

Makine Öğrenmesi Algoritmaları ile Elektrik Dağıtım Şebekeleri Arıza Tahmini

Ali GEYİKOĞLU^{1*}, Mete YAĞANOĞLU²

Öz

Elektrik dağıtım şebekelerinde arıza; kaliteli ve sürekli enerji akışını engelleyici faktörler olarak tanımlanmaktadır. Arızanın meydana gelmesi sonrasında Elektrik Dağıtım Şirketleri, bakım-onarım ve yatırım çalışmaları ile düzeltici faaliyetler gerçekleştirmektedir. Meydana gelen arızalar ve sonrası düzeltici faaliyetler ile teknik kalite parametreleri sistemlerce oluşturulmaktadır. Ancak ortaya çıkan teknik veriler, herhangi bir tahminleme altyapısında kullanılmamakta, düzeltici faaliyetler genel olarak yorum ve taleplere istinaden gerçekleştirilmektedir. Bu çalışmada, sezgisel yaklaşımların önüne geçmek amacıyla, elektrik dağıtım şirketi operatörlerinin saha faaliyetleri sonrası sistemler tarafından örneklenerek kayıt altına alınan Aras EDAŞ'a ait Kesinti Süreleri ve Sıklığı verileri ile ilgili dönemlere ait Aras EDAŞ işletme sorumluluk sahasındaki 7 ile esas meteorolojik veriler kullanılmıştır. Veri seti içerisinde yer alan öznitelikler ve sınıflar üzerinde veri ön işleme, öznitelik seçimi, öznitelik çıkarımı gerçekleştirilmiştir. Regresyon işlemleri ile tahminleme gerçekleştirilecek hale gelen veri setleri %80'i eğitim ve %20'si test verisi olacak şekilde; Hafif Gradyan Artırma Makinesi (LGBM), Aşırı Gradyan Artırma (XGB), Destek Vektör, Rastgele Orman, Kategorik Artırma, k-En Yakın Komşu, Karar Ağacı, Lineer olmak üzere 8 farklı regresyon modeline tabi tutulmuştur. Veri seti üzerinde yer alan iki farklı bağımlı değişkene ait çok sınıflı değerler ayrı ayrı sınıf modeline dahil edilmiş olup toplamda 8 farklı model için 16 adet regresyon çalışması gerçekleştirilmiştir. En iyi model yapısına ulaşabilmek amacıyla hiperparametre optimizasyonu uygulanmıştır. Birincil çok sınıflı regresyon tahmini için en iyi model doğruluğu LGBM Regressor ile %93,305 olarak elde edilirken, ikincil çok sınıflı tahmin için en iyi model doğruluğu XGB Regressor ile %95,812 olarak elde edilmiştir.

Anahtar Kelimeler: Elektrik dağıtım şebeke arızası, Makine öğrenmesi, Regresyon, Tahmin.

Electricity Distribution Networks Fault Prediction with Machine Learning Algorithms

Abstract

Malfunction in electrical distribution networks; They are defined as factors that prevent quality and continuous energy flow. After the failure occurs, Electricity Distribution Companies carry out corrective actions through maintenance-repair and investment works. Failures that occur and subsequent corrective actions and technical quality parameters are created by the systems. However, the resulting technical data is not used in any forecasting infrastructure, and corrective actions are generally carried out based on comments and requests. In this study, in order to avoid intuitive approaches, the Interruption Duration and Frequency data of Aras EDAŞ, which were sampled and recorded by the systems after the field activities of the electricity distribution company operators, and the main meteorological data of 7 provinces in the Aras EDAŞ operational responsibility area for the relevant periods were used. Data preprocessing, feature selection, and feature extraction were carried out on the attributes and classes in the data set. The data sets that will be used for estimation with regression operations were subjected to 8 different regression models, including Light Gradient Boosting Machine (LGBM), Extreme Gradient Boosting (XGB), Support Vector, Random Forest, Categorical Boosting, k-Nearest Neighbor, Decision Tree, and Linear, with 80% of the data being training and 20% being test data. Multi-class values of two different dependent variables on the data set were included separately in the class model, and a total of 16 regression studies were carried out for 8 different models. Hyperparameter optimization was applied to achieve the best model structure. While the best model accuracy for primary multi-class regression prediction was obtained as 93.305% with the LGBM Regressor, the best model accuracy for secondary multi-class prediction was obtained as 95.812% with the XGB Regressor.

Keywords: Electricity distribution network failure, Machine learning, Regression, Prediction.

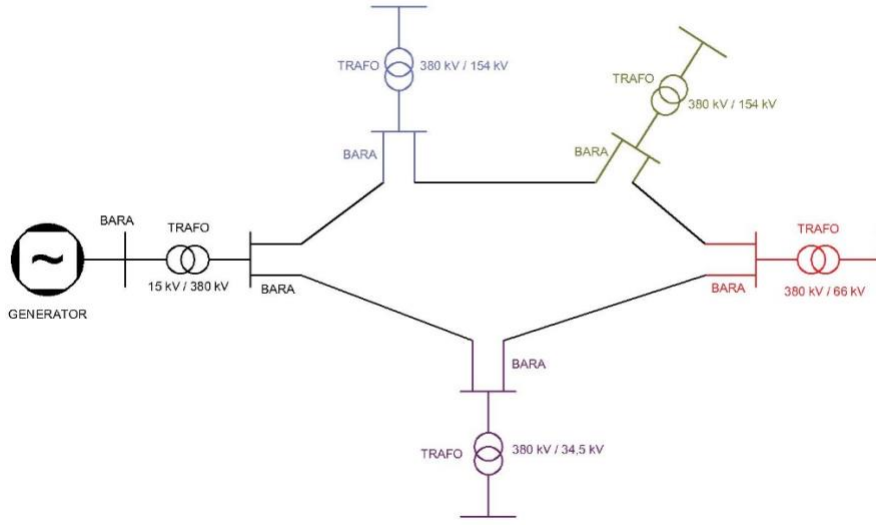
¹Aras Elektrik Dağıtım A.Ş., Akıllı Şebekeler ve AR-GE Müdürlüğü, Erzurum, Türkiye, ali.geyikoglu@arasedas.com

²Atatürk Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Erzurum, Türkiye, yaganoglu@atauni.edu.tr

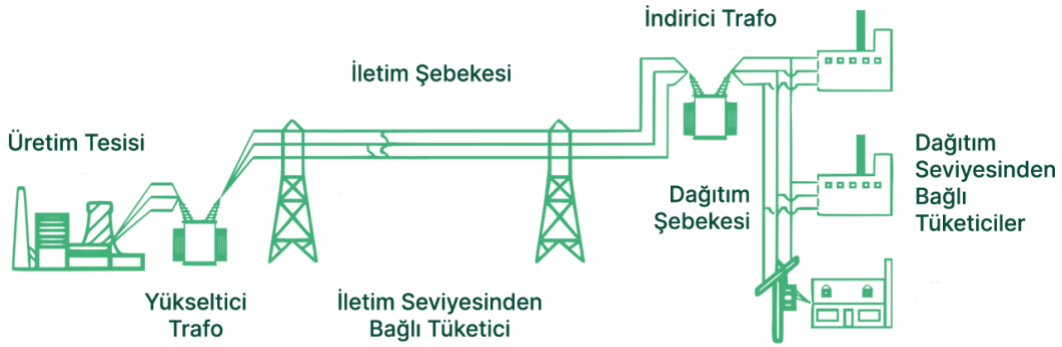
*Sorumlu Yazar/Corresponding Author

1. Giriş

İnsan yaşamının bir parçası olan ve teknolojileri ayakta tutan elektrik enerjisi, hiç kuşkusuz 21.yüzyılın en önemli ve vazgeçilmez enerji kaynaklarından biri olmuştur. Modern teknolojilerin hızlı gelişimi ile birlikte daha fazla ihtiyaç duyulan ve zorunluluk haline gelen elektrik enerjisi için en önemli unsurlar kalite ve sürekliliktir. Temel olarak; elektrik enerjisinin üretim noktasındaki (hidroelektrik, jeotermal, nükleer, rüzgâr, güneş vb.) yükseltici transformatörlerden geçirildikten sonra dağıtım öncesi ilk noktaya (indirici merkez-trafo merkezi) ulaştırılmasına Elektrik Enerjisinin İletimi denilmektedir. Bazı durumlarda elektrik enerjisi üretim tesisleri doğrudan elektrik dağıtım tesisine de bağlı olabilir (Bağlantı Anlaşmaları ve Tesis Sözleşmeleri Doğrultusunda). Elektrik Enerjisinin Yüksek ya da Orta Gerilim seviyesinde iletilmesi ve dağıtılması, temel elektriksel parametrelere esas olarak gerçekleştirilir ve bu şekilde tasarım edilmiş bir sistem ile enerji kayıpları azalır, iletken kesitleri düşürülür ve buna bağlı olarak taşıyıcı-durdurucu-nihayet-zaviye görevli direkler çok daha ekonomik bir hale getirilir. Elektrik Enerjisi indirici güç transformatör merkezlerine ulaştıktan sonra Elektrik Dağıtımına uygun bir gerilim seviyesine dönüştürülür (Ülkemizde T.C. TEİAŞ ve T.C. TEDAŞ mevzuatları doğrultusunda genel olarak 28,5kV-36kV). Elektrik enerjisinin Üretimi, İletimi ve Dağıtımına dair tek hat şeması örnekleme Şekil 1 ve Şekil 2’de sunulmuştur. Bu noktadan itibaren; Ülkemizde Elektrik Dağıtımına esas temel işletme faaliyetleri (bakım-onarım, yatırım) İşletme Hakkı Devir Sözleşmeleri doğrultusunda 21 Elektrik Dağıtım Şirketi tarafından gerçekleştirilmektedir. Elektrik Dağıtımına esas taşınmaz saha varlıkları ise T.C. TEDAŞ (Türkiye Elektrik Dağıtım Anonim Şirketi) mülkiyetindedir. Elektrik Dağıtım Şirketlerinin, Elektrik Piyasası ve Enerji Dönüşümü başlıklı temel faaliyetleri ise T.C. EPDK (Enerji Piyasası Düzenleme Kurumu) tarafından düzenlenmekte ve denetlenmektedir. Bu kapsamda Elektrik Dağıtım Şebekesi üzerinde oluşan arızalar, SCADA (Supervisory Control and Data Acquisition-Merkezi Denetim ve Veri Toplama), OSOS (Otomatik Sayaç Okuma Sistemi) vb. yapılar ile uzaktan izleme-kontrol-kayıt işlemlerine tabi tutulmaktadır. Ayrıca son kullanıcı noktasında oluşan bildirimler ile CBS (Coğrafi Bilgi Sistemi) tabanlı OMS (Outage Management System-Kesinti Yönetim Sistemi) sistemleri doğrultusunda, arıza yeri şebeke üzerinde görselleştirilerek MWM (Mobile Workforce Management System) sistemleri sayesinde ekip atamaları ve arıza giderme faaliyetleri sürdürülmektedir. Böylece grid yapıdaki Elektrik Dağıtım Şebekesi arızaları uzaktan izleme-kontrol-kayıt ve müdahale işlemlerine tabi tutulmaktadır. İlgili kayıtlar T.C. EPDK tarafından yürürlüğe alınan Elektrik Piyasasında Dağıtım ve Perakende Satışına İlişkin Kalite Yönetmeliği ve ilgili Elektrik Piyasası Mevzuatları ile düzenlenmekte ve denetlenmektedir (T.C. EPDK, 2023).



Şekil 1. Elektrik enerjisi iletim ve dağıtım şebekesi tek hat örnekleme şeması.



Şekil 2. Elektrik enerjisi iletim ve dağıtım şebekesi tek hat örnekleme şeması (Erdem ve Karamustafaoğlu, 2021).

İlgi yönetmelik içerisinde yer alan nicel(kantitatif) formüller ile Tablo-1 Kesintiler ve Tablo-5 Ortalama Sistem Kesinti Frekansı/Sıklığı İndisi (SAIFI System Average Interruption Frequency Index) - Ortalama Sistem Kesinti Süresi İndisi (SAIDI System Average Interruption Duration Index) değerleri yukarıda bahsedilen uzaktan izleme ve kayıt sistemleri doğrultusunda ilgili Elektrik Dağıtım Şirketinin işletme sorumluluk sahasına ait arıza kriterlerini yansıtmaktadır (T.C. EPDK, 2023). Bu veriler ve nitel (kalitatif) ölçütler ile bakım-onarım, yatırım faaliyetleri gerçekleştirilmektedir. Nitel ölçütler son kullanıcı başvurularını, şebeke işletmecisi öngörülerini ve saha tespitlerini kapsamaktadır.

Bu bağlamda; nitel ölçütlerin ikinci planda bırakılarak, tamamen nicel matematiksel ölçütlere esas bakım-onarım ve yatırım faaliyetlerinin planlanması ihtiyacı söz konusu edilmektedir. İlgili mevzuat içerisinde yer alan Tablo 1 Kesintiler görseli Şekil 3'te, bu çalışmaya esas alınan Tablo 5

Ortalama Sistem Kesinti Frekansı/Sıklığı İndisi - Ortalama Sistem Kesinti Süresi İndisi görseli ise Şekil 4'te sunulmuştur.

Tablo 5 verileri ham haliyle, orta gerilim ya da alçak gerilim bazında kent, kent altı ve kırsal elektrik abonelerinin/son kullanıcıların hangi sebepler doğrultusunda enerjilerinin gittiğini ortaya koyan Tablo 1 verileri doğrultusunda, T.C. EPDK mevzuat esaslarına dayanan matematiksel bir yaklaşım ile OKSIK öznitelikleri enerji kesintisinden etkilenen bölgedeki abone sayısının o il/ilçe/mahalle kapsamındaki tüm abone sayısına oranını, OKSURE_DK ise enerji kesintisinden etkilenen bölgedeki abone sayısının etkilenen dakika cinsinden süre ile çarpımı sonrası o il/ilçe/mahalle kapsamındaki tüm abone sayısına oranını sayısal veri cinsinden ihtiva etmektedir. Muhtelif izin süreçleri ile temin edilen veri setine esas, mevzuatlarda da yer aldığı şekliyle Tablo 1 üzerinden Tablo 5 verilerinin elde edildiği basitleştirilmiş formül yapısı Şekil 5'te sunulmaktadır.

TABLO-1 KESİNTİLER

KESİNTİ KODU (1)	KADEMERİZ	KESİNTİNİN YERİ (3)				KESİNTİNİN SINIFI (5)				KESİNTİDEN ETKİLENEN KULLANICI SAYISI (9)			TOPLAM ETKİLENME SÜRESİ (10)								
		İL (3A)	İLÇE (3B)	SEBEKE ÜNSÜRÜ TİPİ(3C)	SEBEKE ÜNSÜRÜ KODU (3D)	KESİNTİ NEDENİNE İLİŞKİN AÇIKLAMA (4)	KAYNAĞA GÖRE (5A)	SÜREYE GÖRE (5B)	SEBEBE GÖRE (5C)	BİLDİRİME GÖRE (5D)	BİTLİRME TARİHİ VE ZAMANI (6)	SONA ERME TARİHİ VE ZAMANI (6)	KESİNTİ SÜRESİ (SAAT) (8)-(7)-(6)	KENTSE	KENTAL	KIRSAL	KENTSE	KENTAL	KIRSAL		
														L	TI	L	L	TI	L		
OG (9A)																					
AG (9B)																					
AG (9C)																					
AG (9D)																					
OG (9E)																					
AG (9F)																					
OG (10A)-(9A)(8)																					
AG (10B)-(9B)(8)																					
OG (10C)-(9C)(8)																					
AG (10D)-(9D)(8)																					
OG (10E)-(9E)(8)																					
AG (10F)-(9F)(8)																					

Şekil 3. Elektrik Piyasasında Dağıtım ve Perakende Satış Faaliyetlerine İlişkin Kalite Yönetmeliği Tablo-1 Kesintiler (T.C. EPDK, 2023).

TABLO-5 KESİNTİ SÜRELERİ VE SIKLIĞI

A) OKSÜRE (Bildirimsiz)		KENTSEL KULLANICILAR			KENTALTİ KULLANICILAR			KIRSAL KULLANICILAR			GENEL TOPLAM
KAYNAK	SEBEP	OG	AG	TOPLAM	OG	AG	TOPLAM	OG	AG	TOPLAM	
İLETİM	Şebeke İşletmecisi										
İLETİM	Müşeri Sebep										
DAĞITIM-OG	Şebeke İşletmecisi										
DAĞITIM-OG	Düşük										
DAĞITIM-OG	Müşeri Sebep										
DAĞITIM-OG	Güvenlik										
DAĞITIM-AG	Şebeke İşletmecisi										
DAĞITIM-AG	Düşük										
DAĞITIM-AG	Müşeri Sebep										
DAĞITIM-AG	Güvenlik										
GENEL TOPLAM											
B) OKSÜRE (Bildirimsiz)		KENTSEL KULLANICILAR			KENTALTİ KULLANICILAR			KIRSAL KULLANICILAR			GENEL TOPLAM
KAYNAK	SEBEP	OG	AG	TOPLAM	OG	AG	TOPLAM	OG	AG	TOPLAM	
İLETİM	Şebeke İşletmecisi										
DAĞITIM-OG	Şebeke İşletmecisi										
DAĞITIM-OG	Güvenlik										
DAĞITIM-AG	Şebeke İşletmecisi										
DAĞITIM-AG	Güvenlik										
GENEL TOPLAM											

Şekil 4. Elektrik Piyasasında Dağıtım ve Perakende Satış Faaliyetlerine İlişkin Kalite Yönetmeliği Tablo-5 Kesinti Süreleri ve Sıklığı (T.C. EPDK, 2023).

OKSURE : Ortalama Kesinti Süresi

$$OKSURE = \frac{(Kesintiden Etkilenen Abone Sayısı) \times (Toplam Kesinti Süresi)}{Toplam Abone Sayısı}$$

OKSIK : Ortalama Kesinti Sıklığı-Frekansı

$$OKSIK = \frac{Kesintiden Etkilenen Abone Sayısı}{Toplam Abone Sayısı}$$

Şekil 5. Tablo 1 Kesintiler verilerinin Tablo 5 verileri ile ilişkisi.

Bu doğrultuda, sezgisel yaklaşımların yerine nicel yaklaşım metotları ile gerçekleştirilecek fiziki saha çalışmaları sonucu arıza lokasyonunun tespiti ve arızaya müdahale faaliyetleri, elektrik dağıtım süreçlerini başta T.C. EPDK ve ilgili mevzuat standartlarının çok üzerine çıkarmayı bir hedeften çok uygulanması gereken zorunlu bir amaç haline getirmektedir. Sektörün mevcut uygulamaları incelendiğinde ise özellikle ilgili veri setlerini esas alan, kesin çözüm ve sonuçlarını barındıran herhangi bir yapay zekâ çalışması tespit edilememektedir. Bu çalışmanın amacı; sektörün arıza kaynaklı Tablo-1 veri seti menşeli Tablo-5 veri seti nesnelere yapay zekâ uygulamaları ile işleyerek, güncel teknoloji kullanımının eksik kaldığı metodolojiyi ortaya çıkararak kaliteli, kesintisiz elektrik enerjisini son kullanıcı bazındaki elektrik abonelerine aktarmayı hedeflemektedir. Çalışma kapsamında incelemeye alınan literatür içerikleri ve uygulamaya alınan metotlar ilgili bölümlerde ayrıntılı olarak beyan edilmektedir.

2. Literatür Taraması

Mori ve Yokoyama ilgili çalışmalarında; yenilenebilir enerji kaynaklarının değişken hava koşullarında elektrik dağıtım şebekesinde meydana getirdiği gerilim dalgalanmalarını, arızaları ve enerji kayıplarını ortadan kaldırabilmek amacıyla dağıtım ağı/şebekesi modellemelerinin tahmini için derin öğrenme yöntemlerinden yapay sinir ağları Radyal Temelli Fonksiyon Ağları (RBFN's) kullanılmıştır. Başarılı sonuçların elde edildiği bu çalışmada RBFN's ve Veri Madenciliği Regresyon Ağacı'nın hibrit bir metodu önerilmiştir. Ön filtreleme tekniği olarak Regresyon Ağacı modelleri kullanılmıştır (Mori ve Yokoyama, 2016).

Liu ve diğerleri ilgili çalışmalarında; elektrik dağıtım şebekelerinde meydana gelen buz yükü ve sonucu arızaların afet geçmişlerine bakarak gelecek tahminlemesini, yüksek risk taşıyan hat segmentlerinin tespitini gerçekleştirmişlerdir. Çin'in Shen Yang Şehrine ait elektrik dağıtım ağı üzerinden örneklemeler yapılarak hem geçmiş meteorolojik veriler hem de afet durumları analiz edilmiştir (Xinrui ve ark., 2017).

Yellagoud ve Talluri ilgili çalışmalarında; İleri Beslemeli Sinir Ağları (FFNN's), Radyal Temelli Fonksiyon Ağları (RBFN's) ve Uyarlanabilir Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS's) metotlarının orta gerilim dağıtım şebekelerinde otomatik arıza bulma işlemi ele almışlardır. Muhtelif veriler ışığında dağıtım varlıkları arasında yer alan transformatör bölgelerinde meydana gelebilecek arızaların önceden tahmini, ANFIS's metodu ile daha yakın arıza konumu oluşturmuş, RBFN's araçları ise daha hızlı bir öğrenim ortaya çıkarmıştır. Böylece daha yüksek doğruluk oranı ile otomatik arıza tespiti sağlanmış, etkilenen abone sayısı ve arızalı bölgedeki kayıp enerjiler azaltılmıştır. Güç sisteminin güvenilirliğine ve ekonomisine katkıda bulunulmuştur (Yellagoud ve Talluri, 2019).

Perles ve diğerleri ilgili çalışmalarında; Uyarlanabilir Çoklu Etmen Sistemi Teorisi ile ele alınan elektrik dağıtım şebekesinin, akıllı şebeke formuna geçişindeki makul durum süresini ve hesaplama karmaşıklıklarını tahmin etmeyi ele almıştır. Böylece her yönüyle kendi kendini kontrol edebilen otonom ve akıllı bir güç sistemi yapısının tahminlemesi dile getirilmektedir (Perles ve ark., 2017).

Abdel-Nasser ve diğerleri ilgili çalışmalarında; akıllı şebekelerin akım, gerilim, aktif ve reaktif güç kaybı gibi parametrelerinin yapay sinir ağları kullanılarak durum tahminini ortaya koymaktadır. Bu bağlamda, SE-NN Sinir Ağı Kullanılarak Durum Tahmini olarak adlandırılan yöntem esas alınmıştır. İlgili yöntem yüksek doğrulukla tahmin kapasitesine sahip çok hızlı bir araçtır (Abdel-Nasser ve ark., 2018).

Mai ve diğerleri ilgili çalışmalarında; elektrik alçak gerilim dağıtım şebekelerinde dağıtım transformatörlerindeki doluluğun kullanıcı grupları üzerinden elde edilen teknik veriler ışığında analiz ve tahmin edilmesini, bu tahminler ışığında alçak gerilim şebekesinin kendini yönetebilir bir yapıya kavuşmasını gerçekleştirmişlerdir. Teknik veriler kullanıcıların akıllı sayaçları üzerinden alınmış, dört makine öğrenmesi algoritması olan sırt regresyonu, destek vektörü regresyonu, rastgele orman regresyonu, extrem gradyan artırma regresyonuna (XGBR's) tabi tutulmuştur. Amaçlanan çalışma için en hızlı ve doğru sonuç veren yöntem XGBR's olarak belirlenmiştir (Mai ve ark., 2022).

Kankanala ve diğerleri ilgili çalışmalarında; meteorolojik etkilerin ve çevresel faktörlerin elektrik dağıtım şebekesi enerji akışında kesintilere/arızalara sebep olması nedeniyle ilgili kesintilerin/arızaların önceden tahminini esas almışlardır. ADABOOST(+) hızlandırma algoritması kullanılarak toplu öğrenme yaklaşımı önerilmektedir. Dört ayrı veri setinin kullanıldığı bu yapıda diğer modellerden daha yüksek doğrulukta tahminler elde edilmiştir (Kankanala ve ark., 2013).

Hassani ve diğerleri ilgili çalışmalarında; akıllı şebekelerdeki iletim hatlarının arızalardan korunmasını artırmak amacıyla hesaplamalı bir zekâ yöntemi önermektedir. Arıza oluşumu sırasında meydana gelen arıza sinyalleri sinyal işleme ile nöro-bulanık çıkarım sistemlerine (ANFIS's), destek vektör regresyonuna (SVR's) ve yapay sinir ağlarına (ANN's) aktarılmıştır. Böylece arızanın yeri ve

teknik özellikleri tahmin edilmeye çalışılmıştır. Yöntemler arasında en başarılı sonucu veren yapı ANFIS's olmuştur (Hassani ve ark., 2019).

De Santis ve diğerleri ilgili çalışmalarında; akıllı şebeke yapılarında elektrik operatörlerinden, tüketicilerden ve akıllı sensörlerden alınan bilgilerin örüntü tanıma, analiz işlemlerini esas almaktadır. Arızaların tanınması, zaman serileri, sayısal veriler ve kategorik veriler gibi karmaşık veri türleri ağırlık analizlerine tabi tutulmaktadır. Arıza modeli k küme sayısını tahmin edebilen bir başlatma kuralına sahip kümeleme (k-means) algoritması ile elde edilmektedir. Mevcut çalışma, İtalya'nın Roma Kentinin tamamını besleyen akıllı şebeke parametreleri ile operatör-müşteri geri bildirimlerini veri kümesi olarak barındırmaktadır (De Santis ve ark., 2016).

Majidi ve diğerleri ilgili çalışmalarında; seyrek temsil vektörü, bulanık kümeleme ve makine öğrenmesi aracılığıyla elektrik dağıtım hatlarında arıza yeri tespiti için bir yöntem sunmaktadır. Yöntem arıza öncesi ve anında şebekeden alınacak elektriksel parametrelerin Fuzzy-c ortalama ile analiz edildikten sonra k-en yakın komşu algoritmalarına tabi tutulması ile sonuçlar oluşturmaktadır (Majidi ve ark., 2014).

Kurup ve diğerleri ilgili çalışmalarında; elektrik dağıtım şebekelerindeki kapalı devre topolojisini tahminleme ve farklı konumlardaki arıza oluşumunu analiz etme işlemlerini incelemişlerdir. Evrişimli sinir ağı modelleri, destek vektör makinaları teknikleri ile hatalı ve normal veriler doğrultusunda tahminleme işlemleri esas alınmaktadır (Kurup ve ark., 2021).

Mestav ve Tong ilgili çalışmalarında; elektrik dağıtım şebekelerinde yer alan akıllı sayaçlar ve SCADA ölçüm sonuçları ile Bayes durum tahmincisi, derin üretken düşman ağı ve minimum ortalama kare hatası süreçleri yürütülmektedir. Mevcut çalışmanın geleneksel yöntemlerden daha az maliyetli olduğu tespit edilmektedir (Mestav ve Tong, 2019).

Jamali ve diğerleri ilgili çalışmalarında; dağıtık üretim ve mikro şebeke barındıran elektrik dağıtım şebekelerinde hızlı bir hibrit arıza yeri belirleme faaliyetlerini yürütmüşlerdir. Yöntem, şebekenin ilgili noktalarındaki sensör çıkışlarının ölçüm verilerini kullanmaktadır. Ölçüm noktasından arıza noktasına olan mesafe analizi için tek bir ileri beslemeli sinir ağı kullanılmıştır. KNN ise yapay sinir ağı çıkışındaki verileri yorumlayarak arızaya olan uzaklığı tahminler (Jamali ve ark., 2020).

Dashtdar ve diğerleri ilgili çalışmalarında; elektrik dağıtım şebekelerinde yer alan sekonder aşırı akım koruma röleleri (fider koruma), sekonder aşırı akım-gerilim koruma röleleri (fider yönetim) üzerinden alınan arıza sinyallerinden karakteristik çıkararak yapay sinir ağı algoritmaları ile arıza tespitini esas almaktadır. Girdi verileri birçok elektriksel parametreyi esas almaktadır (Dashtdar ve ark., 2018).

Mori ve diğerleri ilgili çalışmalarında; elektrik iletim sistemlerindeki arıza yeri ve arıza tipinin tahminlenmesi için yapay sinir ağı yöntemini esas alan hibrit bir yapıyı esas almaktadır. FFT girdi

verilerinin özniteliklerini çıkarmak için kullanılır. Girdi ve çıktı değişkenleri arasındaki doğrusal olmayan ilişki radyal tabanlı fonksiyon ağı (RBFN's) ile kontrol edilmektedir. Model doğruluğu artırılmak için D-RBFN adlı yeni bir RBFN yöntemi önerilmiştir. Öneri yöntemi bir sisteme başarı ile uygulanmıştır (Mori ve ark., 2002).

Ferreira ve diğerleri ilgili çalışmalarında; en önemli elektrik dağıtım altyapısı elemanlarından olan dağıtım güç transformatörleri ve bunlara ait OG-AG dağıtım şebekesi teçhizatlarının, talep tahmini yöntemleri ile korunmasını ve en ideal kullanım modellerini ortaya koymayı hedeflemişlerdir. Talep tahmini yönetimi temelde, enerji taleplerinin net bir şekilde modellenmesi ile en ideal enerji arzını ortaya çıkarmak olarak tanımlanabilmektedir. Bu noktadaki en zorlu husus, geleceğe dair enerji talebinin hangi seviyede ve hangi zaman diliminde nasıl gerçekleşeceği. Talep tahmin uygulamalarına dair bu anomali yanlış sonuçlar ile teçhizat kayıplarını (arızalar) ve son kullanıcı noktasındaki müşteri memnuniyetini aksi yönde etkileyecektir. İlgili çalışmada, bu kapsamda ortaya çıkarılan birçok tekniğin geçmiş veriye dayalı olması sebebiyle, karmaşık enerji zaman serilerini ortaya çıkaracağı savunulmaktadır. Yenilikçi bir yöntem olarak kabul gören zamansal füzyon transformatörü ile geçmişe dair anlamlı veriler ortaya çıkarılarak gelecek enerji talep tahminleri yapılmaktadır (Ferreira ve ark., 2025). Ancak ilgili çalışmada da görüleceği üzere, enerji altyapısına dair geçmiş veri için standardizasyon olamayacağı savunulsa da geçmiş verinin kullanıldığı ve ne kadar değerli bir kaynak olduğu anlaşılmaktadır. Enerji altyapılarına dair veri kalitesi, enerjinin üretiminden iletimine ve iletiminden dağıtımına kadar oluşan süreçleri iyi tecrübe etmek ve yorumlamak ile ilgilidir. Böylece veri seti niteliği modellemeye başlamadan önce optimize hale getirilebilir.

Zhou ve diğerleri ilgili çalışmalarında; rüzgâr tribünlerinde meydana gelen geçmiş arıza verilerini kullanarak hafif birçok ölçekli ayrılabilir kalıntı ağı (LMSRN) modelini kullanmayı önermişlerdir. İlgili model mekânsal özellik türetme ve ayrılabilir özellik çıkarma sistematiğine sahiptir. Veri çeşitliliği için kanonik korelasyon analizi ile kanonik kümeleme altyapıları kullanılmaktadır. Küme sayısının optimum hale getirilmesi ve diğer metotlara göre yüksek başarı kriterleri ortaya çıkarılmıştır (Zhou ve ark., 2025). İlgili çalışmanın elektrik dağıtım şebekesinden çok rüzgâr üretim santralleri üzerine etkisi söz konusudur. Mevcut çalışmamız ile ortak alanı, geçmiş veriden analiz ve veri ön işleme adımlarının seçici nitelikte olmasıdır.

Arıza ve arızanın oluşacağı döneme ait tahminler, elektrik enerjisinin sürekliliği açısından önemli bir rol oynamaktadır. Tablo-1 Kesinti verileri ile bölgesel parametreler sonucu ortaya çıkan ve kalite faktörlerini birebir etkileyen Tablo-5 Kesinti Süreleri ve Sıklığı verilerinin tahmini enerji sürekliliğini sağlamanın yanında, kestirimci yaklaşımlar ile ilgili noktaya zamanında ve etkin bir müdahale imkânı sunacağından zamanın, malzemenin ve personelin etkin kullanımını sağlayacaktır.

Böylece sezgisel yaklaşımlar, kişisel öngörü ve görüşler ile söz konusu edilen bakım ve yatırım faaliyetleri de etkin bir şekilde, gerekli olan bölgeye gerekli olduğu şekliyle gerçekleştirilebilir.

Bu çalışmada ortaya konulan yöntem, diğer yöntemlerden farklı olarak yapay zekâ biliminin bir alt dalı olan makine öğrenmesi yöntemleriyle, sınıflandırma yapısına sahip olmayan ham veri seti üzerinde çoklu sınıf modellerinin belirlenmesini, veri setine geçmişe dair reel hava durumu parametrelerinin eklenmesini, aynı veri seti üzerinde iki ana başlığa esas çoklu sınıfın regresyon işlemlerine tabi tutulmasını ihtiva etmektedir. Literatürdeki diğer çalışmalar genel olarak, elektrik dağıtım şebekesi üzerinden muhtelif ölçüm cihazları ile alınan, canlı temel elektriksel parametreler ile canlı hava durumu verilerini sınıflandırma veya regresyon modellerine tabi tutarak tahminleme işlemleri yürütmüştür. Bu çalışma ile geçmişte meydana gelmiş reel veriler ve sonuçlarına dair veri setlerini kullanarak, gelecek öngörüsünü daha gerçekçi bir şekilde ortaya çıkarma amacı söz konusu edilmiştir. Ayrıca, önerilen yöntemin diğer çalışmalardan en büyük farkı ise katar (string) veriler için etiketleme amacıyla kullanılan LabelEncoder fonksiyonunun, ondalık sayısal (float) veriler üzerine uygulanması ile normal sayısal (integer) veriler elde edilerek, veri ön işleme adımlarına farklı bir yaklaşım kazandırılması olmuştur. Ayrıca aynı veri seti içerisindeki özniteliklerin iki farklı sınıf değişkeni olarak değerlendirilmesi ile arıza sayı ve süre tahmini elde edilmiştir. Bu çalışma içerisinde kullanılan veri setleri, gerekli üst yazışmalar ile alınan izinler doğrultusunda Aras Elektrik Dağıtım A.Ş.'ye ait veri ambarından temin edilen Kesinti Süreleri ve Sıklığı ile WeatherSpark resmî web sitesinden temin edilen pilot bölgeye ait geniş kapsamlı meteorolojik verileri ihtiva etmektedir. Veri setine ait öznitelikler ve çoklu sınıflar, ön işleme adımlarından geçirilerek gereksiz verilerin silinmesi, eksik verilerin tamamlanması, normalizasyon, One-hot encoding, verilerin ikili ilişkilerine göre sıralanması gibi çeşitli işlemlere tabi tutulmuştur. Veri ön işleme adımları için kullanılan, Scikit-learn kütüphanesinin preprocessing alt kütüphanesinde yer alan iki farklı encoder mevcuttur. Veri ön işleme adımlarında unutulmaması gereken en önemli aşama, ham veri seti içeriğinde yer alan ve nesne/öznitelik kapsamında kullanılması öngörülen katar(string) verilerin, makinenin/bilgisayarın anlayacağı sayısal dile çevrilmesidir. Çalışmamızda veri seti olarak esas aldığımız Tablo-5 yapısı ham haliyle "IL; DONEM" özniteliklerini barındırmaktadır. Pilot bölge seçimi ile tek alana indirgenen veri seti içeriği sadece Erzurum iline ait nitelikleri içerdiğinden, bu özneliğe ait tüm nesnelere ile birlikte düşürülmüştür. Veri setinin reel saha değişkenleri ile beslenmesi adına ilave edilen meteorolojik nesnelere ise "NEM_DURUMU" özneliğinde kategorik veriler içermektedir. Bu doğrultuda, 12 aylık dönem isimlerini (Ocak, Şubat...) içeren "DONEM" ve havanın nem kalitesini (rahat, kuru, ıslak...) içeren "NEM_DURUMU" öznitelikleri One-hot encoder uygulamasına tabi tutulmuştur. Label encoder her katar için ayrı ayrı sayısal etiket oluşturur. 2'li veya 3'lü değişkene kadar fayda gösteren Label encoder uygulamasının en büyük problemi, değişkenlere atanan değerlerin modeller tarafından tespit edilme zorluğu ile diğer öznitelik verileri ile birebir eşleşme

olması durumundaki aşırı öğrenme (overfitting) yaşanması olasılığıdır. Bu doğrultuda, kategorik verilerin sayısallaştırılması yanında, öznitelik matrisinin komple boyutlarınca (en x boy) binarizasyon işleminin de gerçekleştirilmesidir. Böylece doğal veri çoğaltma işleminin yanında, ilgili öznitelik için yapılan sayısallaştırma sadece o varlık söz konusu iken 1 değerini, olmadığı tüm durumlar için 0 değerini alacaktır. Özetle, “1” o kategorinin varlığını, “0” ise aynı kategorinin yokluğunu esas alacaktır. Uygulama ile oluşan vektör boyutu, kategorik bir sütunun aldığı benzersiz değerlerin sayısına eşittir. Çok değişkenli katar verilerin sayısallaştırılmasında daha nitelikli görümler oluşturan One-hot encoder uygulaması sonrası ortaya çıkan 0-1 değerlerinin genel adı kukla değişkenler (dummy variables) olarak adlandırılmaktadır. Ancak kodlanmış değişkenler arasında mükemmel bir çoklu doğrusal oluşumu (yüksek düzeyli korelasyon) var ise uygulama tarafından otomatik oluşturulan kukla değişkenler bir tuzak oluşturabilir. Bu durum literatürde kukla değişken tuzağı (the dummy variable trap) olarak adlandırılmaktadır. Bu tuzak durumu, bir değişkenin diğerleri üzerinden tahmin edilebileceğini belirtmektedir. Sonuç olarak sistem tahminleri çoklu bağlantılar sebebiyle mantıksal sonuçlar ortaya çıkaramayabilir. One-hot encoder alt kütüphanesi back-end özelinde kukla değişken tuzağını engelleyici ve ortadan kaldıracı otomatik yönetim sistemine sahiptir. Bunun yanında her türlü olasılığı göze alarak tuzak durumunun tam anlamıyla önüne geçilebilmesi adına oluşturulan kukla değişken öznitelik sütunlarında düşürme ve dönüşüm/dizi dilimleme işlemleri yürütülmelidir. Böylece hem model için ihtiyacın az olduğu bir ağırlık düşürülerek öğrenme süreci kolaylaştırmakta hem de öğrenme için asıl gerekli olan ağırlıklar söz konusu edilmektedir. Özellikle makine öğrenmesine ait regresyon modelleri için tahmin ve yorum problemlerine yol açan kukla değişken tuzağı, One-hot encoder uygulamasındaki en önemli problemlerden biri olmaktadır. Böylece eğitilen modele daha sade ve anlaşılır veriler sunulması işlem süresi ve işlem yükü asgari seviyelere indirgenmiş, daha etkin regresyon sonuçları elde edilmiştir. Veri setinde yer alan SAIFI-OKSIK ve SAIDI-OKSURE ana başlıkları, çok sınıflı yapıya sahip olduğundan, bu kapsama dair model seçim ve uygulamaları gerçekleştirilmiştir. Regresyon işlemleri ile tahminleme gerçekleştirilecek hale gelen veri setleri; LGBM Regressor, XGB Regressor, Support Vector Regressor, Random Forest Regressor, Cat Boost Regressor, k-Neighbors Regressor, Decision Tree Regressor, Linear Regressor olmak üzere 8 farklı regresyon modeline tabi tutulmuştur. Veri seti üzerinde yer alan SAIFI-OKSIK ve SAIDI-OKSURE çok sınıflı değerleri ayrı ayrı sınıf modeline dahil edilmiş olup toplamda 8 farklı model için 16 adet regresyon çalışması gerçekleştirilmiştir. En iyi model yapısının eldesi amacıyla hiperparametre optimizasyonu uygulanmış olup 8 adet performans değerlendirme parametresi ortaya konmuştur. Bu parametreler; doğruluk (accuracy), ortalama mutlak hata (mean absolute error), ortalama karesel hata (mean squared error), ortalama karesel hatanın karekökü (root mean square error), belirleme katsayısı (r2), ortalama mutlak yüzde hatası (mean absolute percentage error), maksimum hata (maximum error) ve açıklanabilen varyans

puanı (explained variance score) şeklindedir. Doğruluk (accuracy); modele esas doğru tahmin miktarının toplam tahmin miktarına oranı olarak tanımlansa da tek başına etkin bir başarı performansı değildir. Sınıflandırmadaki dengesizlik, yani yanlış tahmin edilen pozitif ve negatif örnekler kriter hesaplamasını yanlış sonuca götürecektir. Bu sebeple farklı başarı metrikleri kullanmak, doğruluk oranını destekleyici ve doğrulayıcı etkide olacaktır. Ortalama mutlak hata (mean absolute error); modelin gerçek değerler ile tahmin değerleri arasında fark olarak mutlak değerlerinin ortalamasını hesaplar. Böylece aykırı değerler söz konusu olan bir veri setinde, ortalama karesel hatanın karekökü metodundan daha nitelikli sonuçlar verecektir. Karesel hata ve karesel hatanın karekökü endeksli metotlar ise temelde hata oranını küçültme ve standart sapma payları üzerinden sonuçlar üretmektedir. Belirleme katsayısı (r^2); bağımsız değişken ('X') tarafından ortaya çıkarılan bağımlı değişken ('y') varyansının oranını belirler. Bu metrik, veri setinin seçilen modele göre uyumluluğunu ortaya çıkaran en önemli yapıdır. Çalışmada kullanılan 8 başarı metriği, herhangi bir açık uç bırakmamak adına hem birbirini destekleyici hem de anomali durumlarında ayırım sağlayan, makine öğrenmesi metotları için evrensel nitelikli metriklerdir. Bu çalışmada, SAIFI-OXSİK çok sınıflı tahmin için en iyi model doğruluğu LGBM Regressor ile %93,305 olarak elde edilirken, SAIDI-OXSURE çok sınıflı tahmin için en iyi model doğruluğu XGB Regressor ile %95,812 olarak elde edilmiştir. Böylece diğer çalışmalarda söz konusu edilen canlı veriden tahmin ve öneri sistemi oluşturulması yerine geçmişe dönük reel veriler üzerinden net bir şebeke karakteristiği meydana getirilerek saha gerçeğini yansıtacak şekilde tahmin ve öneri sistemlerinin oluşturulması, bu çalışmanın en büyük farkını ortaya koymaktadır. Veri seti esasına dayandırılan bu husus haricinde metot üzerindeki farklılıklar ise hem veri seti için birçok veri altyapısının birleştirilmesi hem de veri ön işleme adımlarında standart olarak uygulanmayan metotlara yer verilmesi ile ortaya konmuştur. Elektrik dağıtım şebekesi karakteristiğini net bir şekilde ortaya koyacak parametreler, geçmişe dönük reel veri setleri ile söz konusu olmaktadır. Arıza ve arızaya müdahale süreçleri de mevcut karakteristik yapıya göre gerçekleştirileceğinden, bu çalışma ile ortaya konulan yöntem ve sonuçları da sahada uygulanan gerçek metodolojiyi en iyi şekilde yansıtmaktadır.

3. Materyal ve Metot

Bu bölümde; regresyon işlemlerinde kullanılan Kesinti Süreleri ve Sıklığı ile meteorolojik parametrelerin tümleşik hale getirildiği veri seti, ilgili veri setine uygulanan veri ön işleme adımları ve regresyon algoritma modelleri hakkında bilgilere yer verilmiştir.

3.1. Veri Seti

Bu çalışma içerisinde kullanılan veri setleri, gerekli üst yazışmalar ile alınan izinler doğrultusunda Aras Elektrik Dağıtım A.Ş.'ye ait veri ambarından temin edilen Kesinti Süreleri ve Sıklığı ile WeatherSpark resmî web sitesinden temin edilen pilot bölgeye ait geniş kapsamlı meteorolojik verileri ihtiva etmektedir. Veri seti, 2019-2023 yılları arasında Aras Elektrik Dağıtım A.Ş. işletme sorumluluk bölgesinde yer alan Erzurum iline ait 267.571 adet örnek içeren Kesintiler veri setinin, T.C. EPDK tarafından uygulamaya alınan matematiksel yaklaşımı ile indirgenmiş Kesinti Süreleri ve Sıklığı veri setinde yer alan 5 yıllık 60 ana örneği esas almaktadır. SAIFI-OKSIK ve SAIDI-OKSURE çok sınıflı iki ana sınıf verisi sırasıyla öznitelikler içine dahil edilerek çapraz regresyon işlemlerine tabi tutulmuştur. Meteorolojik veriler için aylık bazda ortalama sıcaklık, nem durumu, en yüksek rüzgâr hızı (km/sa), en yüksek atmosfer basıncı (mbar) değerleri de veri setine dahil edilmiştir. Çalışmada kullanılan SAIFI-OKSIK ve SAIDI-OKSURE sınıf değişkenlerine, katar (string) veriler için etiketleme amacıyla kullanılan LabelEncoder fonksiyonunun, ondalık sayısal (float) veriler üzerine uygulanması ile normal sayısal (integer) veriler elde edilerek veri ön işleme adımlarına farklı bir yaklaşım kazandırılmıştır.

Aşağıda yer alan Şekil 6'da tümleşik geliştirme ortamına aktarılmış veri setine ait 15 satır, 10 öznitelik ve bunlara ait değerler örneklendirilmiştir. Ayrıca Çizelge 1 ile elektrik dağıtım şirketi veri ambarından temin edilen ham veri seti görseline de yer verilmiştir. İlgili şekil ve tabloda da görüleceği üzere veri kümesi hem sayısal hem de kategorik değerleri ihtiva etmektedir. Makine öğrenmesi algoritmaları, makine diline esas sayısal veriler ile işlem yapabildiğinden, kategorik verilerin sayısal verilere dönüştürülmesi zorunlu bir durumdur.

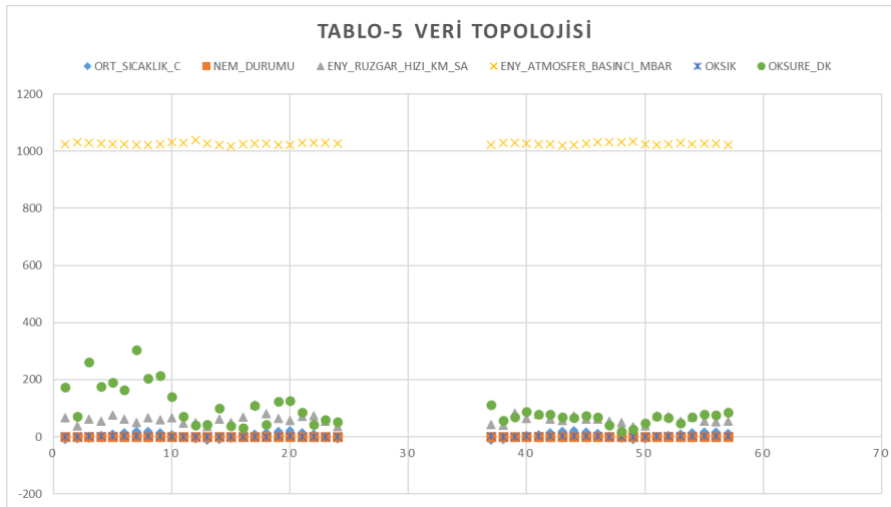
Index	İL	YIL	DONEM	ORT_SICAKLIK_C	NEM_DURUMU	ENY_RUZGAR_HIZI_KM_SA	ENY_ATMOSFER_BASINCI_MBAR	OKSIK	OKSURE_DK
0	ERZURUM	2019	OCAK	-9.4	kuru	67	1025	0.79	174.78
1	ERZURUM	2019	SUBAT	-7.5	kuru	38	1032	0.52	72.41
2	ERZURUM	2019	MART	-4.1	kuru	63	1028	1.96	260.64
3	ERZURUM	2019	NISAN	3.8	kuru	54	1026	1.44	176.96
4	ERZURUM	2019	MAYIS	10.7	kuru	77	1025	1.68	190.5
5	ERZURUM	2019	HAZIRAN	14.9	rahat	62	1024	1.56	163.3
6	ERZURUM	2019	TEMMUZ	19.2	rahat	51	1022	2.71	303.89
7	ERZURUM	2019	AGUSTOS	19.5	rahat	66	1023	1.79	203.9
8	ERZURUM	2019	EYLUL	14.8	kuru	59	1024	1.54	213.8
9	ERZURUM	2019	EKIM	8.1	kuru	67	1032	1.03	139.84
10	ERZURUM	2019	KASIM	1.2	kuru	48	1028	0.59	72.44
11	ERZURUM	2019	ARALIK	-6.2	kuru	50	1038	0.44	40.75
12	ERZURUM	2020	OCAK	-11.1	kuru	36	1027	0.52	43.99
13	ERZURUM	2020	SUBAT	-8.4	kuru	62	1021	1	101.16
14	ERZURUM	2020	MART	-5.2	kuru	51	1017	0.48	39.29
15	ERZURUM	2020	NISAN	4.3	kuru	70	1025	0.37	31.73

Şekil 6. Veri setine ait rasgele 15 örnek görseli (derlemeye alınmış hali).

Çizelge 1. Veri setine ait rastgele 15 örnek görseli (ham hali)

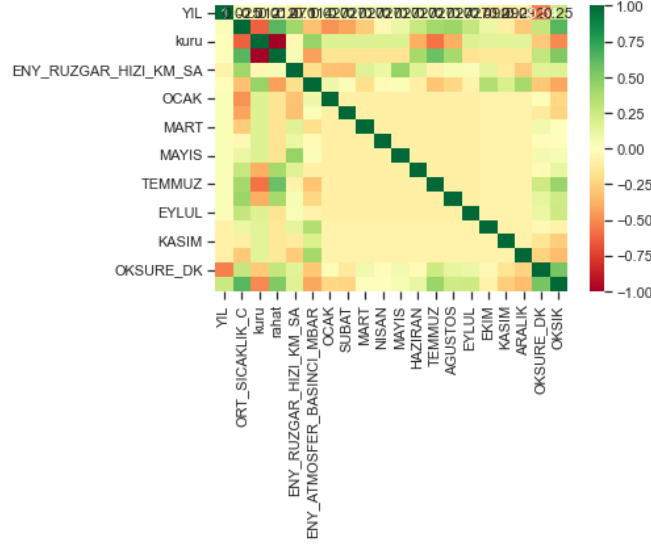
İL	YIL	DONEM	ORT SICAKLIK_C	NEM DURUMU	ENY_RUZGAR HIZI_KM_SA	ENY_ATMOSFER BASINCI_MBAR	OKSIK	OKSURE_DK
ERZURUM	2019	OCAK	-9,4	kuru	67	1025	0,79	174,78
ERZURUM	2019	SUBAT	-7,5	kuru	38	1032	0,52	72,41
ERZURUM	2019	MART	-4,1	kuru	63	1028	1,96	260,64
ERZURUM	2019	NISAN	3,8	kuru	54	1026	1,44	176,96
ERZURUM	2019	MAYIS	10,7	kuru	77	1025	1,68	190,5
ERZURUM	2019	HAZIRAN	14,9	rahat	62	1024	1,56	163,3
ERZURUM	2019	TEMMUZ	19,2	rahat	51	1022	2,71	303,89
ERZURUM	2019	AGUSTOS	19,5	rahat	66	1023	1,79	203,9
ERZURUM	2019	EYLUL	14,8	kuru	59	1024	1,54	213,8
ERZURUM	2019	EKIM	8,1	kuru	67	1032	1,03	139,84
ERZURUM	2019	KASIM	1,2	kuru	48	1028	0,59	72,44
ERZURUM	2019	ARALIK	-6,2	kuru	50	1038	0,44	40,75
ERZURUM	2020	OCAK	-11,1	kuru	36	1027	0,52	43,99
ERZURUM	2020	SUBAT	-8,4	kuru	62	1021	1	101,16
ERZURUM	2020	MART	-5,2	kuru	51	1017	0,48	39,29
ERZURUM	2020	NISAN	4,3	kuru	70	1025	0,37	31,73

Elektrik Dağıtım Şirketi Tablo-5 Kesinti Süreleri ve Sıklığı veri seti içerisinde yer alan OKSIK öznitelikleri enerji kesintisinden etkilenen bölgedeki abone sayısının o il/ilçe/mahalle kapsamındaki tüm abone sayısına oranını, OKSURE_DK ise enerji kesintisinden etkilenen bölgedeki abone sayısının etkilenen dakika cinsinden süre ile çarpımı sonrası o il/ilçe/mahalle kapsamındaki tüm abone sayısına oranını sayısal veri cinsinden ihtiva etmektedir. Birbiriyle ilişkili bu iki öznitelik verisine esas veri yoğunluğu ilişkisi Şekil 7 ile verilmektedir. Ayrıca bu görselde bağımsız ('X') ve bağımlı ('y') değişkenlerinin ham haliyle ortak topolojisi de yer almaktadır.



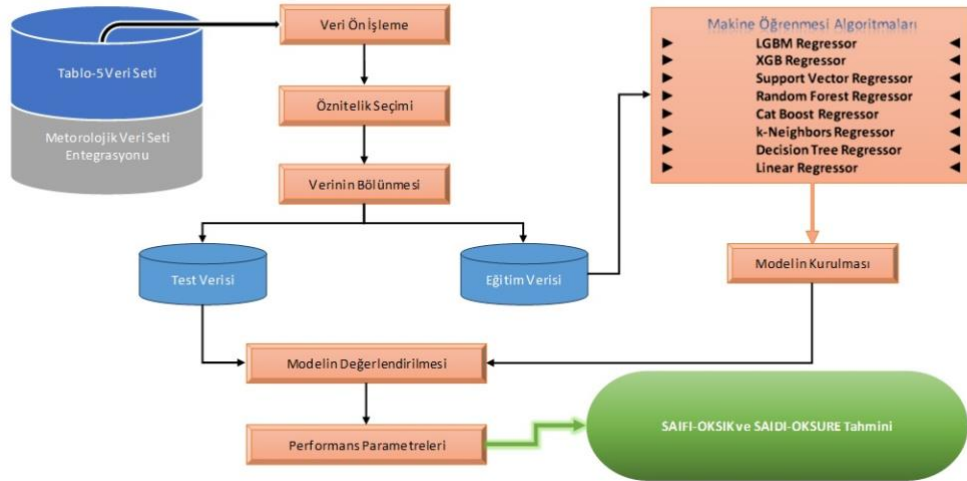
Şekil 7. Tablo-5 Kesinti Süreleri ve Sıklığı veri seti topolojisi.

Çalışmaya konu olan veri seti içerisinde yer alan özniteliklerin birbirleriyle olan korelasyon ikili yakınlıkları/ilişkileri aşağıda yer alan Şekil 8'deki korelasyon matrisi ile görselleştirilmiştir. Böylece bağımsız ('X') ve bağımlı ('y') değişkenler arasındaki birbirine göre var olan etkileşimler ile negatif/pozitif korelasyonlar temsil edilmektedir.



Şekil 8. Veri seti korelasyon ısı haritalama.

Gerçekleştirilen çalışmada genişletilmiş Kesinti Süreleri ve Sıklığı veri seti üzerinde LGBM Regressor, XGB Regressor, Support Vector Regressor, Random Forest Regressor, Cat Boost Regressor, k-Neighbors Regressor, Decision Tree Regressor, Linear Regressor olmak üzere 8 farklı regresyon modeli uygulanmıştır. Veri seti üzerinde yer alan SAIFI-OKSIK ve SAIDI-OKSURE çok sınıflı değerleri ayrı ayrı sınıf modeline dahil edilmiş olup toplamda 8 farklı model için 16 adet regresyon çalışması gerçekleştirilmiştir. Regresyon için önerilen modellere ait mimari diyagram yapısı Şekil 9'da verilmiştir.



Şekil 9. Kesinti Süreleri ve Sıklığı tahmininde farklı modellere ait temel mimari.

3.2. Veri Ön İşleme

Tüm yapay zekâ modellerini kurma aşamasındaki en önemli ve zorunlu adımlardan biri veriyi işlemeye uygun hale getirmektir. Veri ön işleme faaliyetlerini özetleyecek olursak; veri seti üzerinde veri temizleme, veri birleştirme, veri dönüştürme, veri indirgeme gibi birtakım işlemleri gerçekleştirerek, veri setini yapay zekâ biliminin alt kümesi olan makine öğrenmesi ya da derin öğrenme modellerinin çalışma kabiliyetlerine uygun hale getirebilmektir. Veri seti üzerinde gerçekleştirilen bu uygulamalar ile çalışmamız kapsamında da yer alan makine öğrenmesi modellerinin daha güvenilir, doğru ve başarılı regresyon tahminlerinde bulunması sağlanmıştır. Bu çalışmamızda yer alan veri setinde öncelikle, model işleme ve sonuçlara etkisi olmayacak “IL” özniteliği düşürülmüştür/silinmiştir. Çünkü üzerinde çalışılacak komple veri seti, hali hazırda tek pilot bölge olan Erzurum iline aittir. Ayrıca veri setinde bulunan satır x sütun bazındaki muhtelif nesnelere yüksek oranla eksik olanları veri boşluğu sebebiyle ana veri seti üzerinden düşürülmüştür/silinmiştir. Hücre bazında kısmi eksiklikleri bulunan değerler ise ortalama değer ile doldurularak anlamlı hale getirilmiştir.

Tüm yapay zekâ modellerinde olduğu gibi makine öğrenmesi modellerini kurmak için hem giriş hem de çıkış değişkenlerinin sayısal formda olması şarttır. Bu şart ile kategorik forma sahip öznitelik bazındaki bağımsız değişkenler dönem ve nem durumu, ikiden fazla duruma sahip olduğundan çoklu sayısal forma dönüştürülmüştür. Bu işlem için Python’un pandas ve sklearn preprocessing kütüphaneleri (One-Hot Encoder alt kütüphanesi) kullanılmıştır. Çalışmanın giriş-literatür taraması bölümünde anlatıldığı şekliyle, One-hot encoder uygulaması sonrasında ortaya çıkması muhtemel kukla değişken tuzağı (dummy variable trap), öğrenme ağırlığı bakımından yetersiz görülen öznitelik sütunlarında düşürme ve dönüşüm/dizi dilimleme işlemleri ile ortadan kaldırılmaktadır. Özellikle çoklu sınıfa sahip bu çalışmaya esas veri seti için riskli bir durum haline gelecek kukla değişken tuzaqları sebebiyle anlamlandırılmış 0-1 doğrulamasına sahip tanımlı bağımsız değişkenlerin (‘X’) muhtelif bölümleri veri setinden düşürülmüştür/dizi dilimlenmiştir. Böylece aşırı-az öğrenme karmaşasının önüne geçilmesinin yanında veri setinin tahmin mekanizmasını kuvvetlendirecek nesne takımları salt halleriyle bırakılmıştır. Ayrıca bu çalışmada, literatürde sadece katar (string) veriler için etiketleme amacıyla kullanılan sklearn preprocessing kütüphanesindeki LabelEncoder fonksiyonunun, ondalık sayısal (float) veriler üzerine uygulanması ile normal sayısal (integer) veriler elde edilerek, veri ön işleme adımlarına farklı bir yaklaşım kazandırılmıştır. İlgili işlem adımı OKSIK sınıf regresyon işlemlerinde, Linear Regressor ve Support Vector Regressor modelleri için OKSIK sınıf verisine, Decision Tree Regressor modeli için OKSURE öznitelik verisine başarılı bir şekilde uygulanmıştır.

Model tasarımının karmaşıklığını azaltma, yorumlanabilmesini ve sonuç kalitesini artırma amacıyla kullanılan yapı ise veri azaltma teknikleridir. Bu uygulamadan kasıt, muhtelif bileşen analizleri kullanılarak öznitelik çıkarma (feature extraction) işlemleridir. Ancak çalışma kapsamına dahil edilen Tablo 5 Kesinti Süreleri ve Sıklığı veri seti, 2019-2023 yılları arasında Aras Elektrik Dağıtım A.Ş. işletme sorumluluk bölgesinde yer alan Erzurum iline ait 267.571 adet örnek içeren Tablo 1 Kesintiler veri setinin, T.C. EPDK tarafından uygulamaya alınan matematiksel yaklaşımı ile indirgenmiş Kesinti Süreleri ve Sıklığı veri setinde yer alan 5 yıllık 60 ana örneği esas aldığından, böyle bir uygulama adımına gerek kalmamıştır. Kesinti Süreleri ve Sıklığı veri seti, uygulanan modeller açısından oldukça saf ve anlaşılır bir yapıdadır. Ek olarak uygulamaya alınacak öznitelik seçimi ve/veya öznitelik çıkarma faaliyetlerinin, model sonuçlarına aşırı öğrenme ve/veya yetersiz öğrenme şeklinde yansıdığı tespit edilmiştir.

3.3. Regresyon Algoritmaları

Yapay zekâ biliminin bir alt dalı olan makine öğrenimi, giriş veri seti özniteliklerinden sınıf tahminleri yapmak için giriş verilerini analiz ederek işlemleri öğrenen ve optimize eden programlanmış algoritmaları kullanır. Yeni verilerin mevcut veri setine dahil edilmesiyle birlikte algoritmalar daha doğru tahminler yapma ve kaliteli öğrenme eğilimindedir (Uddin ve ark., 2019). Makine öğrenimi algoritmaları, geçmiş deneyimlerden öğrenmek, büyük, yapılandırılmamış ve karmaşık veri kümelerinden faydalı kalıpları tespit etmek amaçlarıyla çeşitli istatistiksel, olasılık ve optimizasyon yöntemleri kullanır (Mitchell, 1997). Özetle makine öğrenmesi, istatistiksel ve matematiksel işlemler ile veriler üzerinden çıkarımlar ve tahminler yapan sistemlerin makineler ile modellenmesidir. Regresyon algoritmaları, bir veya daha fazla bağımsız değişkene dayalı olarak bağımlı bir değişkenin değerini tahmin etmek için kullanılan istatistiksel tekniklerdir. Bu algoritmalar ekonomi, finans ve mühendislik gibi alanlarda yaygın olarak kullanılmaktadır (Parlak ve ark., 2023).

Gerçekleştirilen çalışmada genişletilmiş Kesinti Süreleri ve Sıklığı veri seti üzerinde LGBM Regressor, XGB Regressor, Support Vector Regressor, Random Forest Regressor, Cat Boost Regressor, k-Neighbors Regressor, Decision Tree Regressor, Linear Regressor olmak üzere 8 farklı regresyon modeli uygulanmıştır. Veri seti ile model uyumluluğu göz önüne alınarak oluşturulmuş topolojiler doğrultusunda en yüksek başarı skorları hedeflenmiştir. Makine öğrenmesi özelliği taşıyan birçok model ve parametre, mevcut veri setine uygunluğu ile nihai başarı skorları açısından incelenmiş olup literatür araştırmaları doğrultusunda eleme yöntemine tabi tutulmuştur. Böylece metoda dair en başarılı ve efektif 8 model çalışmaya dahil edilmiştir. Veri seti üzerinde yer alan SAIFI-OKSIK ve SAIDI-OKSURE çok sınıflı değerleri ayrı ayrı sınıf modeline dahil edilmiş olup toplamda 8 farklı model için 16 adet regresyon çalışması gerçekleştirilmiştir.

3.3.1. Light Gradient Boosting Machine Regresyonu (LGBM)

LGBM Algoritması, karar ağacı temelinde bir öğrenme algoritmasıdır. Diğer modellere kıyasla en büyük avantajları yüksek hız, büyük veri setlerini irdeleme ve işleme, düşük bellek kapasitesiyle yüksek işlem gücü ve yüksek başarı oranları şeklinde özetlenebilir. Bir makine öğrenmesi metodu olarak LGBM Regresyon işlemi, veri boyutu ve çeşitliliğinin artması sebebiyle ortaya çıkarılmış, gelişmiş bir Gradient Boosting algoritması alternatifidir. Boosting işlemi temelde zayıf öğrenici düğüm ve katmanların daha kapsamlı öğrenicilere dönüştürülmesi işlemidir. Bu ve benzeri algoritmalar ile daha karmaşık veriler için daha hızlı eğitim ve daha yüksek doğruluk eldesi amaçlanmıştır. Bir makine öğrenmesi metodu olan ışık gradyanını artıran makine modeli (LightGBM) diğer toplu öğrenme metotlarına göre daha verimlidir. Model sıralama, sınıflandırma, regresyon, paralel eğitim ve birçok alanda kullanılır, daha verimli ve kararlı sonuçlar üretir (Guo ve ark., 2023).

3.3.2. eXtreme Gradient Boosting Regresyonu (XGBR)

XGB Algoritması hesaplama yöntemlerinde eğim artırma metotlarını kullanan, yarışmalarda elde ettiği başarılar ile dikkat çeken, yenilikçi yaklaşıma sahip öğrenme algoritmasıdır. Gradient Boosting algoritmasının muhtelif düzenlemeler ile optimize edilmiş yüksek performanslı daha gelişmiş halidir. Temel çalışma mantığı Gradient Boosting ile benzerdir. Model hatalarını azaltmaya yönelik geleneksel yöntemlerden farklı olarak, gereksiz model karmaşıklığını sürekli kontrol altında tutan düzenleme terimlerini içerir. Böylece modelin aşırı öğrenme/eksik öğrenme algısını ortadan kaldırarak genelleme yeteneği kazanmasına yardımcı olur. Büyük veri setlerine rahatça uyarlanabilen sütun blok yapısı ve sıralama yetenekleri mevcuttur. Böylece hız ve verimlilik açısından bir adım önde olmaktadır. Veri setine uydurulabilir esnek yapısı ile problemi çözmeye yönelik en iyi modeli ortaya çıkaracaktır. Out-of-core computing yetisi ile büyük veri seti işlemlerinde RAM sınırlamalarını göz ardı edebilecektir. Bir makine öğrenmesi metodu olarak XGB Regresyon işlemi, yüksek tahmin gücü, aşırı öğrenme engelleme, boş verileri yönetebilme, hızlı işlem gibi özelliklere sahiptir. Tianqi Chen ve Carlos Guestrin'in 2016 yılındaki akademik yayını ile makine öğrenmesi bilimine dahil olmuştur. Ölçeklenebilir yapıya sahip model uçtan uca ölçeklenebilir ağaç güçlendirme yöntemidir. Mevcut sistemlerden çok daha az kaynak kullanarak üstün sonuçlar eldesi için yazılım ve donanım optimizasyonu kullanan sistem ile milyarlarca örneğin ötesine ölçekleme yapılarak popüler algoritmalarından 10 kat daha hızlı çalışma imkânı sağlanmaktadır. Karar ağacı tabanlı algoritmalar arasında performansı en yüksek olanıdır (Chen ve Guestrin, 2016).

3.3.3. Support Vector Regresyonu (SVR)

SVM Algoritması genellikle sınıflandırma ve regresyon problemlerinin çözümünde tercih edilen geleneksel denetimli öğrenme yöntemlerinden biridir. Temelde, en fazla veri seti noktasını bir araya getirebilecek doğruları, polinomları, parabolleri, eğrileri ve hiperbolleri oluşturmaya çalışır. SVM temelli yapı, farklı noktaları sınıflar haline getirerek oluşturulabilecek en büyük marjin hattını oluşturmaya odaklanmaktadır. Bir makine öğrenmesi metodu olarak SV Regresyon işlemi, doğrusal ve doğrusal olmayan regresyon tipleri için tahmin ve eğri uydurma yönünden popüler bir modeldir. SVR, temelde destek vektör makinelerinin (SVM) bileşenlerini temel alır. SVR optimizasyonu giriş bölgesinin boyutuna bağlı olmadan yüksek boyutlu veri uzaylarında avantajlara sahiptir. Hiperdüzlem veri noktalarını belirgin bir şekilde ayıran n öznitelik uzayında meydana gelen hiperdüzleme yönelik daha yakın noktalar meydana getirir (Drucker ve ark., 1996; Parbat ve Chakraborty, 2020).

3.3.4. Random Forest Regresyonu (RFR)

RF Algoritması denetimli sınıflandırma algoritmaları grubunda yer almaktadır. Hem regresyon hem de sınıflandırma problemlerinin çözümünde kullanılabilen yapı, birden fazla karar ağacı üreterek hedef başarı oranını yükseltmeyi esas alır. Meydana getirilen birden çok karar ağacı yapısı içerisindeki en başarılı yapı öne çıkarılır. Oluşturulan ağaç sayısı arttıkça, model çeşitliliği ve başarı vurgusu artırılabilecektir. Karar ağaçları ile arasındaki temel fark, kök düğümün bulunması ve düğüm bölme işlemlerinin gelişi güzel/rastgele yapıyor olmasıdır. Bir makine öğrenmesi metodu olarak RF Regresyon işlemi, birden çok karar ağacı perspektifi ile daha uyumlu modeller ortaya çıkararak yüksek başarı oranlı tahminler oluşturan bir regresyon modelidir. RFR, tahmin edilecek parametre için ikili bölünme dizisini tekrarlı olarak ayırır. Tipik olarak RFR, son uç olan bir yaprakta tek bir tahmin verisi bulunana kadar ikili bölünmelere devam eder. Aşırı uyumu ortadan kaldırmak için yaprak başına minimum örnek uygulaması gerçekleştirilir. Ancak ham veri seti içindeki doğrudan gözlem aralığı dışı tahmin edilemez (Graw ve ark., 2021).

3.3.5. Category Boosting Regresyonu (CBR)

CB Algoritması, standart Gradient Boosting performansını artırmak amacıyla geliştirilen sınıflandırma ve regresyon temelli model topolojisidir. Veri hazırlığı aşaması çok kısa bir sürede tamamlanarak boş verilere kategorik veri kodlama yöntemi uygulanmaktadır. Kategorik veriler ile yüksek performanslı çalışabilen yapı için ön veri hazırlığının yapılmaması tavsiye edilmektedir. Bu

husus hem öğrenme hızını hem de sonuç kalitesini doğrudan etkilemektedir. Bir makine öğrenmesi metodu olarak CB Regresyon işlemi, yüksek öğrenme hızı, sayısal, kategorik ve metin verilerine uygulanabilirlik, GPU desteği ve görsellik barındırması ile diğer model uygulamalarından ayrılmaktadır. Gradyan artırma işlemi ile standart karar ağaçlarının kayıp fonksiyonu düşürülmek için yinelemeli olarak eğitilir. Muhtelif makine öğrenmesi algoritmalarından farklı olarak, CBR küçük veri setleri ile çok iyi çalışır. Ancak küçük veri setlerinde meydana gelebilecek aşırı uyum, ideal parametre ve hiperparametre seçileri ile engellenebilir (Beskopylny ve ark., 2022).

3.3.6. k-Neighbors Regresyonu (kNR)

kN veya kNN Algoritması, veri özniteliklerinin birbirlerine olan benzerliklerini esas alan denetimli öğrenme alt başlığındaki sınıflandırma veya regresyon algoritmasıdır. Bir veri öznitelik grubuna en yakın özelliklere sahip k sayıdaki farklı küme hesaplanır, kümelere ve veriye ait bağımlı değişkenler üzerinden tahminlerde bulunulur. Bir makine öğrenmesi metodu olarak kN Regresyon işlemi, tahmin edilecek verinin diğer verilere olan yakınlığı üzerinden hareket ederek bir komşuluk ilişkisi kurar. Özetle, eğer iki bağımsız değişken, uzaklık tespitleri doğrultusunda diğer bağımsız değişkene göre yakın ise komşuluk grubu oluşturulur ve aynı özelliktedir denir. Anlaşılması çok kolay olan bir algoritma olmasının yanında test aşamasında tüm eğitim verilerini saklayan tembel yapılı bir algoritmadır (Song ve ark., 2017).

3.3.7. Decision Tree Regresyonu (DTR)

DTR Algoritması sınıflandırma ve regresyon olarak iki ana başlığa ayrılabilir. Temel olarak, bağımsız değişkenleri ('X') bilgi kazancına göre belirli aralıklara ayırarak tahmin sürecinde bu aralığa ait muhtelif ortalama tekniklerini kullanır. Bağımlı değişken ('y') sorgusuna karşılık ortaya çıkarılan ortalama hesabı süreklilik içeren yapıda kesintiler ortaya çıkarır. Diğer regresyon modellerine göre en büyük farklılığı bu kesikli yapısıdır. Belli bir tahmin aralığı eğer benzerlik arz ediyorsa, sonuçlarda aynı tahminler doğrultusunda gerçekleşecektir. Bir makine öğrenmesi metodu olarak DT Regresyon işlemi, bir karar aşamasında regresyonu gerçekleştirmek için muhtelif özelliği birlikte kullanan çok aşamalı ve hiyerarşik karar şemasına (ağaç modeline) sahip algoritmalarıdır. Ağaç yapısı tüm verilerin dahil olduğu bir kök düğümden ve sonrasında bölünmüş olduğu dahili düğümlerden oluşur (Xu ve ark., 2005).

3.3.8. Linear Regresyonu (LR)

LR Algoritması, denetimli öğrenme modelleri arasında yer alan, bağımsız değişkenler ('X') ile bağımlı değişkenler ('y') arasındaki uyumu en iyi şekilde ortaya çıkaran doğrusal hattı meydana getiren, kullanım amacına göre sınıflandırma ya da regresyon modelleridir. Girdi ve çıktı değişkenleri arasındaki istatistiksel ilişki uyumuna en uygun matematiksel denklem modeli tanımlanarak ortaya çıkarılan eğri ile sonuçlara dair başarı metrikleri görselleştirilebilir. En ideal yapı, doğrusallık barındıran veri seti modellerinde söz konusu olabilir. Lineerlik, normallik, eş varyanslılık, bağımsızlık, normal hata dağılımı, otokorelasyon varlığının oluşturulmaması/hataların bağımsızlığı esaslarını barındıran model için en uygun veri seti kullanımı göz önünde bulundurulmalıdır. Aksi takdirde, veriye uygun olmayan model seçimi ile başarı metrikleri asgari seviyelerde kalacaktır. Bir makine öğrenmesi metodu olarak L Regresyon işlemi, nicel veri tahmini için basit ve ideal bir yaklaşımdır. Diğer regresyon modelleri genel olarak bu yöntem üzerinden türetilmiştir (James G. ve ark., 2023). Tek değişkenli modeller birden fazla açıklayıcı değişkene sahip modele dönüştürülür. Böylece probleme esas tahminleme işlemi için doğrusal metotlar daha net sonuçlara öncülük eder (Montgomery D. C. ve ark., 2021).

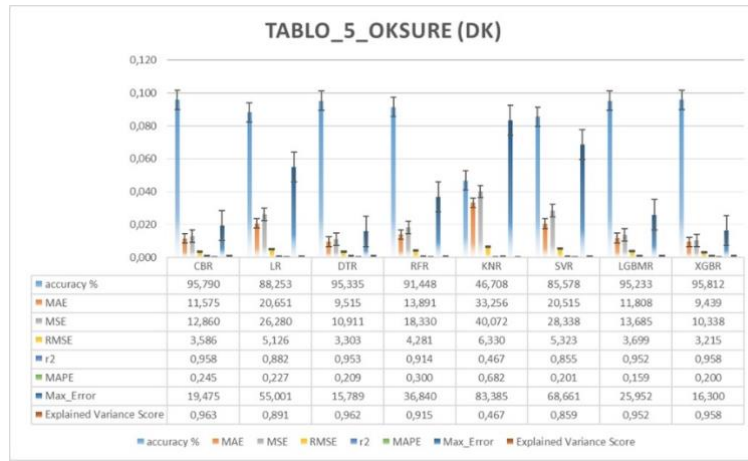
4. Bulgular

Bu çalışmada oluşturulan regresyon modellerinin performansları, 8 ayrı performans değerlendirme parametresi ile karşılaştırılmıştır. Bu parametreler; doğruluk (accuracy), ortalama mutlak hata (mean absolute error), ortalama karesel hata (mean squared error), ortalama karesel hatanın karekökü (root mean square error), belirleme katsayısı (r^2), ortalama mutlak yüzde hatası (mean absolute percentage error), maksimum hata (maximum error) ve açıklanabilen varyans puanı (explained variance score) şeklindedir.

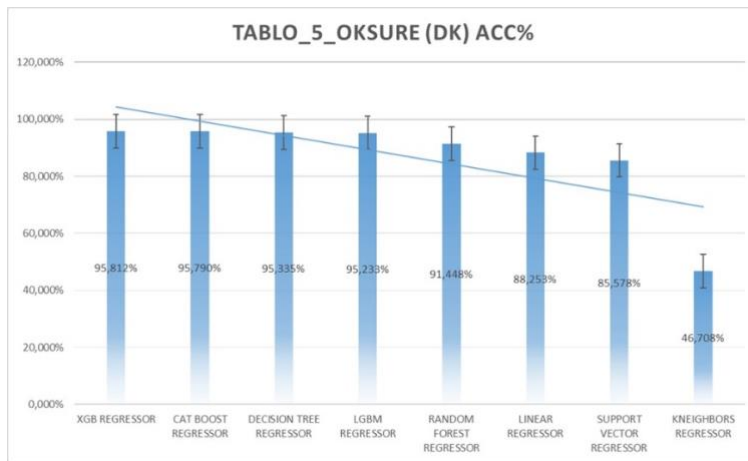
4.1. Farklı Regresyon Modellerinin Karşılaştırılması ve Başarı Değerlendirmesi

Veri seti ön işleme adımları sonrasında model uygulamasına hazır hale getirilen veri seti, tüm regresyon modelleri için %80'i eğitim ve %20'si test verisi olacak şekilde iki bölüme ayrılmıştır. Veri setinde yer alan bağımlı değişkenler ('y') birbirlerinden farklı çok sınıflı nicel (sayısal) yapıda olduğundan, çok sınıflı regresyon modelleri öne çıkarılmıştır. Ayrıca aynı veri seti içerisinde yer alan OKSURE ve OKSIK öznitelikleri ayrı ayrı bağımlı değişken ('y') olarak değerlendirildiğinden, her regresyon modellemesi için farklılıklar söz konusudur. Bu farklılıklar, hiperparametre optimizasyonları ile sağlanmıştır. Uygulanan optimizasyon tekniklerinde GridSearchCV ve

RandomizedSearchCV uygulamaları esas alınmıştır. GridSearchCV, cross-validation ile test edildiğinde her ne kadar yüksek maliyetlerde işlem hacmi ortaya koymuş olsa da bu çalışmaya dahil edilen uygulama olmuştur (Nirmal ve ark., 2024). RandomizedSearchCV uygulamasının işlem maliyeti düşürmesi adına her olasılığı değerlendirmemesi ve ekonomik bir yol izlemeye çalışması başarı oranlarını büyük ölçüde etkilemiş, hatta manuel parametre denemeleri bu kapsamdan daha başarılı olmuştur. Bu çalışmada yer verilen regresyon modellerinin tamamı aynı amaç için işlem yürütse de farklı karakteristiklere sahip olduklarından öğrenme oranları (learnig rate), n tahminciler ($n_{estimators}$), alan derinlikleri (depth), rastgele ölçkleme (random_state) farklı değerlerde seçilmiştir. Bu yöntemler ile ortaya çıkarılan ve bağımlı değişken ('y') olarak OKSURE verilerinin seçimi doğrultusunda oluşan metot ve başarı parametrelerine ait sonuçlar Şekil 10'da ve Şekil 11'de sunulmuştur.

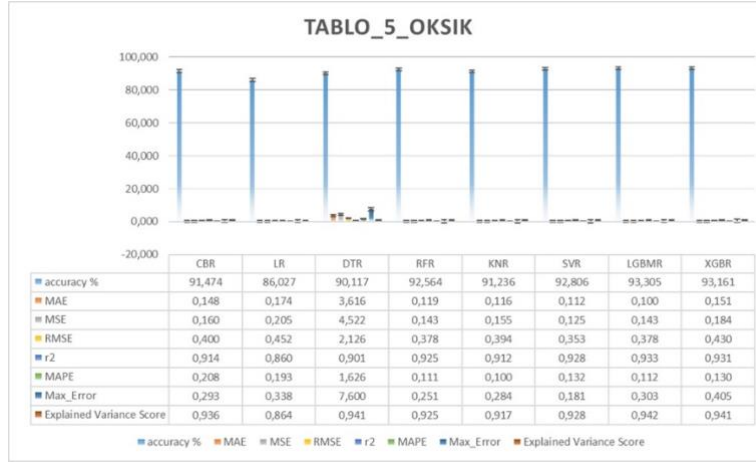


Şekil 10. OKSURE regresyon tahminine esas model seçimi ve başarı parametreleri.

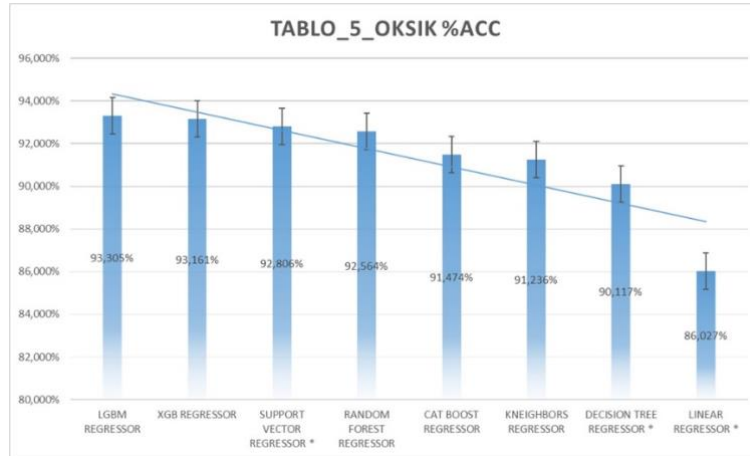


Şekil 11. OKSURE regresyon tahminine esas doğruluk oranları.

Başarı oranları ispat edilen regresyon yöntemleri, bağımlı değişken ('y') olarak OKSIK verilerinin seçimi doğrultusunda yeniden hiperparametre optimizasyonuna tabi tutulmuştur. Yeni metoda ait başarı parametreleri Şekil 12'de ve Şekil 13'te sunulmuştur.

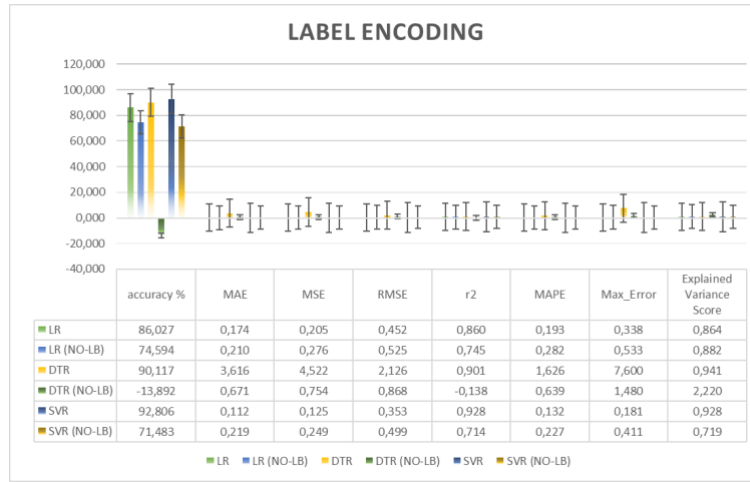


Şekil 12. OKSIK regresyon tahminine esas model seçimi ve başarı parametreleri.

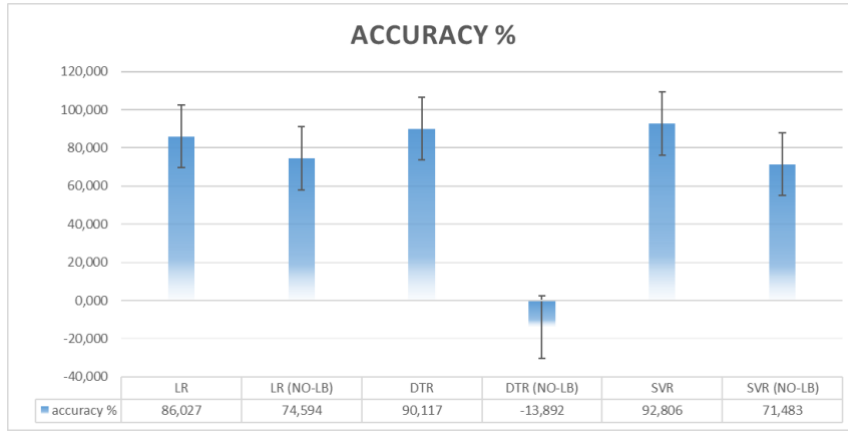


Şekil 13. OKSIK regresyon tahminine esas doğruluk oranları.

OKSIK verileri bağımlı değişken olarak modellere dahil edildiğinde karşılaşılan en büyük problem LR, DTR ve SVR metotlarındaki düşük başarı oranı ile karşılaşılmıştır. Önerilen yöntemin diğer çalışmalardan en büyük farkı ise bu noktada gerçekleştirilmiştir. Katar (string) veriler için etiketleme amacıyla kullanılan LabelEncoder fonksiyonunun, ondalık sayısal (float) veriler üzerine uygulanması ile normal sayısal (integer) veriler elde edilerek, veri ön işleme adımlarına farklı bir yaklaşım kazandırılmış, sayısal verilere düzeltme yürütülerek Şekil 14 ve Şekil 15'teki başarı oranları ortaya konmuştur. Bağımlı değişken ('y') OKSIK verisiyken DTR modelinde, bağımsız değişken ('X') OKSURE verisiyken LR ve SVR modellerinde LabelEncoder uygulaması öncesi başarı parametreleri Çizelge 2'de ayrıntılı olarak sunulmuştur.



Şekil 14. OKSIK regresyon tahminine esas DTR-LR-SVR modellerine label encoding uygulaması sonrası başarı parametreleri.



Şekil 15. OKSIK regresyon tahminine esas DTR-LR-SVR modellerine label encoding uygulaması sonrası doğruluk oranları.

Çizelge 2. LabelEncoder fonksiyonu uygulaması öncesi ve sonrası başarı parametreleri

BAŞARI PARAMETRESİ	LR (LE)	LR (NON-LE)	DTR (LE)	DTR (NON-LE)	SVR (LE)	SVR (NON-LE)
ACCURACY %	86,027	74,594	90,117	-13,892	92,806	71,483
MAE	0,174	0,210	3,616	0,671	0,112	0,219
MSE	0,205	0,276	4,522	0,754	0,125	0,249
RMSE	0,452	0,525	2,126	0,868	0,353	0,499
R2	0,860	0,745	0,901	-0,138	0,928	0,714
MAPE	0,193	0,282	1,626	0,639	0,132	0,227
MAX_ERROR	0,338	0,533	7,600	1,480	0,181	0,411
EXPLAINED VARIANCE SCORE	0,864	0,882	0,941	2,220	0,928	0,719

Çalışma sonucunda, SAIFI-OKSIK çok sınıflı tahmin için en iyi model doğruluğu LGBM Regressor ile %93,305 olarak elde edilirken SAIDI-OKSURE çok sınıflı tahmin için en iyi model doğruluğu XGB Regressor ile %95,812 olarak elde edilmiştir. Nihai sonuçlar göstermektedir ki aynı

sistematik olarak kurulan modeller için bağımlı değişken ('y') sınıf değerleri OKSIK iken başarı ortalaması, OKSURE iken elde edilen başarı ortalamasından daha yüksek sonuçlandırılmıştır.

5. Sonuçlar ve Öneriler

Bu çalışmada elektrik dağıtım sektöründe meydana gelmiş şebeke arızalarına ait veriler üzerinden regresyon işlemleri uygulanarak, geleceğe yönelik kesinti sıklığı ve süresi tahminlenmiş olup etkin müdahale ve kestirimci-önleyici faaliyetler için öngörü oluşturulmuştur. Aras Elektrik Dağıtım A.Ş. veri envanteri ve WeatherSpark web sayfasından elde edilen veri seti üzerinde farklı regresyon algoritmaları kullanılarak modeller geliştirilmiştir. Veri seti, 2019-2023 yılları arasında Aras Elektrik Dağıtım A.Ş. işletme sorumluluk bölgesinde yer alan Erzurum iline ait 267.571 adet örnek içeren Kesintiler veri setinin T.C. EPDK tarafından uygulamaya alınan matematiksel yaklaşımı ile indirgenmiş Kesinti Süreleri ve Sıklığı veri setinde yer alan 5 yıllık 60 ana örneği esas almaktadır. Veri setini modele hazırlama için uygulanan muhtelif veri ön işleme adımları sonrasında LGBMR, XGBR, SVR, RFR, CBR, kNR, DTR, LR olmak üzere 8 farklı regresyon modeli uygulanmıştır. Veri seti üzerinde yer alan 2 farklı çok sınıflı değerleri ayrı ayrı sınıf modeline dahil edilmiş olup toplamda 8 farklı model için 16 adet regresyon çalışması gerçekleştirilmiştir. Hiperparametre optimizasyonları haricinde, uygulanan modellerin diğer çalışmalardan en büyük farkı katar (string) veriler için etiketleme amacıyla kullanılan LabelEncoder fonksiyonunun, ondalık sayısal (float) veriler üzerine uygulanması ile normal sayısal (integer) veriler elde edilmesi, veri ön işleme adımlarına farklı bir yaklaşım kazandırılması ile sayısal verilerde düzeltme uygulaması olmuştur. Çalışma sonunda en iyi doğruluk oranları, LGBMR ile %93,3 ve XGBR ile %95,8 olarak elde edilmiştir. Literatürdeki çalışmalar genel olarak canlı şebeke üzerinde konuşlandırılmış sensör verileri ile haberleşme düzenleri üzerinden yapay zekâ uygulamalarını söz konusu etmektedir. Ancak canlı verilerin sensör hataları sebebiyle yanlış alınması, haberleşme düzenlerinde meydana gelen kesintiler gibi sebepler düşük başarı skorları veya sahanın gerçeğinden uzak önerileri söz konusu edecektir. Geçmişte meydana gelmiş olaylara dair reel sonuçların modellere dahil edilmesi, karakteristiği oluşmuş şebeke yapılarına ait gerçeğe en yakın öngörülerini ve durum yaklaşımlarını ortaya koyacaktır. Gelecekteki çalışmalarda, yoğun ve karmaşık veri sistematiğine sahip Kesintiler değerleri de veri kümesine dahil edilerek farklı yaklaşımlar elde edilebilir. Böylece il geneli makro yaklaşımların yanında en küçük yerleşim bölgesine dair mikro yaklaşımlarda söz konusu olacak, bakım-onarım ve yatırım-tesis faaliyetleri daha etkin bir şekilde gerekli noktaya gerekli olduğu zamanda gerçekleştirilecektir. Elektrik enerjisi tedariki kaliteli ve kesintisiz bir şekilde son kullanıcı noktasındaki abonelere ulaştırılarak, memnuniyet seviyesi de artırılabilecektir. Gelecekte bu problemler derin öğrenme ve yapay sinir ağları optimizasyonları ile çözülebilecektir.

Yazarların Katkısı

Tüm yazarlar çalışmaya eşit katkıda bulunmuştur.

Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada araştırma ve yayın etiğine uyulmuştur.

Kaynaklar

- Abdel-Nasser, M., Mahmoud, K., & Kashef, H. (2018). A novel smart grid state estimation method based on neural networks. *IJIMAI*, 5(1), 92-100.
- Beskopylny, A. N., Stel'makh, S. A., Shcherban', E. M., Mailyan, L. R., Meskhi, B., Razveeva, I., ... & Beskopylny, N. (2022). Concrete strength prediction using machine learning methods CatBoost, k-Nearest Neighbors, Support Vector Regression. *Applied Sciences*, 12(21), 10864.
- 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).
- Dashtdar, M., Dashti, R., & Shaker, H. R. (2018, May). Distribution network fault section identification and fault location using artificial neural network. *In 2018 5th international conference on electrical and electronic engineering (ICEEE)* (pp. 273-278). IEEE.
- De Santis, E., Mascioli, F. M. F., Sadeghian, A., & Rizzi, A. (2016). A dissimilarity learning approach by evolutionary computation for faults recognition in smart grids. *In Computational Intelligence: International Joint Conference, IJCCI 2014 Rome, Italy, October 22-24, 2014 Revised Selected Papers* (pp. 113-130). Springer International Publishing.
- Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. *Advances in neural information processing systems*, 9.
- Erdem, E., & Karamustafaoğlu, M., Elektrik Dağıtım Sektör Raporu. (2021, 31 Aralık) <https://www.erdem-erdem.av.tr/bilgi-bankasi/elektrik-dagitim-sektor-raporu>
- Ferreira, A. B., Leite, J. B., & Salvadeo, D. H. (2025). Power substation load forecasting using interpretable transformer-based temporal fusion neural networks. *Electric Power Systems Research*, 238, 111169.
- Graw, J. H., Wood, W. T., & Phrampus, B. J. (2021). Predicting global marine sediment density using the random forest regressor machine learning algorithm. *Journal of Geophysical Research: Solid Earth*, 126(1), e2020JB020135.
- Guo, J., Yun, S., Meng, Y., He, N., Ye, D., Zhao, Z., ... & Yang, L. (2023). Prediction of heating and cooling loads based on light gradient boosting machine algorithms. *Building and Environment*, 236, 110252.
- Hassani, H., Razavi-Far, R., & Saif, M. (2019, October). Locating faults in smart grids using neuro-fuzzy networks. *In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 3281-3286). IEEE.
- Jamali, S., Bahmanyar, A., & Ranjbar, S. (2020). Hybrid classifier for fault location in active distribution networks. *Protection and Control of Modern Power Systems*, 5, 1-9.
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Linear regression. *In An Introduction to Statistical Learning: With Applications in Python* (pp. 69-134). Cham: Springer International Publishing.

- Kankanala, P., Das, S., & Pahwa, A. (2013). AdaBoost $\{+\}$: An Ensemble Learning Approach for Estimating Weather-Related Outages in Distribution Systems. *IEEE Transactions on Power Systems*, 29(1), 359-367.
- Kurup, A. R., Martínez-Ramón, M., Summers, A., Bidram, A., & Reno, M. J. (2021, October). Deep learning-based circuit topology estimation and fault classification in distribution systems. In *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)* (pp. 01-05). IEEE.
- Mai, T. T., Nguyen, P. H., Haque, N. A., & Pemen, G. A. (2022). Exploring regression models to enable monitoring capability of local energy communities for self-management in low-voltage distribution networks. *IET Smart Grid*, 5(1), 25-41.
- Majidi, M., Etezadi-Amoli, M., & Fadali, M. S. (2014). A novel method for single and simultaneous fault location in distribution networks. *IEEE Transactions on Power Systems*, 30(6), 3368-3376.
- Mestav, K. R., & Tong, L. (2019, October). State estimation in smart distribution systems with deep generative adversary networks. In *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)* (pp. 1-6). IEEE.
- Mitchell, T. M., & Mitchell, T. M. (1997). *Machine learning* (Vol. 1, No. 9). New York: McGraw-hill.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- Mori, H., & Yokoyama, H. (2016). A hybrid intelligent method for estimating distribution network reconfigurations. *IFAC-PapersOnLine*, 49(27), 152-157.
- Mori, H., Aoyama, H., Yamanaka, T., & Urano, S. (2002, October). A fault detection technique with preconditioned ANN in power systems. In *IEEE/PES transmission and distribution conference and exhibition* (Vol. 2, pp. 758-763). IEEE.
- Niranjana, T., Swetha, D., Charitha, V., & Stephen, A. J. (2019). Predicting Burned Area Of Forest Fires. *IRJCS:: International Research Journal of Computer Science*, 6, 132-136.
- Nirmal, M. S., Patil, P., & Kumar, J. R. R. (2024). CNN-AdaBoost based Hybrid Model for Electricity Theft Detection in Smart Grid. *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, 100452.
- Parbat, D., & Chakraborty, M. (2020). A python-based support vector regression model for prediction of COVID19 cases in India. *Chaos, Solitons & Fractals*, 138, 109942.
- PARLAK, B. O., & YAVAŞOĞLU, H. A. (2023). Comparison of Regression Algorithms to Predict Average Air Temperature. *International Journal of Engineering Research and Development*, 15(1), 312-322.
- Perles, A., Camilleri, G., & Gleizes, M. P. (2017). Self-adaptive distribution system state estimation. In *Multi-Agent Systems and Agreement Technologies: 14th European Conference, EUMAS 2016, and 4th International Conference, AT 2016, Valencia, Spain, December 15-16, 2016, Revised Selected Papers* (pp. 202-216). Springer International Publishing.
- Song, Y., Liang, J., Lu, J., & Zhao, X. (2017). An efficient instance selection algorithm for k nearest neighbor regression. *Neurocomputing*, 251, 26-34.
- T.C. EPDK-Enerji Piyasası Düzenleme Kurumu. (Erişim Tarihi: 2 Mayıs 2023). <https://www.epdk.gov.tr>
- Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*, 19(1), 1-16.
- Xinrui, L., Yaoyao, Z., Peng, J., & Tianqi, L. (2017, May). Analysis of ice disaster failure considering the multi angle information modification for distribution network. In *2017 29th Chinese Control And Decision Conference (CCDC)* (pp. 6685-6690). IEEE.
- Xu, M., Watanachaturaporn, P., Varshney, P. K., & Arora, M. K. (2005). Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*, 97(3), 322-336.
- Yellagoud, S. K., & Talluri, P. R. (2019). A comparative evaluation of AI based fault location tools for electric distribution networks. *Int. J. Power Energy Syst*, 39(4), 177-183.
- Zhou, R., Li, Y., & Lin, X. (2025). A clustered federated learning framework for collaborative fault diagnosis of wind turbines. *Applied Energy*, 377, 124532.