



Tüketicilerde Kaçak Elektrik Kullanımının Akıllı Sayaç Verisi Üzerinden Gradyan Artırmalı Karar Ağacı Tabanlı Makine Öğrenmesi Yöntemleriyle Tespiti

Yasin Kara¹, Ayşe Aksu¹

(Alınış / Received: 07.04.2023, Kabul / Accepted: 23.05.2023, Online Yayınlanma / Published Online: 26.07.2023)

Anahtar Kelimeler

Akıllı sayaçlar
Güç sistemleri
Kaçak elektrik
Karar ağaçları
Makine öğrenmesi
Teknik olmayan kayıplar

Öz: Elektrik enerjisinin iletimi ve dağıtımında meydana gelen teknik olmayan kayıplardan başlıcası olan kaçak elektrik kullanımı, tüm dünyada karşılaşılan ciddi bir sorundur. Kaçak elektrik kullanımı, toplumun tüm kesimlerine ekonomik ve sosyal zararlar verdiği gibi elektrik arz güvenliğini de tehdit etmektedir. Son zamanlarda dünya genelinde kullanımı giderek artan akıllı sayaç teknolojisi, kaçak elektrik kullanımının veri güdümlü tespiti için yeni olanaklar sağlamaktadır. Bu çalışmada, tüketicilerde gerçekleşen kaçak elektrik kullanımının akıllı sayaç verileri üzerinden tespiti amacıyla, gradyan artırmalı karar ağacı tabanlı üç makine öğrenmesi modeli ele alınmıştır. Bu anlamda, LightGBM, XGBoost ve CatBoost modelleri kullanılmış ve modeller, akıllı sayaçlardan okunan tarihsel günlük elektrik tüketim verileri üzerinden ilgili tüketicinin elektrik kullanımını "yasal kullanım" veya "kaçak kullanım" olarak sınıflandırması amacıyla eğitilmiştir. Çalışmada, Çin Devlet Elektrik Şirketi'nin yayımladığı, eksik ve dengesiz veriler içeren gerçek akıllı sayaç veri kümesi kullanılmıştır. Başarım denetleme çalışmaları, CatBoost modelinin doğruluk ve AUC ölçevlerinde sırasıyla %78,10 ve %74,17, LightGBM modelinin ise duyarlılık ölçevinde %72,48 değeri ile diğer modellere kıyasla daha yüksek başarımlı olduğunu göstermiştir.

Detection of Electricity Theft in Consumers through Gradient Boosting Decision Tree-based Machine Learning Methods using Smart Meter Data

Keywords

Smart meters
Power systems
Electricity theft
Decision trees
Machine learning
Non-technical losses

Abstract: Electricity theft, which is one of the major non-technical losses in the transmission and distribution of electricity, is a serious problem encountered worldwide. Electricity theft causes economic and social losses to all segments of society and also threatens the security of electricity supply. Recently, increasing use of smart meter technology worldwide provides new opportunities for data-driven detection of electricity theft. In this work, three machine learning models based on gradient boosting decision trees, LightGBM, XGBoost, and CatBoost, are adopted for the detection of electricity theft in consumers through smart meter data. The models are trained to classify the consumer's electricity usage as "legal usage" or "theft

¹ Çözümevi Yönetim Danışmanlığı ve Bilgisayar Yazılım Ticaret A.Ş., Orta Mah. Ordu Sok. İzpark Plaza No: 23/A İç Kapı No: 13, 34880 Kartal İstanbul/Türkiye

usage" using historical daily electricity consumption data obtained from smart meters. In the study, a real smart meter dataset published by the State Grid Corporation of China, which contains missing and unbalanced data, is used. The performance evaluation studies reveal that the CatBoost model achieved higher accuracy and AUC of 78.10% and 74.17%, respectively, while the LightGBM model achieved higher recall value of 72.48%, compared to the other models.

1. Giriş

Elektrik enerjisinin üretiminden başlayarak son kullanıcıya ulaştırılmasına kadar gerçekleşen tüm süreçlerde meydana gelen enerji kayıpları, tamamen bertaraf edilmesi mümkün olmayan bir olgudur. Elektrik enerjisi kayıpları, teknik ve teknik olmayan kayıplar olarak iki sınıfta incelenebilir. Teknik kayıplar temel olarak, iletim hattı, kablo, transformatör ve sayaç gibi elektriksel elemanların direnci nedeniyle açığa çıkan ısı enerjisinden ileri gelmektedir. Teknik kayıplar, verimli eleman seçimi ve uygun planlama ile en aza indirilebilse de tamamen sıfırlanması mümkün olmayan fiziksel bir fenomendir.

Teknik olmayan kayıplar ise, elektrik hırsızlığı (kaçak elektrik kullanımı), sayaçların hatalı çalışması ya da okunması veya eksik faturalama gibi durumların sonucudur [1]. Ülkemizde yürürlükte olan 6446 numaralı Elektrik Piyasası Kanunu, teknik ve teknik olmayan kayıpları,

"Dağıtım sistemine giren enerji ile dağıtım sisteminde tüketicilere tahakkuk ettirilen enerji miktarı arasındaki farkı oluşturan ve maliyeti etkileyen; teknik kayıp ve/veya kaçak kullanım gibi sebeplerden kaynaklanan ve teknik bir sebebe dayanmayan kayıp"

olarak tanımlamıştır [2]. Teknik olmayan kayıpların, teknik kayıpların aksine, sıfıra indirilmesi kuramsal olarak mümkündür [3].

Teknik olmayan kayıplardan başlıcası olan kaçak elektrik kullanımı, tüm dünyada, özellikle de gelişmekte olan ülkelerde karşılaşılan ciddi bir sorundur. Yapılan araştırmalar, dünya genelinde kaçak elektrik kullanımının yılda yaklaşık 90 milyar \$ maddi zarar meydana getirdiğini ortaya koymaktadır [4]. TEDAŞ verilerine göre Türkiye genelinde 2019 yılında elektrik enerjisi kayıpları oranı %12,7 olarak gerçekleşmiştir [5]. Teknik olmayan kayıpların bu yüzde içindeki ağırlığı belirtilmemiş olsa da kaçak elektrik kullanımının özellikle kimi illerde %60 mertebelerine ulaştığı bilinmektedir [6].

Kaçak elektrik kullanımı bir yandan toplam elektrik üretim yükünü, bir yandan da meşru (yasal) elektrik tüketicilerinin enerji giderlerini (kayıp/kaçak bedeli benzeri kalemler üzerinden) arttırdığından, toplumun hemen her kesimine ekonomik zararlar vermektedir [7]. Bunların ötesinde, kaçak kullanım için elektrik tesisatı üzerinde yapılan sayaç baypası ve kaçak hat çekilmesi gibi yetkisiz müdahaleler, elektrik tüketicileri için hayati tehlike oluşturmaktadır [8,9]. Ayrıca ülkemizde yapılan bir çalışmada, kaçak elektrik kullanımının enterkonnekte sistem genelinde gerilim çökmesi gibi enerji kalitesi problemlerine yol açabildiği gösterilmiştir [10]. Bu bilgilerden hareketle, kaçak elektrik kullanımının önlenmesinin, ekonomik ve toplumsal yitimlerin önlenmesi açısından büyük önem arz ettiği görülmektedir. Kaçak kullanımların en aza indirilebilmesi için atılacak adımların başında, süregelen bir kaçak elektrik kullanımının tespit edilmesi gelmektedir.

Dünyada kullanımı giderek artan akıllı sayaç (smart meter) teknolojisi, tüketici-şebeke işletmecisi arası iki yönlü haberleşme ile yüksek doğruluk ve sıklıkta veri aktarımına imkân sağlamaktadır [11,12]. Böylelikle, şebeke işletmecileri tarafından temin edilen tüketim verileri, veri güdümlü yöntemlerle işlenerek, tüketici tarafında var olan çeşitli anomalilerin tespiti için kullanılabilir [13]. Bu anlamda, son yıllarda başarımları giderek artan ve hemen her alanda yaygın kullanım alanı bulunan makine öğrenmesi yaklaşımları, tüketici tarafında meydana gelen kaçak elektrik kullanımlarının akıllı sayaç verisine dayalı olarak tespiti için uygun bir aday haline gelmiştir. Makine öğrenmesi modelleri, akıllı sayaçlardan elde edilen tüketim verileri ile eğitilerek, kaçak ve yasal kullanım örüntülerini öğrenebilir [14]. Eğitilen makine öğrenmesi modeli, eğitim kümesinde bulunmayan tüketim verileri üzerinde tahmin yaparak kaçak elektrik kullanımını tespit amacıyla kullanılabilir.

Kaçak elektrik kullanımının makine öğrenmesi modelleri ile tespiti problemi, özellikle son yıllarda literatürde geniş çaplı olarak irdelenmiştir. Bu anlamda, gözetimli (supervised), yarı-gözetimli (semi-supervised) ve gözetimsiz (unsupervised) öğrenme problemleri kurgulanmış ve çeşitli makine öğrenmesi modelleri ele alınmıştır [15].

Kullanılan gözetimli öğrenme modelleri arasında, yapay sinir ağı tabanlı uzun kısa-dönem hafızalı sinir ağları [15], çok katmanlı derin sinir ağları [16], evrişimli sinir ağları [17] ve yinelemeli sinir ağları [18] bulunmaktadır. Öte

yandan, karar ağaçları [19], rastsal ormanlar [20] ve destek vektör makineleri [21] gibi modeller de ele alınmıştır. Yarı gözetimli problemler için RDAE-AG-TripleGAN [22] ve derin öğrenme [23] tabanlı hibrid modeller önerilmiştir. Kurgulanan gözetimsiz öğrenme problemlerinin çözümünde ise hiyerarşik kümeleme [24] ve k-ortalamar [25] gibi kümeleme yöntemleri kullanılmıştır.

Bu çalışmada, tüketicilerde gerçekleşen kaçak elektrik kullanımlarının akıllı sayaç verileri üzerinden tespiti amacıyla, gradyan artırılmış karar ağacı (gradient boosting decision trees) tabanlı üç ayrı makine öğrenmesi modeli önerilmiştir. Bu anlamda, LightGBM, XGBoost ve CatBoost modelleri ele alınmıştır. Modeller, akıllı sayaçlardan okunan tarihsel günlük elektrik tüketim verileri üzerinden, ilgili tüketicinin elektrik kullanımını “yasal kullanım” veya “kaçak kullanım” olarak sınıflandırması amacıyla eğitilmiştir. Çalışmada, Çin Devlet Elektrik Şirketi’nin yayımladığı gerçek akıllı sayaç verileri kullanılmıştır.

Makalenin geri kalan kısımları şu şekilde düzenlenmiştir: Kısım 2’de çalışmada kullanılan veri kümesi tanıtılmış, ele alınan makine öğrenmesi modelleri ile bu modellerin eğitimi, üstparametre optimizasyonu, sınanması ve başarımlarını değerlendirme süreçleri detaylandırılmıştır. Kısım 3’te üstparametre optimizasyonu ve sınanma kümesi üzerindeki bulgular paylaşılmıştır. Son olarak Kısım 4’te tartışma ve sonuçlara yer verilmiştir.

2. Materyal ve Metot

2.1. Veri kümesi

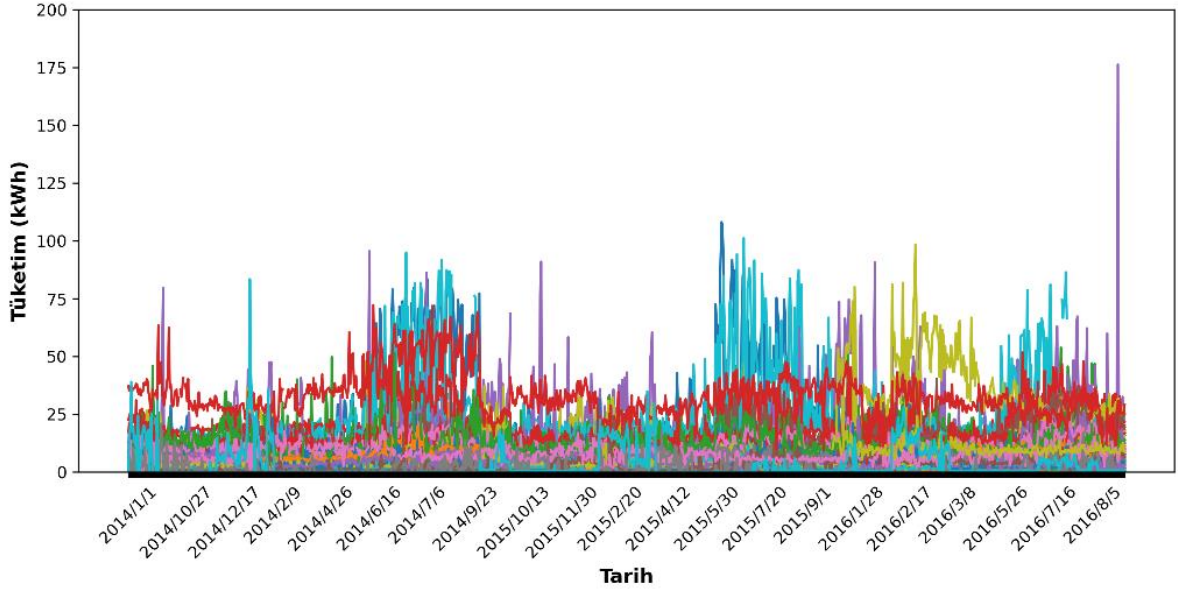
Çalışmaya konu olan kaçak elektrik kullanımının akıllı sayaç verisi üzerinden tespiti problemi, bir sınıflandırma problemi olarak kurgulanmıştır. Bir gözetimli öğrenme problemi olan sınıflandırma problemlerinde, makine öğrenmesi modelinin eğitimi için etiketlenmiş bir veri kümesine ihtiyaç vardır. Bu amaçla, [26]’da sunulan veri kümesinden faydalanılmıştır. Veri kümesi, Çin Devlet Elektrik Şirketi’nin [27] yayımladığı gerçek akıllı sayaç verilerini içermektedir. Veri kümesine, ilgili GitHub çevrimiçi deposundan [28] erişilebilmektedir.

Bu veri kümesinde, 42.374 elektrik tüketicisinden 2014 – 2016 yılları arasında 1.034 gün boyunca alınan günlük elektrik enerjisi tüketimi verileri yer almaktadır. Dolayısıyla her bir tüketici bir örnek, her bir günlük tüketim verisi ise bir öznitelik olarak alınmıştır. Veri kümesinde ayrıca, her bir tüketiciye ilişkin “yasal kullanım” (sınıf 0) ve “kaçak kullanım” (sınıf 1) etiketi yer almaktadır. Veri kümesine ilişkin çeşitli istatistikler, Tablo 1’de sunulmuştur. Yasal ve kaçak kullanımlara ilişkin 100’er adet veri örneği, sırasıyla Şekil 1 ve 2’de gösterilmiştir.

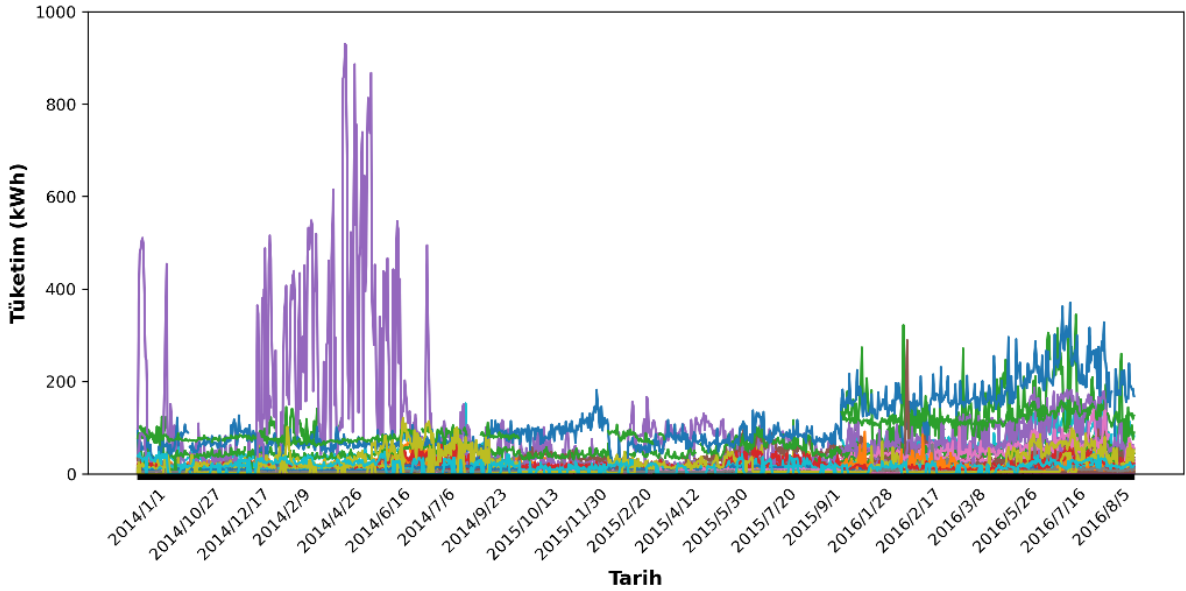
Tablo 1. Kullanılan veri kümesine ilişkin istatistikler

Örnek sayısı	Öznitelik Sayısı	Sınıf sayısı (etiketler)	Eksik veri oranı	Kaçak/yasal kullanım örnek oranı
42.374	1.034	2 (kaçak/yasal)	1/3,90	1/10,72

Tablo 1’de görüldüğü üzere veri kümesinde pek çok eksik veri (missing data) bulunmaktadır, yaklaşık her dört veri hücresinden biri eksiktir. Ayrıca, bekleneceği üzere, kaçak kullanıma ilişkin örnekler, yasal kullanıma ilişkin örneklerden çok daha azdır. Bu durum, veri dengesizliği (imbalanced data) durumunu doğurmaktadır. Eksik veri ve veri dengesizliği koşullarında makine öğrenmesi modellerinin eğitiminin güçleştiği ve başarımlarının düştüğü bilinmektedir [29,30]. Dahası, kimi makine öğrenmesi modellerinin (örneğin lojistik regresyon, yapay sinir ağları, destek vektör makineleri) eğitimi, eksik verinin bulunduğu veri kümeleri ile gerçekleştirilememekte ve eksik veri doldurma (imputation) yöntemlerine ihtiyaç duyulmaktadır.



Şekil 1. Yasal kullanım sınıfına ilişkin 100 adet örnek veri.



Şekil 2. Kaçak kullanım sınıfına ilişkin 100 adet örnek veri.

2.2. Ele alınan makine öğrenmesi modelleri

Eldeki veri kümesinin özellikleri (eksik ve dengesiz veri koşulları) göz önüne alındığında, gradyan artırmalı karar ağacı (gradient boosting decision trees – GBDT) tabanlı makine öğrenmesi modellerinin, problemin çözümünde etkin olacağı düşünülmüştür. GBDT tabanlı modeller, eksik veri varlığında dahi eğitilebilmekte ve veri doldurma yöntemlerine gerek duymamaktadır [31]. Ayrıca yapılan deneysel çalışmalar, GBDT tabanlı modellerin dengesiz veri koşullarında yüksek başarımlar sağladığını göstermiştir [32].

GBDT tabanlı yöntemler, birer topluluk öğrenme (ensemble learning) modelidir [33]. Bu modellerde, karar ağacı tabanlı pek çok “zayıf öğrenici” eğitilir. Eğitim algoritmasının temel mantığı, öğrencilerin hatalarına odaklanılarak topluluğa yeni karar ağaçlarının eklenmesine dayanır. Özellikle, topluluktaki her yeni ağaç, önceki ağaçların artıkları (yani, tahmin edilen ve gerçek değerler arasındaki farklar) üzerinde eğitilir. Eğitim sonunda elde edilen GBDT tabanlı modelin nihai tahmini, topluluktaki tüm ağaçların tahminlerinin bir kombinasyonu olur.

Literatürde pek çok farklı GBDT tabanlı makine öğrenmesi modeli önerilmiştir. Bu çalışmada, bu modellerden üç tanesi ele alınmıştır. Bunlar, Microsoft tarafından geliştirilen LightGBM [34], Chen ve Guestrin tarafından önerilen XGBoost [35] ve Yandex tarafından geliştirilen CatBoost [36] modelleridir.

2.3. Model eğitimi ve üstparametre optimizasyonu

Makine öğrenmesi modellerinin eğitimi için, eldeki veri kümesinin %80'i eğitim kümesi olarak alınmış, geri kalan %20'lik bölümü sınav kümesi olarak ayrılmıştır. Model eğitimine geçilmeden önce, model başarımını önemli ölçüde etkileyen üstparametre (hyperparameter) değerlerine karar verilmesi gerekmektedir. Bu amaçla, her bir model için üç ayrı üstparametre ele alınmış ve bunlar için ayrık ve sonlu bir üstparametre uzayı oluşturulmuştur (bkz. Tablo 2).

Tablo 2. Ele alınan üstparametreler ve üstparametre uzayları.

Model	Üstparametre 1	Üstparametre 1 uzayı	Üstparametre 2	Üstparametre 2 uzayı	Üstparametre 3	Üstparametre 3 uzayı
LightGBM	max_depth	{4, 8, 16}	num_leaves	{8, 16, 32}	class_weight	{1, 5, 10, 15}
XGBoost	max_depth	{4, 8, 16}	eta	{0,1, 0,3, 0,5}	scale_pos_weight	{1, 5, 10, 15}
CatBoost	max_depth	{4, 6, 8}	learning_rate	{0,03, 0,05, 0,1}	scale_pos_weight	{1, 5, 10, 15}

Tablo 2'de verilen "max_depth", karar ağacı derinliği için, "num_leaves" ise ağaçlardaki yaprak sayısı için bir üst sınır belirler. Bu değerler model karmaşıklığını düzenler. Öte yandan, "eta" ve "learning_rate", öğrenme hızını belirlemektedir. Yüksek öğrenim hızı, eğitim sürecini hızlandırabilir ancak başarımı düşürebilir. Son olarak LightGBM için "class_weight" ve XGBoost ile CatBoost için "scale_pos_weight" üstparametreleri, az sayıda örnek içeren sınıfa (yani, kaçak kullanım sınıfına) bir ağırlık değeri atayarak, var olan veri dengesizliği durumunun yarattığı başarım düşümünü bertaraf etme amacıyla kullanılmaktadır.

Oluşturulan üstparametre uzayında, eğitim kümesinin %20'si kullanılmış ve ayırma geçerlemesi (hold-out validation) ile ızgara arama (grid search) yöntemi kullanılarak optimizasyon gerçekleştirilmiştir. Tablo 2'de görüldüğü üzere, her bir makine öğrenmesi modeli için, her biri farklı üstparametre kombinasyonu içeren 36'şar adet aday model söz konusudur. Bu aday modeller, farklı başarım ölçüleri kullanılarak sınanmış ve en iyi ölçü değerlerini veren aday modeller "nihai model" olarak alınmıştır. Bu amaçla ele alınan başarım ölçüleri, Kısım 2.4'te ifade edilmiştir.

2.4. Başarım denetleme ölçüleri

Bu çalışmada, sınıflandırıcıların başarımlarının denetlenmesi için üç ayrı başarım ölçüleri ele alınmıştır. Bunlar, doğruluk (accuracy), duyarlılık (recall) ve AUC (ROC eğrisi altında kalan alan – Area Under the ROC Curve) olarak belirlenmiştir. Her üç ölçü de 0 ile 1 arası değer alır ve daha yüksek değerler daha iyi başarımları gösterir.

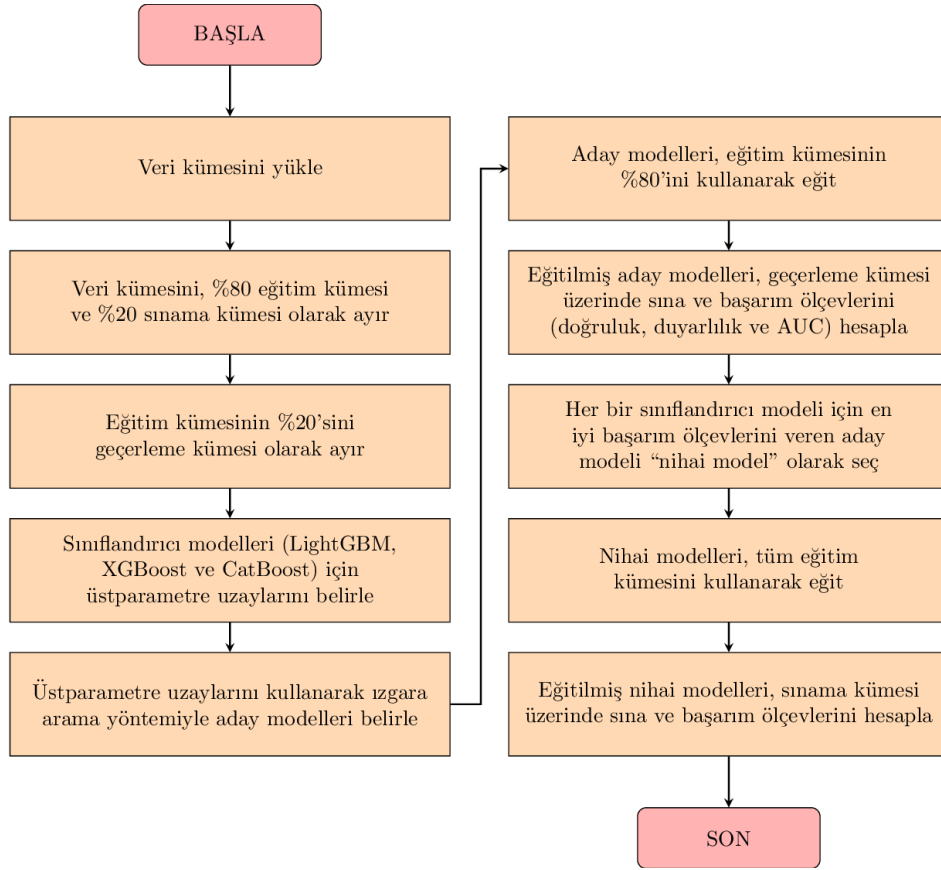
Doğruluk ve duyarlılık değerleri, sırasıyla (1) ve (2) 'de verilen ifadeler ile hesaplanabilir:

$$\text{Doğruluk} = \frac{DY + DK}{DY + DK + YY + YK} \quad (1)$$

$$\text{Duyarlılık} = \frac{DK}{DK + YY} \quad (2)$$

Burada DY ve DK, sırasıyla doğru olarak tahmin edilmiş yasal ve kaçak kullanımları; YY ve YK ise sırasıyla yanlış olarak tahmin edilmiş yasal ve kaçak kullanımları ifade etmektedir. AUC ölçüleri ise, ROC (Receiver Operating Characteristic – alıcı işletim karakteristiği) eğrisi altında kalan alanı ifade eder. ROC eğrisi, doğru pozitif oranı ve yanlış pozitif oranı arasındaki ilişkiyi tarifleyen bir eğridir. AUC ise, ROC eğrisi çizildikten sonra sayısal integrasyon yöntemleri ile hesaplanır.

Çalışma kapsamında model geliştirme ve sınav süreçlerinde izlenen metodoloji, Şekil 3'te verilen akış şeması ile özetlenmiştir.



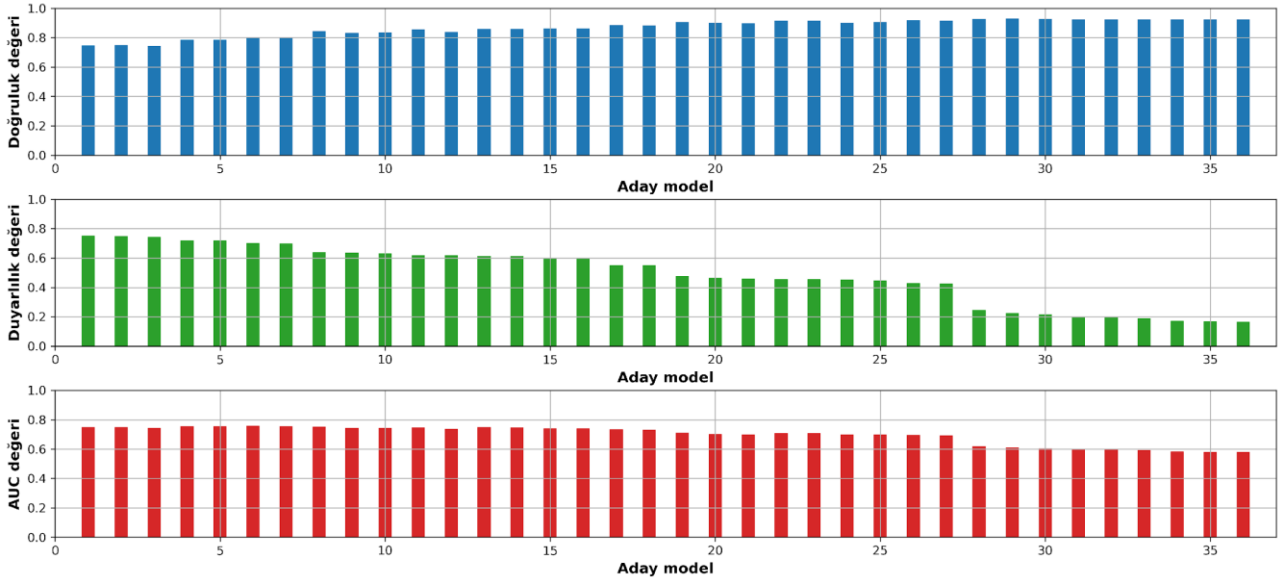
Şekil 3. Model geliştirme ve sına metodolojisine ilişkin akış şeması.

3. Bulgular

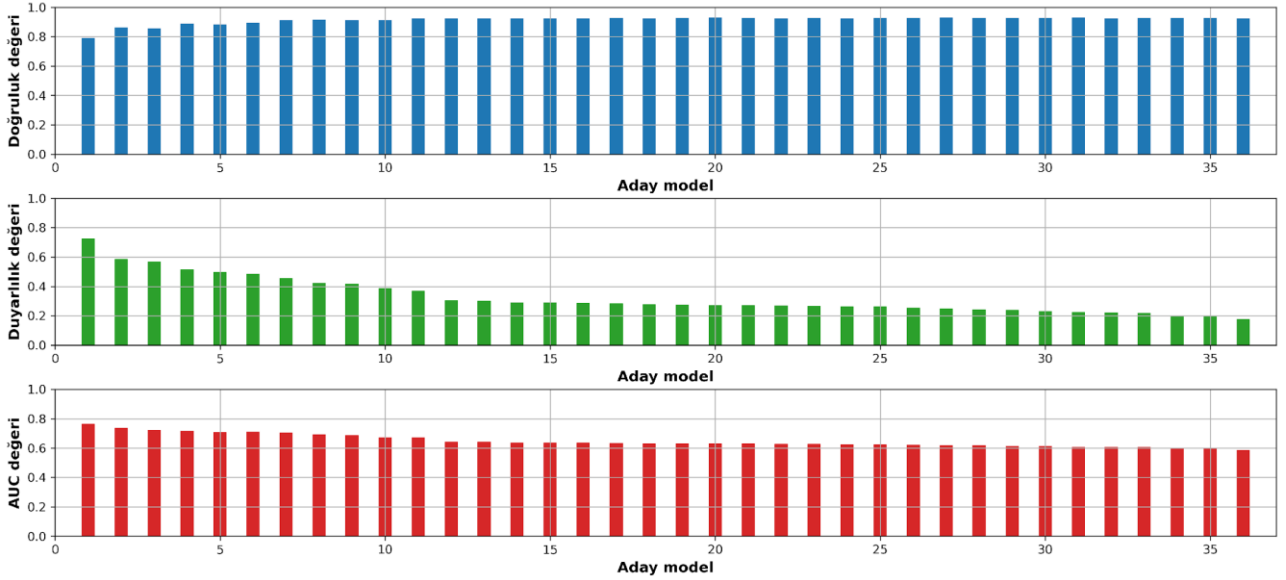
Bu kısımda, üstparametre optimizasyonu süreci ile edinilen aday model başarıml değerleri, seçilen nihai modeller ve bu nihai modellerin sına kümesi üzerindeki başarımları gösterilmiştir. Model geliştirme süreçleri Python ortamında pandas, numpy, scikit-learn, lightgbm, xgboost ve catboost kütüphaneleri kullanılarak gerçekleştirilmiştir.

3.1. Üstparametre optimizasyonu bulguları

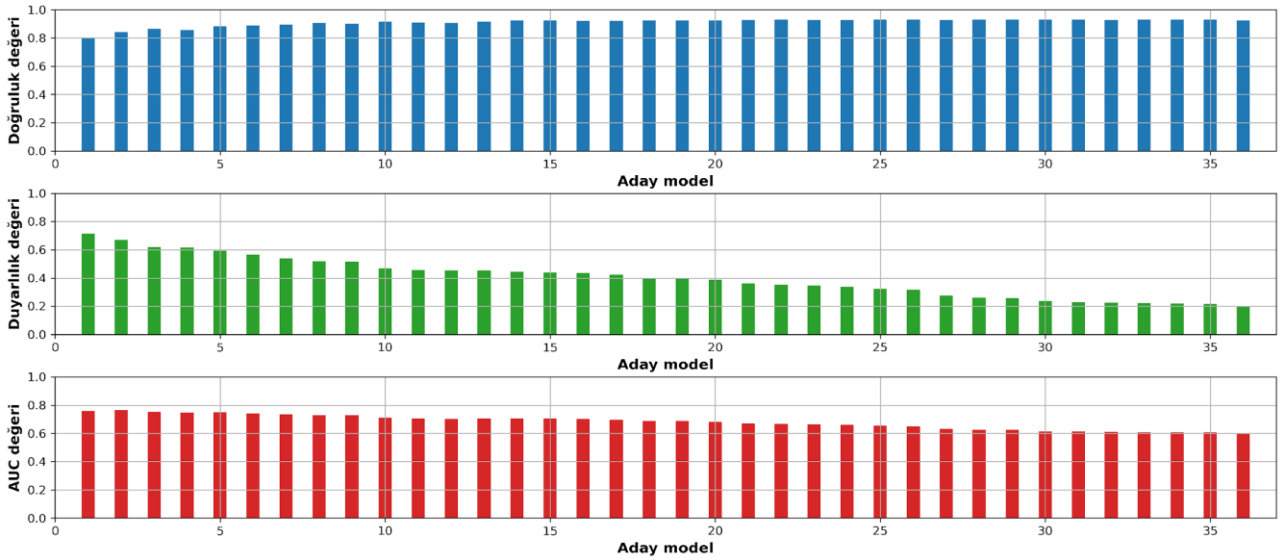
LightGBM, XGBoost ve CatBoost için farklı üst parametre değerleri üzerinden oluşturulan 36'şar adet aday modelin, önceki kısımda ifade edilen başarıml ölçveleri kullanılarak elde edilen geçerleme sonuçları, sırasıyla Şekil 4, 5 ve 6'da verilmiştir. Aday modeller, duyarlılık ölçveine göre azalan şekilde sıralanmıştır.



Şekil 4. LightGBM aday modelleri için geçerleme sonuçları.



Şekil 5. XGBoost aday modelleri için geçerleme sonuçları.



Şekil 6. CatBoost aday modelleri için geçerleme sonuçları.

Verilen grafiklere bakıldığında, aynı model için farklı üstparametrelerin kullanılması, başarımlarını %70'e varan mertebelerde değiştirmektedir. Bu da üstparametre optimizasyonunun önemini göstermektedir. Söz konusu veriler incelendiğinde, LightGBM, XGBoost ve CatBoost için 1 numaralı aday modellerin, her üç başarı ölçü de düşünülüğünde, uygun ve dengeli başarımlarını verdiği sonucuna varılmış ve bu aday modellerin nihai model olarak alınmasına karar verilmiştir. Seçilen nihai modeller için geçirme sonuçları ile ilgili çeşitli detaylar, Tablo 3'te sunulmuştur.

Tablo 3. Seçilen nihai modeller için geçirme sonuçları
(en iyi değerler, kalın italik olarak verilmiştir).

Model	Üstparametre 1 değeri	Üstparametre 2 değeri	Üstparametre 3 değeri	Doğruluk	Duyarlılık	AUC
LightGBM	4	8	15	%74,65	%75,26	%74,92
XGBoost	4	0,1	15	%79,26	%72,66	%76,27
CatBoost	4	0,03	15	%79,48	%71,45	%75,84

Tablo 3'teki veriler incelendiğinde, doğruluk ölçüde CatBoost, duyarlılık ölçüde LightGBM ve AUC ölçüde ise XGBoost modelinin en yüksek başarımlarını verdiği görülebilir. Bununla beraber, nihai modellerin başarımları, modellerin tüm eğitim kümesinde eğitilmesi sonrası sınama kümesi üzerinde sınanması ile bir sonraki kısımda verilmiştir.

3.2. Sınama kümesi üzerinde başarımların denetleme

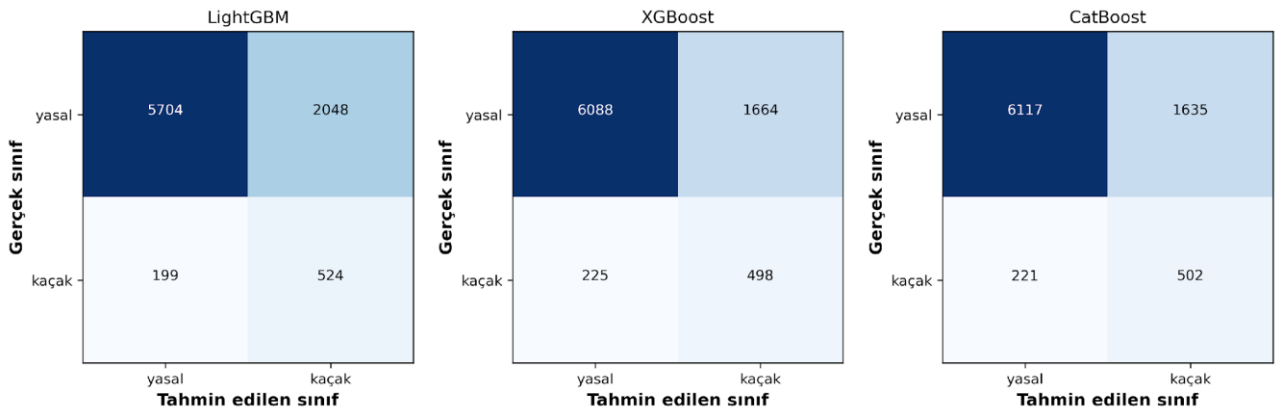
Elde edilen nihai modeller, üstparametre optimizasyonu ve eğitim süreçlerinde kullanılmayan %20'lik sınama kümesi üzerinde sınanmış ve başarımları denetlenmiştir. Gözlemlenen başarımlar ölçü değerleri ve modellerin eğitim süreleri, Tablo 4'te sunulmuştur.

Tablo 4. Nihai modeller için sınama kümesi üzerinde elde edilen değerler
(en iyi değerler, kalın italik olarak verilmiştir)

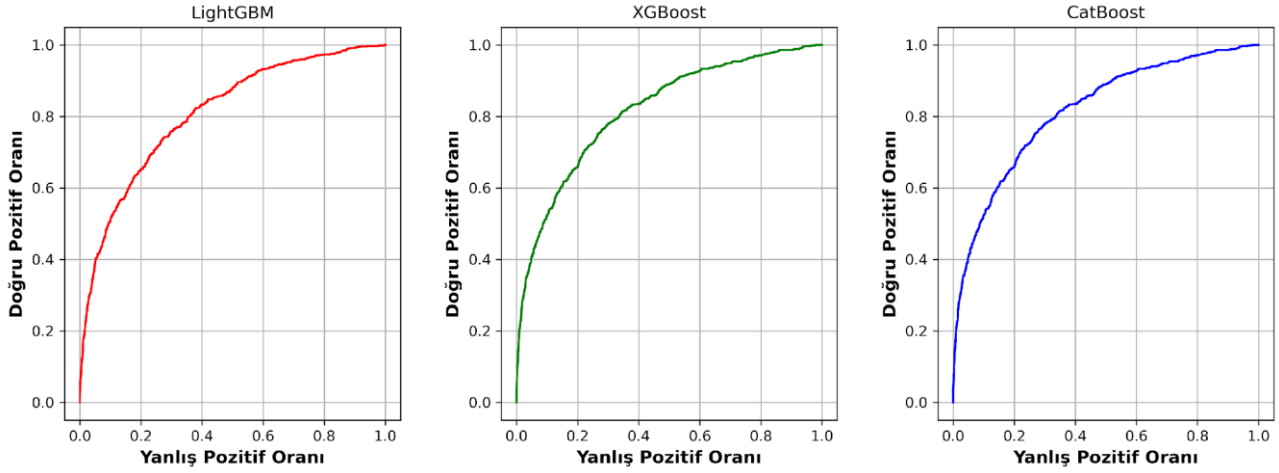
Model	Doğruluk	Duyarlılık	AUC	Eğitim süresi (sn.)
LightGBM	%73,49	%72,48	%73,02	4,38
XGBoost	%77,71	%68,88	%73,70	37,79
CatBoost	%78,10	%69,43	%74,17	66,20

Verilere bakıldığında, CatBoost tabanlı sınıflandırıcı modelinin doğruluk ve AUC ölçüde de diğer iki modelden daha yüksek değerler elde ettiği görülebilir. CatBoost diğer modellere, doğruluk ölçüde 4,61 puan ve AUC ölçüde 1,15 puana varan bir üstünlük sağlamıştır. Duyarlılık ölçüde ise en iyi değere LightGBM modeli ulaşmıştır.

Ele alınan ölçülerde CatBoost modelinin başarımlarını öne çıkarsa da bu model uzun eğitim süresi ile göze çarpmaktadır. LightGBM'in yaklaşık 15 katı bir işlem süresi yükü olan CatBoost eğitiminin, milyonlarca örnek içeren bir veri kümesi varlığında günler sürebileceği tahmin edilebilir. Bu durumda GPU tabanlı eğitim yapılabilir veya daha az işlem yükü olan LightGBM modeli tercih edilebilir. Son olarak, her üç model için de elde edilen doğruluk matrisleri Şekil 7'de, ROC eğrileri ise Şekil 8'de verilmiştir.



Şekil 7. Nihai modeller için sınama kümesi üzerinde elde edilen doğruluk matrisleri.



Şekil 8. Nihai modeller için sına kümesi üzerinde elde edilen ROC eğrileri.

Elde edilen bulgulardan hareketle, her üç model kullanılarak elde edilen başarımların değerlerinin, özellikle veri kümesinde yaklaşık 1/4 değerindeki eksik veri ve 1/10 değerindeki veri dengesizliği oranları da düşünüldüğünde, kaçak elektrik kullanımlarının tespiti için uygun nitelikte olduğu değerlendirilebilir. Enerji dağıtım şirketleri, bu modelleri kullanarak kaçak elektrik kullanım şüphesi bulunan tüketicileri tespit edebilir, daha sonra sayaç ve tesisat incelemesi gibi işlemleri başlatıp kaçak elektrik kullanımının var olup olmadığı konusunda nihai karara varabilir. Bu sayede, hem milyonlarca tüketiciden elde edilen yüksek boyutlu verilerin analizi otomatikleştirilerek insan gücü kaynağı etkin kullanılmış, hem de kaçak elektrik kullanımından doğan ekonomik ve sosyal kayıplar en aza indirilmiş olur.

4. Tartışma ve Sonuç

Bu çalışmada, tüketicilerde kaçak elektrik kullanımının akıllı sayaçlardan alınan tarihsel günlük elektrik tüketimi verilerine dayalı olarak tespiti için üç farklı gradyan artırılmış karar ağacı tabanlı makine öğrenmesi sınıflandırıcı modeli önerilmiştir. Ele alınan modeller olan LightGBM, XGBoost ve CatBoost sınıflandırıcılarının üstparametreleri optimize edilmiş, bu süreçte hesaplanan üç farklı başarımların ölçümleri göz önüne alınarak birer nihai model elde edilmiştir.

Nihai modeller kullanılarak sına kümesinde yapılan başarımların denetleme çalışmalarında, CatBoost tabanlı sınıflandırıcı modelinin doğruluk ve AUC ölçümlerinde sırasıyla %78,10 ve %74,17 değerleri ile, LightGBM modelinin ise duyarlılık ölçümünde %72,48 değeri ve 4,38 saniyelik eğitim süresi ile diğer modellere kıyasla daha yüksek başarımlı olduğu sonucuna varılmıştır. Elde edilen bulgular, akıllı sayaçlardan alınan verilerin makine öğrenmesi yöntemleri kullanılarak işlenmesiyle kaçak elektrik kullanımının yüksek doğrulukla tespit edilebileceğini gösterir niteliktedir. Bu tespit işlemi, eksik ve dengesiz veri koşullarında dahi başarıyla gerçekleştirilebilmektedir.

Gelecekte yapılacak çalışmalarda, sınıflandırma başarımlarının daha da artırılmasına odaklanılarak, işaret işleme tabanlı veri ön işleme yöntemleri, dengesiz veri durumunun üstesinden gelmek için aşırı örnekleme yöntemleri ve birden fazla makine öğrenmesi modelinin bir araya getirilmesiyle oluşturulacak topluluk modelleri gibi yaklaşımlar kullanılacaktır. Dünya genelinde kullanımı giderek artan akıllı sayaç teknolojisinin ülkemizde de yaygınlaşmasıyla beraber, toplumun her kesimini maddi manevi kayba uğratan kaçak elektrik kullanımının düşürülmesinde önemli rol oynayacağına inanıyoruz.

Teşekkür

Bu çalışma T.C. Sanayi ve Teknoloji Bakanlığı tarafından AGTMPR94340 no'lu proje ile desteklenmiştir. Çalışma boyunca desteklerini esirgemeyen Çözüm evi Ar-Ge Merkezi Direktörü Sn. Hakan Çolak, Çözüm evi Ar-Ge Merkezi Araştırmacıları Sn. Şükran Batur, Sn. Mustafa Özkan ve Sn. Burak Müderrisoğlu'na ve tüm Çözüm evi Yönetim Danışmanlığı ve Bilgisayar Yazılım Ticaret A.Ş. ailesine teşekkürlerimizi sunarız.

Yazar Katkı Oranları

Y. Kara – Fikir, tasarım, denetleme, kaynaklar, veri toplama/işleme, analiz, literatür tarama, yazım, eleştirel inceleme. A. Aksu – fikir, tasarım, denetleme, kaynaklar, yazım, eleştirel inceleme.

Çıkar Çatışması

Tüm yazarlar, Çözümevi Yönetim Danışmanlığı ve Bilgisayar Yazılım Ticaret A.Ş. firmasında çalışmaktadır.

Kaynakça

- [1] 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.
- [2] Enerji Piyasası Düzenleme Kurumu (EPDK). (2022). Güncel 6446 Sayılı Elektrik Piyasası Kanunu. <https://www.epdk.gov.tr/Detay/Icerik/3-14650/guncel-6446-sayili-elektrik-piyasasi-kanunu> (Erişim tarihi: 01.04.2023).
- [3] Yıldız, E., & Çetinkaya, N. (2022). Elektrik güç sistemlerindeki kaçak kullanımların yapay sinir ağları ile tahmini. *Journal of Investigations on Engineering and Technology*, 5(1), 1-10.
- [4] Biswas, P. P., Cai, H., Zhou, B., Chen, B., Mashima, D., & Zheng, V. W. (2019). Electricity theft pinpointing through correlation analysis of master and individual meter readings. *IEEE Transactions on Smart Grid*, 11(4), 3031-3042.
- [5] Türkiye Elektrik Dağıtım Anonim Şirketi (TEDAŞ). (2020). 2020 Yılı Türkiye Elektrik Dağıtım Sektör Raporu. https://www.tedas.gov.tr/sx.web.docs/tedas/docs/Stratejikplan/2020_Yili_Turkiye_Elektrik_Dagitimi_Sektor_Raporu.pdf (Erişim tarihi: 01.04.2023).
- [6] T.C. Hakkari Valiliği. (2020). İlimiz Türkiye Genelinde Kaçak Elektriğin En Çok Kullanıldığı İlk 5 il arasında Yer Alıyor. <http://www.hakkari.gov.tr/ilimiz-turkiye-genelinde-kacak-elektrigin-en-cok-kullanildigi-ilk-5-il-arasinda-yer-aliyor> (Erişim tarihi: 01.04.2023).
- [7] Hasan, M. N., Toma, R. N., Nahid, A. A., Islam, M. M., & Kim, J. M. (2019). Electricity theft detection in smart grid systems: A CNN-LSTM based approach. *Energies*, 12(17), 3310.
- [8] Leite, J. B., & Mantovani, J. R. S. (2016). Detecting and locating non-technical losses in modern distribution networks. *IEEE Transactions on Smart Grid*, 9(2), 1023-1032.
- [9] Haq, E. U., Pei, C., Zhang, R., Jianjun, H., & Ahmad, F. (2023). Electricity-theft detection for smart grid security using smart meter data: A deep-CNN based approach. *Energy Reports*, 9, 634-643.
- [10] Mertoğlu, Z., & Tezcan, S. S. (2019). Kaçak Elektrik Kullanımının Enterkonnekte Sisteme Etkileri. *El-Cezeri*, 6(3), 571-584.
- [11] Otuoze, A. O., Mustafa, M. W., Abioye, A. E., Sultana, U., Usman, A. M., Ibrahim, O., ... & Abu-Saeed, A. (2022). A rule-based model for electricity theft prevention in advanced metering infrastructure. *Journal of Electrical Systems and Information Technology*, 9(1), 1-17.
- [12] Kocaman, B. (2018). Teknik Olmayan Enerji Kayıplarının Azaltılmasında PLC Sayaçlarının Önemi. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, 7(2), 220-230.
- [13] Yan, Z., & Wen, H. (2021). Performance analysis of electricity theft detection for the smart grid: An overview. *IEEE Transactions on Instrumentation and Measurement*, 71, 1-28.
- [14] Huang, Y., & Xu, Q. (2021). Electricity theft detection based on stacked sparse denoising autoencoder. *International Journal of Electrical Power & Energy Systems*, 125, 106448.
- [15] Adil, M., Javaid, N., Qasim, U., Ullah, I., Shafiq, M., & Choi, J. G. (2020). LSTM and bat-based RUSBoost approach for electricity theft detection. *Applied Sciences*, 10(12), 4378.
- [16] Lepolesa, L. J., Achari, S., & Cheng, L. (2022). Electricity theft detection in smart grids based on deep neural network. *IEEE Access*, 10, 39638-39655.

- [17] Pereira, J., & Saraiva, F. (2021). Convolutional neural network applied to detect electricity theft: A comparative study on unbalanced data handling techniques. *International Journal of Electrical Power & Energy Systems*, 131, 107085.
- [18] Lin, G., Feng, H., Feng, X., Wen, H., Li, Y., Hong, S., & Ni, Z. (2021). Electricity theft detection in power consumption data based on adaptive tuning recurrent neural network. *Frontiers in Energy Research*, 9, 773805.
- [19] Tehrani, S. O., Moghaddam, M. H. Y., & Asadi, M. (2020, September). Decision tree based electricity theft detection in smart grid. In *2020 4th International conference on smart city, internet of things and applications (SCIOT)* (pp. 46-51). IEEE.
- [20] Li, S., Han, Y., Yao, X., Yingchen, S., Wang, J., & Zhao, Q. (2019). Electricity theft detection in power grids with deep learning and random forests. *Journal of Electrical and Computer Engineering*, 2019, 1-12.
- [21] Toma, R. N., Hasan, M. N., Nahid, A. A., & Li, B. (2019, May). Electricity theft detection to reduce non-technical loss using support vector machine in smart grid. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)* (pp. 1-6). IEEE.
- [22] Aslam, Z., Ahmed, F., Almogren, A., Shafiq, M., Zuair, M., & Javaid, N. (2020). An attention guided semi-supervised learning mechanism to detect electricity frauds in the distribution systems. *IEEE Access*, 8, 221767-221782.
- [23] Lu, X., Zhou, Y., Wang, Z., Yi, Y., Feng, L., & Wang, F. (2019). Knowledge embedded semi-supervised deep learning for detecting non-technical losses in the smart grid. *Energies*, 12(18), 3452.
- [24] Boucetta, C., Flauzac, O., Nassour, A. N. M., & Nolot, F. (2020, June). Multi-level Hierarchical Clustering Algorithm For Energy-theft Detection in Smart Grid Networks. In *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)* (pp. 1-6). IEEE.
- [25] Lin, Q., Li, M., Feng, S., Yang, J., Surn, X., Li, J., ... & Xie, X. (2022, September). Identification of electricity theft based on the k-means clustering method. In *2022 IEEE 9th International Conference on Power Electronics Systems and Applications (PESA)* (pp. 1-6). IEEE.
- [26] Zheng, Z., Yang, Y., Niu, X., Dai, H. N., & Zhou, Y. (2017). Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. *IEEE Trans. on Industrial Informatics*, 14(4), 1606-1615.
- [27] State Grid Cooperation of China. (2023). <http://www.sgcc.com.cn/> (Erişim tarihi: 01.04.2023).
- [28] henryRDlab. (2018). <https://github.com/henryRDlab/ElectricityTheftDetection> (Erişim tarihi: 01.04.2023).
- [29] Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data*, 8(1), 1-37.
- [30] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
- [31] Aydin, Z. E., & Ozturk, Z. K. (2021). Performance analysis of XGBoost classifier with missing data. *Manchester Journal of Artificial Intelligence and Applied Sciences (MJAIAS)*, 2(02), 2021.
- [32] Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3), 3446-3453.
- [33] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189-1232.
- [34] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30.
- [35] Chen, T., & Guestrin, C. (2016, August). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).

- [36] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31.