sınıflamada $\hat{y}_i$ lojit ve $p_i = \sigma(\hat{y}_i)$ olmak üzere lojistik kayıp $\ell(y_i, \hat{y}_i) = -y_i \log p_i - (1-y_i)\log(1-p_i)$; buradan $g_i = p_i - y_i$, $h_i = p_i(1-p_i)$ elde edilir. Bir bölmenin getirisi, düğüm toplam gradyan/Hessian'ları G, H ile

$$Gain = \frac{1}{2}\left(\frac{G_L^{\,2}}{H_L + \lambda} + \frac{G_R^{\,2}}{H_R + \lambda} + \frac{G_T^{\,2}}{H_T + \lambda} \right) - \gamma \tag{15}$$

şeklinde hesaplanır ve λ L2 ceza, γ yaprak cezasını temsil etmektedir. LightGBM yaprak-odaklı olarak, getirisi en yüksek yaprağı büyütür, aşırı uyumu sınırlamak için max_*depth* ve *num_leaves* kısıtları uygulanır. *Gain* ifadesi sonrasında LightGBM, yaprak-odaklı stratejiyle getirisi en yüksek yaprağı büyütür; aşırı uyumu sınırlamak için *num_leaves*, max_*depth* ve min_*data_in_leaf* kısıtları ile $L1/L2$ cezası uygulanır [37-39].

4. BULGULAR VE TARTIŞMA

Bu çalışmada, tüm hesaplamalar MacOS Sequoia işletim sistemi sürüm 15.6, 2.4 GHz işlemci (Intel Core i5) ve 8 GB bellek ile donatılmış bir Macintosh bilgisayarda gerçekleştirilmiştir. Tüm hesaplama görevleri için entegre bir geliştirme ortamı olarak Python programlama dili kullanılmıştır. Ayrıca, önerilen metodolojiyi uygulamak için sklearn, lightgbm ve optuna paketleri kullanılmıştır. CER veri setine uygulanan ön işleme süreci, ham verilerin Python ortamına aktarılmasıyla başlamıştır. Veri kümesi; (i) 1000–1999 aralığında hane kimlik numaraları, (ii) gün ve 30 dakikalık periyotları temsil eden kodlanmış zaman bilgisi ve (iii) kW cinsinden tüketim değerlerinden oluşmaktadır. Ön işleme kapsamında, kodlanmış zaman bilgisi okunabilir formata dönüştürülmüş, veriler hane bazında gruplandırılmış ve her grup içindeki kayıtlar kronolojik olarak sıralanarak analiz için tutarlı bir zaman serisi yapısı elde edilmiştir.

Veri ön işlemenin ikinci aşamasında, eksik gözlemler KA yöntemi ile tahmin edilmiştir. Uygulama öncesinde, eksik veri oranı %5'in üzerinde olan tüketici serileri analiz dışı bırakılmıştır. Kalan serilerde, kısa dönemli bağımlılıkları (p=1), durağanlaştırmayı (d=1) ve geçmiş hata terimlerinin etkilerini (q=1) dikkate alan KA modeli kullanılmıştır. Bu yaklaşım, zaman serisinin dinamik yapısını yansıtarak özellikle dalgalı tüketim bölgelerinde tutarlı sonuçlar sağlamış ve serinin sürekliliğini koruyarak sonraki analizlerin güvenilirliğini artırmıştır. Üçüncü aşamada, aykırı değerlerin belirlenmesi amacıyla Küresel MAD yöntemi uygulanmış; eşik değeri 3,5'in üzerinde kalan gözlemler aykırı olarak etiketlenerek veri setinden çıkarılmıştır [40]. Bu işlem, tüketim profilindeki olağandışı artış ve düşüşlerin başarılı bir şekilde ayrıştırılmasını sağlamış, temizlenen veri seti ise sonraki öznitelik çıkarımı ve sınıflandırma adımlarında daha istikrarlı ve güvenilir sonuçlar elde edilmesine katkı sunmuştur.

Dördüncü aşamada, gerçek tüketim profilleri üzerinden literatürde tanımlanan altı farklı saldırı tipi kullanılarak yapay veri setleri oluşturulmuş ve dürüst olmayan tüketici davranışlarının çeşitli senaryoları güvenilir biçimde benzetilmiştir. Beşinci aşamada ise, veri setinden günlük bazda öznitelikler çıkarılarak dürüst ve dürüst olmayan tüketim profilleri arasındaki farklılıkların belirlenmesine ve sonraki sınıflandırma sürecinin desteklenmesine katkı sağlanmıştır. Son aşama olan veri boyut indirgeme aşamasında karşılıklı bilgi temelli en iyilerin seçim yöntemi kullanılmış ve elde edilen sonuçlara göre en iyi performans, minimum değer, değer aralığı, medyan, maksimum değer, ortalama, standart sapma ve RMS özniteliklerinden oluşan yedi öznitelik seçildiğinde elde edilmiştir. Böylece yüksek boyutluluğun getirdiği gürültü azaltılmış, daha açıklayıcı ve ayırt edici bir öznitelik kümesi elde edilmiştir.

Makine öğrenmesi yöntemi ile modelleme aşamasında, Bayes optimizasyon temelli hiperparametre seçimi, LightGBM sınıflandırıcısının performansını anlamlı biçimde iyileştirmiştir. Optuna kütüphanesi ile

gerçekleştirilen 60 deneme sonucunda, doğruluk, duyarlılık, F1-skor ve kesinliğe göre en uygun parametreler elde edilmiştir. Bu süreçte, öğrenme oranı, yaprak sayısı, derinlik ve düzenlileştirme katsayıları gibi kritik hiperparametrelerin otomatik ayarlanması, modelin hem doğruluk hem de genelleme yeteneğini artırmıştır. Çapraz doğrulama sonuçları, modelin farklı veri bölünmelerinde kararlı performans gösterdiğini ortaya koymuştur. Sonuç olarak, Bayes optimizasyon yaklaşımı ile desteklenen LightGBM modeli, elektrik hırsızlığı tespitinde yüksek doğruluk ve güvenilirlik sunarak, yöntemin pratik uygulamalara uygunluğunu göstermiştir.

Modelin performansını değerlendirmek için doğruluk, kesinlik, duyarlılık ve F1-skoru metrikleri kullanılmıştır. Bu metrikler sırasıyla aşağıdaki eşitliklerde tanımlanmaktadır. Çalışmada, her bir model çıktısı bu ölçütler üzerinden değerlendirilmiş ve elde edilen sonuçlar tablo ve şekillerde karşılaştırmalı olarak sunulmuştur.

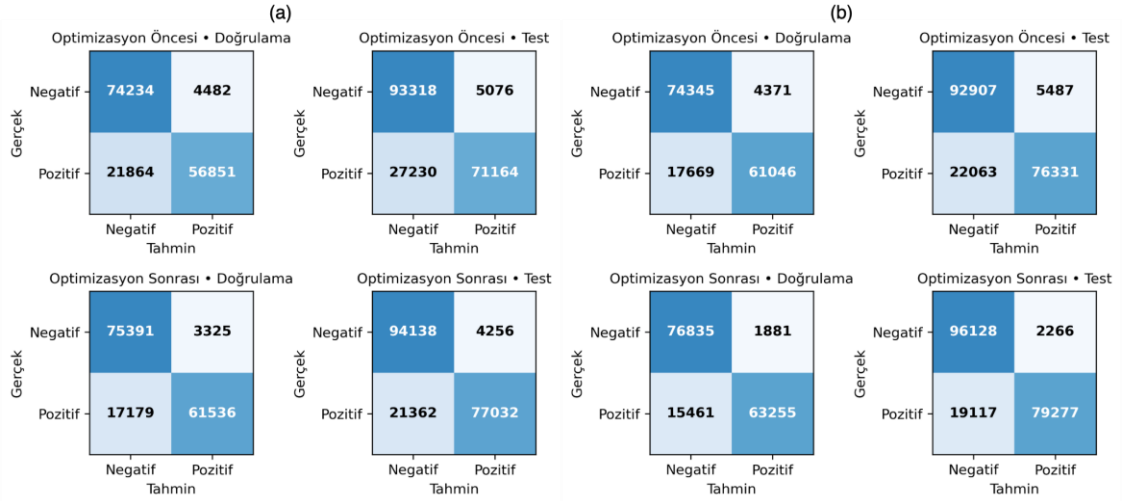
$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

$$\text{Kesinlik} = \frac{TP}{TP + FP} \tag{17}$$

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \tag{18}$$

$$\text{F1-Skor} = \frac{2 * \text{Kesinlik} * \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \tag{19}$$

Şekil 4 (a) ve (b)'de boyut indirgeme öncesi ve sonrası için doğrulama ve test setlerine ait karmaşıklık matrisleri verilmiştir. Optimizasyon öncesinde yanlış sınıflandırmalar özellikle negatif sınıflarda yoğunlaşırken, optimizasyon sonrasında doğru pozitif (TP) ve doğru negatif (TN) değerleri artmış, yanlış pozitif (FP) ve yanlış negatif (FN) değerleri azalmıştır. Bu durum, boyut indirgeme ve hiperparametre optimizasyonunun modelin ayırt ediciliğini güçlendirdiğini, sınıflandırma dengesini artırdığını ve genelleme yeteneğini koruduğunu göstermektedir. Çizelge 3 incelendiğinde, optimizasyon ve boyut indirgeme adımlarının tüm performans metriklerinde anlamlı iyileşmeler sağladığı görülmektedir. Doğruluk oranı %83,3–86,0 seviyelerinden %87,0–89,1 aralığına yükselmiştir. Kesinlik %92,7–93,3'ten %94,9–97,2'ye, duyarlılık %72,2–77,6'dan %78,2–80,6'ya çıkmış, F1-Skor ise %81,2–84,7'den %85,7–88,1 aralığına ulaşmıştır. Bu sonuçlar, yanlış pozitiflerin azaldığını, pozitif sınıfın daha doğru tespit edildiğini ve genel model performansının önemli ölçüde güçlendiğini göstermektedir.



Şekil 4. (a) Boyut indirgeme öncesinde, optimizasyon öncesi ve sonrası için doğrulama ile test setlerine ait karmaşıklık matrisi sonuçları, **(b)** Boyut indirgeme sonrasında, optimizasyon öncesi ve sonrası için doğrulama ile test setlerine ait karmaşıklık matrisi sonuçları

Çizelge 3. Boyut indirgeme ve optimizasyon öncesi/sonrası için doğrulama ve test setlerinde elde edilen performans metrikleri

Boyut indirgeme	Optimizasyon	Veri seti	Doğruluk	Kesinlik	Duyarlılık	F1-Skor
Öncesi	Öncesi	Doğrulama	83,3	92,7	72,2	81,2
	Sonrası		87,0	94,9	78,2	85,7
	Öncesi	Test	83,6	93,3	72,3	81,5
	Sonrası		87,0	94,8	78,3	85,7
Sonrası	Öncesi	Doğrulama	86,0	93,3	77,6	84,7
	Sonrası		89,0	97,1	80,4	87,9
	Öncesi	Test	86,0	93,3	77,6	84,7
	Sonrası		89,1	97,2	80,6	88,1

5. SONUÇ

Bu çalışmada, akıllı şebekelerde elektrik hırsızlığının tespiti için veri odaklı ve gelişmiş bir makine öğrenmesi yaklaşımı önerilmiştir. Önerilen metodoloji, KA tabanlı eksik veri doldurma, Küresel MAD ile aykırı değer tespiti, K-ortalamalar tabanlı veri dengeleme, karşılıklı bilgiye dayalı özellik seçimi ve Bayes optimizasyon ile iyileştirilmiş LightGBM sınıflandırıcı adımlarını içeren çok aşamalı bir yapıya dayanmaktadır. Elde edilen sonuçlar, önerilen yaklaşımın tüm performans metriklerinde anlamlı iyileştirmeler sağladığını göstermektedir. Boyut indirgeme ve optimizasyon öncesinde doğruluk oranı %83,3–86,0 aralığında iken, bu değer optimizasyon ve boyut indirgeme sonrasında %87,0–89,1 seviyelerine yükselmiştir. Benzer şekilde, kesinlik %92,7'den %97,2'ye, duyarlılık %72,2'den %80,6'ya, F1-skoru ise %81,2'den %88,1'e ulaşmıştır. Bu iyileşmeler, yanlış sınıflandırma oranlarının azaldığını, özellikle pozitif sınıfın daha güvenilir biçimde tespit edildiğini ve modelin genelleme yeteneğinin güçlendiğini ortaya koymaktadır.

Sonuçlar, geliştirilen yöntemin yalnızca akademik açıdan değil, aynı zamanda dağıtım şirketleri için pratik uygulamalarda da kullanılabilecek düzeyde güvenilir olduğunu göstermektedir. Yöntem, farklı kaçak kullanım senaryolarını başarıyla ayırt edebilmiş ve elektrik hırsızlığının tespitinde mevcut yaklaşımlara kıyasla daha yüksek doğruluk sunmuştur. Ayrıca önerilen çerçevenin esnek yapısı, farklı veri setleri ve dağıtım bölgelerine uyarlanabilir olma potansiyeli taşımaktadır.

Gelecek çalışmalar kapsamında, yöntemlerin akıllı şebekelerde gerçek zamanlı karar destek sistemlerine entegre edilmesi, siber saldırı senaryolarıyla birlikte değerlendirilmesi ve açıklanabilir yapay zekâ yaklaşımlarıyla desteklenerek şeffaflığının artırılması hedeflenmektedir. Böylece, elektrik hırsızlığının tespitine yönelik daha dayanıklı, ölçeklenebilir ve güvenilir çözümler geliştirilmesi mümkün olacaktır.

6. TEŞEKKÜR

Bu çalışma, Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi Bilimsel Araştırma Proje Birimi tarafından proje numarası 24803005 ile desteklenmiştir.

7. KAYNAKLAR

1. Kebotogetse, O., Samikannu, R. & Yahya, A. (2022). A concealed based approach for secure transmission in advanced metering infrastructure. *IEEE Access, 10*, 84809-84817.
2. Esmael, A.A., Da Silva, H.H., Ji, T. & da Silva Torres, R. (2021). Non-technical loss detection in power grid using information retrieval approaches: A comparative study. *IEEE Access, 9*, 40635-40648.
3. de Souza Savian, F., Siluk, J.C.M., Garlet, T.B., do Nascimento, F.M., Pinheiro, J.R. & Vale, Z. (2021). Non-technical losses: A systematic contemporary article review. *Renewable and Sustainable Energy Reviews, 147*, 111205.
4. Northeast Group, (2021). *Electricity theft and non-technical losses: Global markets, solutions, and vendors.* https://northeast-group.com/2021/10/20/electricity-theft-non-technical-losses/.
5. Tasdoven, H., Fiedler, B.A. & Garayev, V. (2012). Improving electricity efficiency in Turkey by addressing illegal electricity consumption: A governance approach. *Energy Policy, 43*, 226-234.
6. T.C. Enerji Piyasası Düzenleme Kurumu (EPDK), (2024). *2023 yılı elektrik piyasası gelişim raporu.* https://www.epdk.gov.tr/Detay/Icerik/4-14475/duyuru. Erişim tarihi: 30 Ağustos 2024.

7. Zulu, C.L. & Dzobo, O. (2023). Real-time power theft monitoring and detection system with double connected data capture system. *Electrical Engineering*, *105*(5), 3065-3083.

8. Kawoosa, A.I., Prashar, D., Faheem, M., Jha, N. & Khan, A.A. (2023). Using machine learning ensemble method for detection of energy theft in smart meters. *IET Generation, Transmission & Distribution*, *17*(21), 4794-4809.

9. Abdulaal, M.J., Ibrahem, M.I., Mahmoud, M.M., Khalid, J., Aljohani, A.J., Milyani, A.H. & Abusorrah, A.M. (2022). Real-time detection of false readings in smart grid AMI using deep and ensemble learning. *IEEE Access*, *10*, 47541-47556.

10. Žarković, M. & Dobrić, G. (2024). Artificial intelligence for energy theft detection in distribution networks. *Energies*, *17*(7), 1580.

11. Tripathi, A.K., Pandey, A.C. & Sharma, N. (2024). A new electricity theft detection method using hybrid adaptive sampling and pipeline machine learning. *Multimedia Tools and Applications*, *83*(18), 54521-54544.

12. Zhuang, W., Jiang, W., Xia, M. & Liu, J. (2024). Dynamic generative residual graph convolutional neural networks for electricity theft detection. *IEEE Access*, *12*, 42737-42750.

13. Peng, Y., Yang, Y., Xu, Y., Xue, Y., Song, R., Kang, J. & Zhao, H. (2021). Electricity theft detection in AMI based on clustering and local outlier factor. *IEEE Access*, *9*, 107250-107259.

14. Qi, R., Zheng, J., Luo, Z. & Li, Q. (2022). A novel unsupervised data-driven method for electricity theft detection in AMI using observer meters. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1-10.

15. El-Toukhy, A.T., Badr, M.M., Mahmoud, M.M., Srivastava, G., Fouda, M.M. & Alsabaan, M. (2023). Electricity theft detection using deep reinforcement learning in smart power grids. *IEEE Access*, *11*, 59558-59574.

16. Chen, J., Nanehkaran, Y.A., Chen, W., Liu, Y. & Zhang, D. (2023). Data-driven intelligent method for detection of electricity theft. *International Journal of Electrical Power & Energy Systems*, *148*, 108948.

17. Zhang, W. & Dai, Y. (2024). A multiscale electricity theft detection model based on feature engineering. *Big Data Research*, *36*, 100457.

18. Althobaiti, A., Rotsos, C. & Marnerides, A.K. (2023). Adaptive energy theft detection in smart grids using self-learning with dual neural network. *IEEE Transactions on Industrial Informatics*, *20*(2), 2776-2786.

19. Liao, W., Yang, D., Ge, L., Jia, Y. & Yang, Z. (2025). Electricity theft detection in integrated energy systems considering multi-energy loads. *International Journal of Electrical Power & Energy Systems*, *164*, 110428.

20. Sharma, A. & Tiwari, R. (2024). Anomaly detection in smart grid using optimized extreme gradient boosting with SCADA system. *Electric Power Systems Research*, *235*, 110876.

21. Zhu, S., Xue, Z. & Li, Y. (2024). Electricity theft detection in smart grids based on omni-scale cnn and autoxgb. *IEEE Access*, *12*, 15477-15492.

22. Norouzi, A., Ahmadi, I., Nazari, M., Pourrostami, H. & Hoseini-Kordkheili, H. (2025). A novel framework to detect electricity theft in regions with traditional metering; applying a hybrid machine learning-based approach to low-sampling-rate data. *Results in Engineering*, *25*, 104478.

23. Sun, X., Hu, J., Zhang, Z., Cao, D., Huang, Q., Chen, Z. & Hu, W. (2023). Electricity theft detection method based on ensemble learning and prototype learning. *Journal of Modern Power Systems and Clean Energy*, *12*(1), 213-224.

24. Nirmal, S., Patil, P. & Shinde, S. (2025). Adversarial measurements for convolutional neural network-based energy theft detection model in smart grid. *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, 100909.

25. Ullah, A., Khan, I.U., Younas, M.Z., Ahmad, M. & Kryvinska, N. (2025). Robust resampling and stacked learning models for electricity theft detection in smart grid. *Energy Reports*, *13*, 770-779.

26. Kılınç, E. (2024). Comparison of feature extraction methods in high dimensional time series. *Çukurova Üniversitesi Mühendislik Fakültesi Dergisi, 39*(4), 991-997.

27. CER Smart Metering Project - *Electricity Customer Behaviour Trial, commission for energy regulation (cer), 2009-2010*, https://www.ucd.ie/issda/data/commissionforenergyregulationcer/.

28. Fei, K., Li, Q., Zhu, C., Dong, M. & Li, Y. (2022). Electricity frauds detection in Low-voltage networks with contrastive predictive coding. *International Journal of Electrical Power & Energy Systems*, *137*, 107715.

29. Takiddin, A., Ismail, M., Zafar, U. & Serpedin, E. (2022). Deep autoencoder-based anomaly detection of electricity theft cyberattacks in smart grids. *IEEE Systems Journal*, *16*(3), 4106-4117.

30. K. Zor, (2019). Research and application of real-time short-term electrical energy consumption forecasting using artificial intelligence based techniques. *Ph.D. dissertation*, Cukurova University, Adana.
31. Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764-766.
32. Lohrer, A., Kazempour, D., Hünemörder, M. & Kröger, P. (2024). CoMadOut—a robust outlier detection algorithm based on CoMAD. *Machine Learning*, *113*, 3553-3580.
33. Xu, Z. & Shen, D. (2021). A cluster-based oversampling algorithm combining SMOTE and k-means (KNSMOTE).
34. Vergara, J.R. & Estévez, P.A. (2014). A review of feature selection methods based on mutual information. *Neural Computing and Applications*, *24*(1), 175-186.
35. Snoek, J., Larochelle, H. & Adams, R.P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, *25*, 2960-2968.
36. Bentéjac, C., Csörgő, A. & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review, 54*, 1937-1967.
37. LightGBM Developers, (2025). *LightGBM documentation (Parameters & LGBMClassifier)*.
38. Omotehinwa, T.O., Oyewola, D.O. & Dada, E.G. (2023). A light gradient-boosting machine algorithm with tree-structured parzen estimator for breast cancer diagnosis. *Healthcare Analytics*, *4*, 100218.
39. Atalay, B.A. & Zor, K. (2025). Xgboost (aşırı gradyan artırımlı karar ağaçları) ile hidroelektrik enerji tahmini. *Çukurova Üniversitesi Mühendislik Fakültesi Dergisi, 40*(1), 205-218.
40. Yaro, A.S., Maly, F., Prazak, P. & Malý, K. (2024). Outlier detection performance of a modified z-score method in time-series rss observation with hybrid scale estimators. *IEEE Access*, *12*, 12785-12796.