

**T.C.
MANİSA CELAL BAYAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**YÜKSEK LİSANS
Elektrik – Elektronik Mühendisliği Anabilim Dalı
Elektrik – Elektronik Mühendisliği Bilim Dalı**

**GÜÇ TRAFOLARINDAKİ PLANSIZ KESİNTİLERİ ÖNLEMELİK İÇİN
ÖRNEK BİR KESTİRİMCİ BAKIM TABANLI ANOMALİ TAHMİNLEME
YÖNTEMİ GELİŞTİRİLMESİ**

TUGAY EREN GÜZELYOL

Doç. Dr. Yavuz ATEŞ



MANİSA-2025

**TUGAY
EREN
GÜZELYOL**

**Güç Trafolarındaki Plansız Kesintileri Önlemek İçin Örnek Bir Kestirimci
Bakım Tabanlı Anomali Tahminleme Yöntemi Geliştirilmesi**

2025

TAAHHÜTNAME

Bu tezin Manisa Celal Bayar Üniversitesi Lisansüstü Eğitim Enstitüsü Elektrik Elektronik Mühendisliği Ana Bilim Dalında akademik ve etik kurallara uygun olarak yazıldığını ve kullanılan tüm literatür bilgilerinin referans gösterilerek tezde yer aldığını, tamamen kendi çalışmam olduğunu, her alıntıya kaynak gösterdiğimi, tezin yazımında akademik ve etik kurallara aykırı herhangi bir yapay zekâ ve program kullanmadığımı beyan ederim.

Tugay Eren GÜZELYOL



İÇİNDEKİLER

İÇİNDEKİLER.....	III
SİMGELER VE KISALTMALAR DİZİNİ.....	IV
ŞEKİLLER DİZİNİ.....	V
TABLO DİZİNİ.....	VI
TEŞEKKÜR.....	VII
ÖZET.....	VIII
ABSTRACT.....	X
1. GİRİŞ.....	1
1.1 Elektrik Dağıtım Sistemlerinde Dönüşüm İhtiyacı.....	1
1.2 Akıllı Şebeke Teknolojileri Ve Küresel Gelişim Süreci.....	4
1.3 Türkiye’de Akıllı Şebeke Vizyonu ve Dijitalleşme Uygulamaları.....	6
2. LİTERATÜR TARAMASI.....	12
3. MATERYAL VE YÖNTEMLER.....	32
3.1 Veri Hazırlığı Ve Ön İşlem Süreçleri.....	33
3.1.1 Verilerin Temini Ve Kullanılan Özellikler.....	33
3.1.2 Veri Dağılımı.....	41
3.1.3 Veri Ön İşleme Adımları.....	45
3.1.4 Zaman Gecikmesi ve Kayan Pencere Metodolojisi (Lag Rolling)	
52	
3.2 Model Oluşturma Süreçleri.....	54
3.2.1 Eğitim Verisi Düzenlemeleri.....	54
4. SONUÇ VE ÖNERİLER.....	103
5. KAYNAKÇA.....	105

SİMGELER VE KISALTMALAR DİZİNİ

ABD	Amerika Birleşik Devletleri
ADT	Aykırı Değer Tespiti
ANN	Artificial Neural Networks
AUC	Area Under Curve
CNN	Convolutional Neural Networks
DGA	Dissolved Gas Analysis
DL	Deep Learning
DNN	Derin Sinir Ağları
ELDER	Elektrik Dağıtım Hizmetleri Derneği
EPDK	Enerji Piyasası Düzenleme Kurumu
EPIAŞ	Enerji Piyasaları İşletme A.Ş.
FN	False Negative
FP	False Positive
GRU	Gated Recurrent Unit
IEA	Uluslararası Enerji Ajansı
IEEE	Institute Of Electrical And Electronics Engineers
IQR	Interquartile Range
KNN	K-Nearest Neighbors
LGBM	Light Gradient Boosting Machine
LMLSM	Linear Multifractional Lévy Stable Motion
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MCC	Matthews Correlation Coefficient
ML	Machine Learning
MLP	Çok Katmanlı Algılayıcı, Multi-Layer Perceptron
ORVAE	Residual Variational Autoencoder
PR–AUC	Precision–Recall Curve – Area Under Curve
RF	Random Forest
RMSE	Root Mean Square Error
ROC–AUC	Receiver Operating Characteristic – Area Under Curve
SCADA	Supervisory Control and Data Acquisition
SMOTE	Synthetic Minority Over-Sampling Technique
SNN	Spike Neural Network
SVM	Support Vector Machine
TDO	Tasmanian Devil Optimization
TFT	Temporal Fusion Transformer
TN	True Negative
TP	True Positive
TSS	Time Series Split
VAE	Variational Autoencoder

ŞEKİLLER DİZİNİ

Şekil 1 Elektrik Tüketiminin Yıllara Göre Sektör Bazında Artışı [2].....	1
Şekil 2 Veri İşlem Süreçleri.....	32
Şekil 3 Transformatörlerin Haritalandırılması.....	42
Şekil 4 Transformatör Özelinde Kesinti Sayıları	43
Şekil 5 Yıllara Göre Kesinti Dağılımı	43
Şekil 6 Şehirlere Göre Mevsimsel Kesinti Dağılımı	44
Şekil 7 Transformatör Bazında Mevsimsel Kesinti Dağılımı	44
Şekil 8 Ayın Günlerine Göre Transformatör Bazlı Kesinti Dağılımı	45
Şekil 9 Aykırı Değer Tespiti (Outlier Detection, OD).....	47
Şekil 10 Sayısal Değişkenlerde Aykırı dağılımın BoxPlot ile Gösterimi.....	48
Şekil 11 Her Transformatör için Aykırı Değer Dağılımı	49
Şekil 12 Aykırı Değer Ve Aykırı Değer Olmayan Verilerin İlişkisi	50
Şekil 13 Aykırı Değer İçin Kullanılan Sütunların Korelasyonu Isı Haritası	52
Şekil 14 Sinus Verileri ile Haftanın Günleri	53
Şekil 15 G-Mean Parametresine Göre En Başarılı Modeller (Fold-1).....	74
Şekil 16 G-Mean Parametresine Göre En Başarılı Modeller (Fold-2).....	74
Şekil 17 G-Mean Parametresine Göre En Başarılı Modeller (Fold-3).....	75
Şekil 18 G-Mean Parametresine Göre En Başarılı Modeller (Fold-4).....	75
Şekil 19 G-Mean Parametresine Göre En Başarılı Modeller (Fold-5).....	76
Şekil 20 G-Mean Parametresine Göre En Başarılı Modeller (Fold-6).....	76
Şekil 21 G-Mean Parametresine Göre En Başarılı Modeller (Fold-7).....	77
Şekil 22 G-Mean Parametresine Göre En Başarılı Modeller (Fold-8).....	77
Şekil 23 G-Mean Parametresine Göre En Başarılı Modeller (Fold-9).....	78
Şekil 24 G-Mean Parametresine Göre En Başarılı Modeller (Fold-10).....	78
Şekil 25 G-Mean Parametresine Göre En Başarılı Modeller (Fold-11).....	79
Şekil 26 G-Mean Parametresine Göre En Başarılı Modeller (Fold-12).....	79
Şekil 27 G-Mean Parametresine Göre En Başarılı Modeller (Fold-13).....	80
Şekil 28 G-Mean Parametresine Göre En Başarılı Modeller (Fold-14).....	80
Şekil 29 G-Mean Parametresine Göre En Başarılı Modeller (Fold-15).....	81
Şekil 30 G-Mean Parametresine Göre En Başarılı Modeller (Fold-16).....	81
Şekil 31 Ham Veri ile G-Mean Parametresine göre En Başarılı Modeller	83
Şekil 32 Under Sampling ile G-Mean Parametresine göre En Başarılı Modeller.....	84
Şekil 33 SMOTE ile G-Mean Parametresine göre En Başarılı Modeller	85
Şekil 34 Class_Weight ile G-Mean Parametresine göre En Başarılı Modeller	86
Şekil 35 Model Bazında Ortalama Başarı Değerleri	88
Şekil 36 Varyant Bazında Ortalama Başarı Değerleri.....	89

TABLO DİZİNİ

Tablo 1 World Meteorological Organization – Hava Durumu Koşulları	34
Tablo 2 Sayısal Değerlerin Kesinti Sınıfı ile Korelasyonu	51
Tablo 3 Fold'ların Eğitim - Test Tarihleri.....	73
Tablo 4 Under Sampling ile G-Mean Parametresine göre En Başarılı Modeller.....	85
Tablo 5 SMOTE ile G-Mean Parametresine göre En Başarılı Modeller	86
Tablo 6 Class_Weight ile G-Mean Parametresine göre En Başarılı Modeller.....	87
Tablo 7 TSS ile G-Mean Parametresine göre En Başarılı Modeller.....	87
Tablo 8 Model Bazında Ortalama Başarı Değerleri	88
Tablo 9 Model Bazında Ortalama Başarı Değerleri	89



TEŐEKKÜR

Arařtırmamdaki her ařamada bana yardımcı olan deęerli tez danıřmanım Do. Dr. Yavuz ATEŐ 'e, Arř. Gör. Emir Kaan TUTUŐ 'a ve yüksek lisans eęitimim boyunca benden desteklerini esirgemeyen eřim Melisa GÜZELYOL 'a, annem Nur GÜZELYOL 'a, babam Erdal GÜZELYOL 'a, ablam EDA GÜZELYOL 'a ve Arř. Gör. Ceyda Boz'a teőekkürlerimi sunarım.

TUGAY EREN GÜZELYOL

Manisa, 2025



ÖZET

Yüksek Lisans

Tugay Eren GÜZELYOL
Manisa Celal Bayar Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik – Elektronik Mühendisliği Anabilim Dalı

Danışman: DOÇ.DR. YAVUZ ATEŞ

Elektrik dağıtım sistemlerinde güç transformatörlerinde meydana gelen arızalar, enerji sürekliliğini olumsuz etkileyerek müşteri memnuniyetinin azalmasına, işletme maliyetlerinin artmasına ve enerji arz güvenliğinin zayıflamasına neden olmaktadır. Bu durum, özellikle iklim koşullarına duyarlı altyapıya sahip bölgelerde daha kritik hale gelmektedir. Plansız arızaların büyük bir kısmının önlenabilir nitelikte olması, arıza gerçekleşmeden önce risklerin belirlenmesine ve bakım faaliyetlerinin proaktif bir yaklaşımla yürütülmesine duyulan ihtiyacı ortaya koymaktadır.

Bu tez, İzmir ve Manisa bölgelerinde bulunan güç trafolarında hava durumu verileri, altyapı özellikleri ve geçmiş kesinti kayıtları kullanılarak plansız kesintilerin önceden tahmin edilmesine yönelik özgün bir yaklaşıma odaklanmaktadır. Çalışmanın temel amacı, güç transformatörlerinde oluşabilecek arızaları önceden belirleyerek bakım süreçlerinin daha etkin yönetilmesini sağlamak ve enerji arz güvenliğini artırmaktır. Literatürdeki çalışmaların çoğu, tek bir algoritma veya sınırlı veri kullanımıyla yürütülmekte olup kapsam açısından dar kalmaktadır. Bu tezde ise birden fazla makine öğrenmesi ve derin öğrenme algoritması karşılaştırmalı olarak değerlendirilmiş ve model performansları farklı optimizasyon teknikleri ile geliştirilmiştir.

Çalışmada, 2021-2024 yılları arasında 34 güç trafosuna ait toplam 8.671 kesinti kaydı, bölgesel meteorolojik parametreler ve trafoya özgü teknik özellikler bir araya getirilerek kapsamlı bir veri seti oluşturulmuştur. Ön işleme sürecinde dengesiz sınıf dağılımı, aykırı değerlerin tespiti ve değişken seçimi gibi kritik adımlar uygulanmış;

veri seti hem geleneksel makine öğrenmesi algoritmaları (XGBoost, LightGBM, CatBoost) hem de derin öğrenme tabanlı modeller (LSTM, GRU) üzerinde test edilmiştir. Model başarıları G-Mean ve ROC-AUC gibi metriklerle değerlendirilmiş; en yüksek performans gösteren modellerin hiper parametreleri Optuna tabanlı optimizasyon yöntemleriyle iyileştirilmiştir. Elde edilen sonuçlar, hava durumu verileri ile arıza riskleri arasında ne tür bir ilişki olduğunu ortaya koymuş ve mevsimsel etkilerin kesinti ihtimalleri ile ilişkilerini göstermiştir. Bu çalışma, elektrik dağıtım sistemlerinde kestirimci bakımın etkinliğini arttırmaya yönelik, kesinti sürelerini azaltacak ve enerji sürekliliğine katkı sağlayacak önemli bir referans niteliği taşımaktadır.

Anahtar Kelimeler:

Kestirimci Bakım, Güç Transformatörü, Anomali Tespiti, Kesinti Tahmini, Makine Öğrenmesi, Hava Durumu Verisi Entegrasyonu, Dağıtım Şebekesi
2025, 111 sayfa

ABSTRACT

M.Sc.

Tugay Eren GUZELYOL

**Manisa Celal Bayar University
Graduate School of Applied and Natural Sciences
Department of Electrical Electronics Engineering**

Supervisor: Assoc. Prof. Dr. YAVUZ ATEŞ

Failures occurring in power transformers within electrical distribution systems adversely affect energy continuity, resulting in reduced customer satisfaction, increased operational costs, and weakened energy supply security. This issue becomes particularly critical in regions where infrastructure is sensitive to climatic conditions. Given that a significant portion of unplanned outages is preventable, it underscores the need for identifying risks prior to failures and managing maintenance activities through a proactive approach.

This thesis focuses on a novel approach aimed at predicting unplanned outages in power transformers located in the İzmir and Manisa regions by utilizing weather data, infrastructure characteristics, and historical outage records. The primary objective of this study is to anticipate potential transformer failures in advance, thereby enabling more efficient maintenance management and enhancing energy supply security. Most studies in the literature rely on either a single algorithm or limited data use, resulting in restricted scope. In contrast, this thesis comparatively evaluates multiple machine learning and deep learning algorithms and improves model performances through various optimization techniques.

In the study, a comprehensive dataset was created by integrating a total of 8,671 outage records from 34 power transformers between the years 2021 and 2024, regional meteorological parameters, and transformer-specific technical attributes. Critical preprocessing steps such as addressing class imbalance, detecting outliers, and feature selection were applied; subsequently, the dataset was tested with both conventional

machine learning algorithms (XGBoost, LightGBM, CatBoost) and deep learning-based models (LSTM, GRU). Model performances were evaluated using metrics such as G-Mean and ROC-AUC, and hyperparameters of the best-performing models were optimized using Optuna-based optimization methods. The obtained results revealed the type of relationship existing between weather data and failure risks and demonstrated the associations between seasonal effects and outage probabilities. This study serves as an important reference for enhancing predictive maintenance effectiveness in electrical distribution systems, reducing outage durations, and contributing to energy continuity.

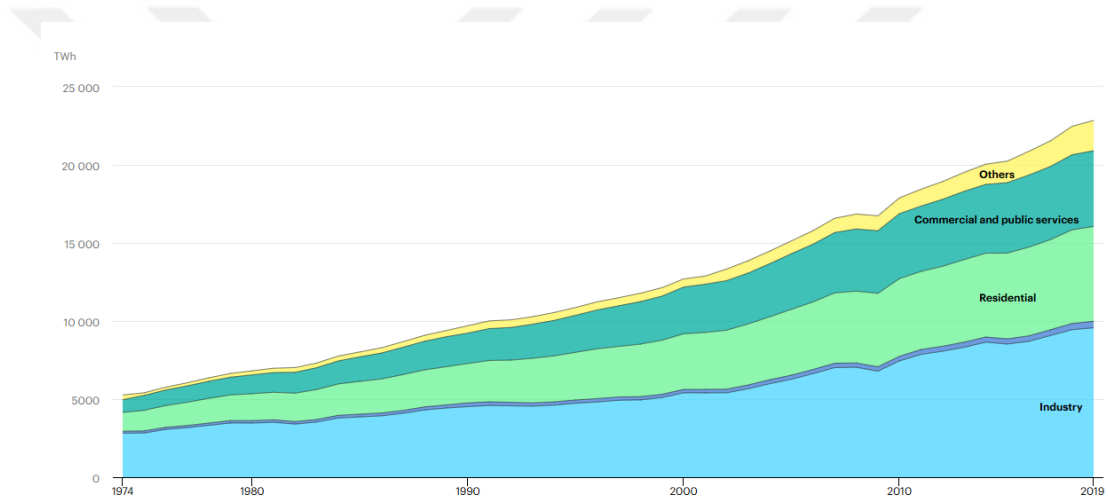
Keywords:

Predictive Maintenance, Power Transformer, Anomaly Detection, Outage Prediction, Machine Learning, Weather Data Integration, Distribution Network
2025, 111 pages

1. GİRİŞ

1.1 Elektrik Dağıtım Sistemlerinde Dönüşüm İhtiyacı

Elektrik enerjisi; modern toplumlarda ekonomik, teknolojik ve sosyal ihtiyaçların sürekliliğini sağlayan temel altyapı bileşenlerinden biridir. Sanayi ve endüstri faaliyetleri, sağlık hizmetleri, haberleşme, ulaşım gibi temel ihtiyaçlarımız ve internet gibi diğer ihtiyaçlarımızın sürdürülebilirliğini sağlayan elektrik enerjisi; güvenilir, kaliteli ve sürekli olarak temin edilmesi büyük önem taşımaktadır [1].



Şekil 1 Elektrik Tüketiminin Yıllara Göre Sektör Bazında Artışı [2]

Şekil 1’de belirtildiği gibi elektrik tüketimi dünya genelinde her geçen gün artmaktadır. Artan nüfus ve kentleşmenin yanı sıra, iklim değişikliği kaynaklı mevsimsel göçler, sanayileşme ve teknolojik ilerlemeler nedeniyle enerji tüketiminde yaşanan düzensiz artış ve kullanıcıların tüketim alışkanlıklarındaki dönüşüm, mevcut dağıtım sistemlerini ciddi anlamda zorlamaktadır [2].

Dağıtım sistemleri, enerji zincirinin nihai halkası olarak doğrudan son kullanıcı deneyimini etkileyen kritik bir kademeyi temsil etmektedir. Geleneksel elektrik üretim santrallerinin yerleşim birimlerinden uzak konumlarda kurulması nedeniyle, üretilen enerjinin iletim ve dağıtım şebekeleri aracılığıyla son kullanıcılara ulaştırılması; enerji kayıplarına karşı savunmasız, arıza riski taşıyan ve esneklik açısından sınırlı bir süreçtir. İletim ve dağıtım hatlarındaki teknik kayıplar yalnızca sistem verimliliğini düşürmekle kalmayıp, aynı zamanda enerji maliyetlerinin artmasına ve karbon emisyonlarının yükselmesine yol açmaktadır [3].

Üstelik iletim mesafesi arttıkça sistem kırılganlaşmaktadır; örneğin fırtına veya yıldırım düşmesi gibi olaylar geniş çaplı kesintilere neden olabilmektedir [4], [5]. Dolayısıyla, geleneksel tek yönlü, merkezi üretime dayalı elektrik şebekeleri artık çağın gereklerini karşılayamamakta, bu sistemlerin yerini daha esnek, etkileşimli, dayanıklı ve çevreyle uyumlu şebeke altyapılarına bırakması kaçınılmaz bir zorunluluk haline gelmiştir.

Bir diğer zayıf nokta ise merkezi üretime olan bağımlılıktır. Merkezi üretim tesislerinde meydana gelen bir arıza ya da iletim hatlarında yaşanan bir kopma, geniş çaplı kesintilere yol açabilmektedir [5]. Bu tür yapısal problemler, elektrik dağıtım sistemlerinde köklü bir dönüşümün gerekliliğini ortaya koymaktadır. Söz konusu dönüşümün merkezinde ise “akıllı şebeke” kavramı yer almaktadır. Bu bağlamda, enerji sektöründe yaşanan dönüşüm ihtiyacı üç temel faktör etrafında şekillenmektedir:

- (i) Teknik gereklilikler,
- (ii) Ekonomik sürdürülebilirlik ve
- (iii) Çevresel sorumluluk.

Teknik olarak sistemin daha hızlı karar verebilen, kendini yenileyebilen, gerçek zamanlı izleme ve müdahale kapasitesine sahip olması beklenmektedir. Ekonomik olarak ise; verimlilik artışı, enerji kayıplarının azaltılması, işletme ve bakım maliyetlerinin düşürülmesi gibi hususlar ön plana çıkmaktadır. Çevresel açıdan ise; düşük karbon salınımı, yenilenebilir enerji kaynaklarının entegrasyonu ve enerji tüketiminde bilinçli yönetim esas alınmaktadır [1].

Akıllı şebekelerin küresel düzeyde benimsenmesinde yalnızca teknik değil, aynı zamanda politik ve toplumsal faktörler de belirleyici olmuştur [6]. Enerji dönüşümü politikaları, karbon nötr hedefleri, tüketici hakları ve enerjiye erişim gibi alanlardaki regülasyonlar, akıllı şebeke yatırımlarını teşvik eden bir ortam yaratmıştır. Uluslararası Enerji Ajansı'nın (IEA) 2024 Dünya Enerji Görünümü raporu bu dönüşümün temel dinamiklerini ve küresel enerji sektöründeki jeopolitik riskleri detaylı biçimde ortaya koymaktadır [7]. Raporda, özellikle gelişmekte olan ülkelerde hızla artan elektrik talebi, yenilenebilir enerji kapasitesinin her geçen gün artması ve şebeke altyapılarının modernizasyonuna yönelik yatırımlar ön plana çıkmaktadır.

Bu kapsamda, akıllı şebeke teknolojileri, enerji depolama sistemleri ve talep tarafı yönetimi gibi dijitalleşme süreçleri enerji sistemlerinin esnekliğini ve güvenilirliğini artırmakta; Amerika Birleşik Devletleri (ABD), Avrupa Birliği, Çin ve Güney Kore gibi büyük ekonomiye sahip ülkeler, temiz enerji teknolojilerinde yerli üretim teşvikleriyle birlikte elektrikli araç altyapısı ve şebeke esnekliği yatırımlarını artırmaktadır. Bu ülkelerde akıllı şebeke uygulamaları sayesinde; sistemin kendini toplama kapasitesi artmakta, arıza süresi ve sıklığı azalmaktadır. Ayrıca veri temelli bakım stratejileri, tüketici profillemesi ve enerji arz-talep optimizasyonu gibi işlevler sayesinde sistem performansı anlamlı şekilde iyileştirilmektedir [8] , [9].

Elektrik Dağıtım Hizmetleri Derneği (ELDER) tarafından hazırlanan “Türkiye Akıllı Şebekeler 2023 Vizyon ve Strateji Belirleme Projesi” ne göre akıllı şebekelerin özellikleri şu şekilde sıralanmıştır [10]:

- (i) Dağıtık üretimin yaygınlaştırılması,
- (ii) Tüketicinin piyasaya katılmasına teşvik edilmesi,

- (iii) Depolama sistemlerinin, elektrikli araç şarj istasyonlarının şebekeye çift yönlü bağlanabilmesi,
- (iv) Esnek ve uyarlanabilir olmasıyla, asgari düzeyde insan müdahalesi gerektirmesi,
- (v) Tedarik sürekliliğini artışı ile enerji kalitesinin korunması.

Türkiye’de elektrik dağıtım sektörüne son 15 yıl içinde önemli yatırımlar yapılmış olmakla birlikte; Enerji ve Tabii Kaynaklar Bakanlığı ve Enerji Piyasası Düzenleme Kurumu (EPDK) öncülüğünde dijitalleşme stratejileri geliştirilmiş olsa da sahada uygulama kapasitesi hâlen gelişim aşamasındadır. Bu nedenle Türkiye’de, akıllı şebeke dönüşümünün hızlandırılması, dijital altyapının güçlendirilmesi ve kestirimci bakım gibi ileri düzey uygulamaların yaygınlaştırılması, dağıtım sektörünün geleceği açısından kritik öneme sahiptir.

Enerji sektörünün daha güvenli, verimli, esnek ve çevre dostu hale gelmesi için şebeke yapılarının dijitalleşmesi ve akıllı altyapılarla desteklenmesi kaçınılmazdır. Bu dönüşümde dağıtım sistemleri özelinde uygulanacak politikalar, teknolojik yatırımlar ve veri temelli modeller belirleyici olmaktadır. Kestirimci bakım uygulamaları ve bu uygulamalara benzeyen makine öğrenmesi (ML) veya derin öğrenme (DL) temelli modeller sayesinde, yapılacak veri analizi çalışmaları dönüşümün merkezinde yer alan yapılar olarak ortaya çıkmaktadır.

1.2 Akıllı Şebeke Teknolojileri Ve Küresel Gelişim Süreci

ABD Enerji Bakanlığı tarafından “Akıllı Şebeke” için yapılmış olan tanımda; “Elektrik sistemini büyük ölçekli üretimlerden dağıtım ve nihai elektrik tüketicisine kadar modernize etmek için dijital teknolojileri kullanan ve 7 adet performans bazlı fonksiyon ile tanımlanan sistemdir.” İfadesi kullanılmıştır [11]. Tanımda belirtilen 7 fonksiyon şu şekildedir:

- (i) Tüketici katılımının artırılması
- (ii) Tüm üretim ve depolama kaynaklarının entegrasyonu,
- (iii) Yeni pazar ve operasyon modellerinin desteklenmesi,

- (iv) Yüksek kalitede güç sunumu,
- (v) Varlık optimizasyonu ve operasyonel verimlilik,
- (vi) Bozucu olaylar sonrası sistemin kendi kendini toparlayabilmesi,
- (vii) Fiziksel ve siber saldırılara karşı dirençli ve esnek bir yapı oluşturulmasıdır.

Bu tanım çerçevesinde değerlendirildiğinde, akıllı şebekelerin sadece bir otomasyon sistemi değil; veri odaklı, müşteri merkezli ve çevresel etkileri minimize eden bir enerji ekosistemi olduğu anlaşılmaktadır. Bu sistemler, klasik elektrik şebekelerinin sınırlı etkileşimli yapısının aksine, her bir bileşenin bilgi ürettiği, birbirini etkilediği ve sistemin bütünüyle gerçek zamanlı olarak yönetilebildiği bir yapı sunmaktadır [11].

Akıllı şebeke teknolojilerinin temel bileşenleri arasında; akıllı sayaç sistemleri, enerji yönetim sistemleri, dağıtım yönetim sistemleri, talep tarafı yönetimi, mikro şebekeler, yenilenebilir enerji kaynaklarının entegrasyonu, enerji depolama sistemleri, SCADA (Supervisory Control and Data Acquisition), kesinti yönetim sistemleri, elektrikli araçlar ve şarj altyapısı, siber güvenlik çözümleri ve yapay zekâ temelli karar destek sistemleri yer almaktadır. Bu bileşenlerin her biri, sistemin farklı yönlerine etki ederek operasyonel verimliliği artırmakta, enerji arz-talep dengesini optimize etmekte ve kullanıcı deneyimini iyileştirmektedir.

Akıllı şebekelerin küresel gelişim süreci, büyük ölçüde enerji sistemlerinin karşı karşıya olduğu ortak sorunlara yanıt verme amacıyla şekillenmiştir. Enerji arz güvenliği, iklim değişikliği ile mücadele, karbon salımının azaltılması ve yenilenebilir enerji kaynaklarının artan payı, geleneksel şebekelerin yetersiz kaldığı alanlar olarak öne çıkmaktadır. Örneğin, fotovoltaik sistemler veya rüzgâr türbinleri gibi yenilenebilir kaynakların üretim kapasitesi anlık değişikliklere maruz kalabildiğinden, bu kaynakların elektrik şebekesine etkin entegrasyonu esnek ve duyarlı şebeke altyapılarını gerektirmektedir [12].

Avrupa Yeşil Mutabakatı 2050 Karbon Nötr Hedefi doğrultusunda enerji dönüşümünün bir aracı olarak konumlandırmıştır. Horizon 2020 ve Horizon Europe programları kapsamında birçok akıllı şebeke projesi desteklenmiş; özellikle mikro şebekeler, enerji toplulukları, üreten tüketici (prosumer) yapılar ve blok zincir tabanlı enerji paylaşım sistemleri gibi yenilikçi yaklaşımlar ön plana çıkmıştır. Örneğin, Grid4EU gibi kapsamlar sayesinde Avrupa genelinde kullanıcıların enerji üretim ve tüketim davranışları üzerine veri temelli optimizasyonlar geliştirilmiştir [13].

Güney Kore, akıllı şebeke alanında ulusal stratejilerini SKT, KEPCO gibi şirketlerin liderliğinde yürüterek, enerji verimliliği, uzaktan izleme, siber güvenlik ve elektrikli araç entegrasyonu gibi alanlarda bütüncül bir dönüşüm gerçekleştirmiştir. KEPCO, 2030 yılına kadar ülke genelinde akıllı şebeke uygulamasına tamamen geçmeyi hedeflemektedir [1]. Bu kapsamda, akıllı sayaçların kurulumu, yenilenebilir enerji kaynaklarının entegrasyonu ve elektrikli araçların şebekeye bağlanması gibi projeler yürütülmektedir.

1.3 Türkiye’de Akıllı Şebeke Vizyonu ve Dijitalleşme Uygulamaları

Türkiye’nin enerji politikaları, 2000’li yıllardan itibaren arz güvenliğinin temini, piyasa mekanizmalarının serbestleştirilmesi ve hizmet kalitesinin artırılması hedefleri doğrultusunda şekillenmiştir. Bu kapsamda elektrik dağıtım sektöründe önemli yapısal reformlar gerçekleştirilmiştir. 2001 yılında yürürlüğe giren 4628 sayılı Elektrik Piyasası Kanunu ile başlatılan serbestleşme süreci, 2013 yılında 6446 sayılı Elektrik Piyasası Kanunu’nun yürürlüğe girmesiyle güncel ve kapsamlı bir düzenleyici çerçeveye kavuşturulmuştur [14], [15] Özellikle dağıtım hizmetlerinin özelleştirilmesi ve kamu-özel ortaklığı modellerinin benimsenmesiyle, sektörde rekabetçi, verimli ve şeffaf bir yapı oluşturulması amaçlanmıştır.

Bu dönüşüm süreci, elektrik sektörünün finansal yükünün hafifletilmesinin yanı sıra, hizmetlerin etkinliği ve kalitesinin artırılmasını da hedeflemektedir. Ancak nüfus artışı, kentleşme oranlarının yükselmesi, iklim değişikliği etkileri ve enerji talebindeki artış gibi dinamikler, mevcut dağıtım altyapısının sürdürülebilirliğine yönelik yeni gereksinimleri beraberinde getirmiştir. Bu gelişmeler ışığında Türkiye’nin akıllı şebeke vizyonu, yalnızca teknik bir dönüşüm değil aynı zamanda stratejik ve yönetimsel bir gereklilik olarak önem kazanmıştır.

“Türkiye Akıllı Şebekeler 2023 Vizyon ve Strateji Belirleme Projesi” hedeflerine ulaşmak adına dijitalleşme, veri yönetimi, uzaktan izleme ve karar destek sistemlerinin dağıtım sektörüne entegre edilmesi temel strateji olarak benimsenmiştir [10]. Türkiye, bu kapsamda enerji altyapısını modernleştirmek konusunda hem regülasyon hem de saha uygulamaları açısından önemli adımlar atmıştır. Bu adımlar arasında dijital sistemlerin yaygınlaştırılması, operasyonel verimliliği artırmakla ile müşteri deneyimini de iyileştirmeyi hedeflemektedir. Akıllı sayaçlar ile fatura okuma, tüketim analizleri ve kullanıcı bazlı talep yönetimi mümkün hale gelmiştir. Özellikle pandemi sonrası dönemde sahada temasın azaltılması ve uzaktan müdahale ihtiyacı, dijital sistemlerin önemini daha da artırmıştır. Dağıtım şirketleri, mobil uygulamalar ve mesaj yoluyla bilgilendirme sistemleri ile kullanıcıya kesinti bilgisi, enerji tüketim grafikleri ve sistem durumu hakkında anlık bildirimler sunabilmektedir.

Dijitalleşme sürecinin bir diğer boyutu da veri analitiği ve yapay zekâ uygulamalarıdır. Yapılan [16] , [17] , [18] ve benzeri çalışmalarda yer alan kesinti tahminleme süreçlerinde; hava durumu verileri, arıza geçmişi, trafo yüklülüğü, direk tipi, hat uzunluğu gibi çok sayıda değişken bir araya getirilerek arıza tahminleme modelleri oluşturulmaktadır. Bu modellerde XGBoost (eXtreme Gradient Boosting), CatBoost (Categorical Boosting), LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit) gibi algoritmalar kullanılarak hem arıza frekansı hem de müdahale süresi azaltılması; bu sayede SAIDI (Sistem Ortalama Kesinti Süresi Endeksi, System Average Interruption Duration Index) ve SAIFI (Sistem Ortalama Kesinti Sıklığı Endeksi, System Average Interruption Frequency Index) göstergelerinde iyileşme sağlanması incelenmektedir. Ayrıca, bu kapsamda yapılacak projeler ile kesintilerin hava durumu sebebi ile “mücbir” kesinti olarak nitelendirilmesinde kolaylık sağlanabilmektedir. Özellikle EPDK tarafından yapılan denetimlerde, hava koşullarına dayalı arıza raporları veri destekli biçimde sunulabilmekte ve böylece dağıtım şirketleri, şeffaflık ilkesi kapsamında regülatif risklerini azaltmaktadır.

Enerji ve Tabii Kaynaklar Bakanlığı'nın 2023 ve 2024 yılı stratejik planlarında; yenilenebilir enerji kaynaklarının daha etkin entegrasyonu, dağıtımda kayıpların azaltılması, müşteri memnuniyetinin artırılması ve yerli teknoloji kullanımı gibi hedeflerle birlikte akıllı şebeke altyapısının genişletilmesi öngörülmektedir. Özellikle Türkiye Elektrik İletim A.Ş. (TEİAŞ) ve Enerji Piyasaları İşletme A.Ş. (EPIAŞ) tarafından desteklenen sistemlerde; veri standardizasyonu, piyasa tahminleme modelleri ve şebeke esnekliği gibi kavramlar giderek daha fazla yer bulmaktadır [19].

Türkiye'nin mevcut enerji altyapısı ve sahip olduğu potansiyel göz önünde bulundurulduğunda, akıllı şebeke dönüşümünün ivme kazanması halinde sektörde önemli kazanımlar sağlanacağı açıkça görülmektedir. Bu dönüşümün etkileri; daha düşük kesinti frekansı, operasyonel maliyetlerde azalma, verimlilik artışı, müşteri memnuniyetinde yükseliş, çevresel sürdürülebilirlik ve regülasyonlara uyum olarak birden fazla alana etki edeceği şekilde ortaya çıkmaktadır.

Ancak tüm bu gelişmelere rağmen, Türkiye'de akıllı şebeke uygulamalarının önünde bazı yapısal engellerin bulunduğu da unutulmamalıdır. Bunlar arasında; dijital altyapı eksiklikleri, bölgesel eşitsizlikler, veri paylaşımındaki yetersizlikler, teknik personel eksikliği ve mevzuatsal boşluklar sayılabilir. Üniversite-sanayi iş birliklerinin teşvik edilmesi, AR-GE desteklerinin artırılması ve ulusal veri politikalarının netleştirilmesi önem arz etmektedir. Ayrıca, dağıtım şirketlerinin dijitalleşme performanslarının periyodik olarak ölçülmesi, karşılaştırmalı raporlar hazırlanması ve en iyi uygulamaların paylaşılması; sektör genelinde dönüşüm sürecini hızlandıracaktır.

Elektrik dağıtım sistemlerinde uygulanan bakım stratejileri günümüzde sistem güvenilirliği, hizmet sürekliliği ve müşteri memnuniyetini sağlama açısından kritik öneme sahiptir. Geleneksel yöntemlerin yerini almaya başlayan kestirimci bakım yaklaşımları, büyük veri analitiği, yapay zekâ ve makine öğrenmesi tekniklerinin gelişimiyle giderek daha erişilebilir hale gelmiştir. Kestirimci bakım, ekipman ve sistemlerde meydana gelebilecek arızaların gerçekleşmeden önce tahmin edilmesi ve bakım faaliyetlerinin en uygun zamanda yapılmasını amaçlamaktadır.

Bu yaklaşım sayesinde ekipmanların çalışma süreleri artırılmakta, bakım maliyetleri düşürülmekte, kesinti süreleri minimize edilmekte ve iş gücü verimliliği iyileştirilmektedir. Enerji dağıtımında yoğun olarak kullanılan transformatörler, enerji hatları, röleler, kesiciler ve direkler gibi ekipmanların sürekli izlenebilmesi ve geçmiş arıza kayıtlarıyla analiz edilmesi mümkün hale gelirken; sıcaklık, nem, yağış miktarı ve rüzgâr hızı gibi hava durumu verilerinin de bakım kararlarına entegrasyonu, süreci daha dinamik ve veriye dayalı hale getirmektedir.

Kestirimci bakım uygulamalarının önemli avantajlarından biri, saha ekiplerinin vardiya planlamalarında optimizasyon sağlamasıdır. Geçmiş arıza kayıtları, hava durumu tahminleri ve ekipman performans verileri birlikte analiz edilerek, hangi bölgelerde ve hangi zaman dilimlerinde arıza riskinin yüksek olacağı öngörülebilmektedir. Böylece, bakım ekipleri proaktif bir şekilde, arıza gerçekleşmeden önce gerekli bölgelere yönlendirilmekte ve iş gücü kaynaklarının verimli kullanımı sağlanmaktadır. Ayrıca, yüksek talep dönemlerinde iş gücü ihtiyaçlarının önden belirlenmesiyle dönemsel olarak vardiya düzenlemeleri yapılabilmektedir. Bu durum hem çalışan verimliliğini artırmakta hem de kesinti sürelerini azaltmaktadır.

Enerji dağıtım hizmetleri, özellikle geniş coğrafyada bulunan ya da kırsal bölgelerde bulunan alanlarda bakım ekiplerinin rota seçimleri nedeniyle zaman, yakıt ve maliyet kayıpları yaşayabilmektedir. Makine öğrenmesi tabanlı kestirimci bakım sistemleri, olası arıza risklerini belirleyerek bakım müdahalelerini önceliklendirilmiş olarak sıralar ve Google Haritalar, Uygulama Programlama Arayüzleri (Application Programming Interface), Coğrafi Bilgi Sistemleri (Geographic Information System) ve gerçek zamanlı trafik verileri gibi kaynaklarla entegre çalışarak, bakım ekiplerinin en verimli rotayı kullanmalarına olanak tanır. Bu yöntem, müdahale sürelerini kısaltırken fiziksel kaynakların kullanımında da optimizasyon sağlamaktadır. Ayrıca, genellikle düzenli aralıklarla veya arıza sonrası müdahale şeklinde gerçekleşen mevcut bakım stratejileri, yüksek bakım maliyetlerine ve plansız kesintilere neden olmaktadır.

Kestirimci bakımın uygulanmasıyla birlikte, bakım faaliyetleri yalnızca ihtiyaç duyulan durumlarda gerçekleştirilerek, gereksiz operasyonel maliyetlerin önüne geçilmektedir. Örneğin, transformatör gibi kritik ekipmanların düzenli periyotlarla değil, anormal sıcaklık artışı, titreşim veya yağ analizi sonuçlarına göre kontrol edilmesi hem işçilik hem de ekipman yıpranmasını azaltmaktadır. Bunun yanı sıra yedek parça planlaması optimize edilerek stok maliyetleri düşürülmekte ve tekrarlayan arızaların kök neden analizleri yapılarak belirli ekipman gruplarının kullanımından kaçınılabilmektedir.

Enerji dağıtım şirketleri için müşteri memnuniyeti, hizmet kalitesinin sürdürülebilirliği ve mevzuata uyum açısından önem arz etmektedir. Plansız elektrik kesintileri özellikle mesken kullanıcıları, sanayi tesisleri ve hastaneler gibi kritik tüketicilerde ciddi memnuniyetsizlik yaratmaktadır. Kestirimci bakım uygulamalarının sağladığı proaktif yaklaşım sayesinde potansiyel arızalar önceden tespit edilerek müdahale edildiğinde, kesinti sıklığı ve süreleri azalmakta, bu durum müşterilerin memnuniyet düzeyini yükseltmektedir. Ayrıca, planlı bakım çalışmaları öncesinde kullanıcıların bilgilendirilmesine benzer bir uygulamanın potansiyel arıza durumlarında yapılması, müşteri şikâyetlerini ve çağrı merkezleri üzerindeki yükü önemli ölçüde azaltmaktadır.

Enerji sektöründe hizmet kalitesinin en temel göstergeleri olan SAIDI ve SAIFI değerleri, sırasıyla kullanıcıların yıllık ortalama kesinti sürelerini ve sıklıklarını ölçmektedir. Kestirimci bakım uygulamaları, bu performans göstergelerini iyileştirme potansiyeline sahiptir. Özellikle meteorolojik verilerle desteklenen kestirimci bakım sistemleri; fırtına, kar yağışı, yıldırım gibi mücbir sebeplerden kaynaklanan arızaların analiz edilerek raporlanmasını kolaylaştırmaktadır. Bu sayede dağıtım şirketleri, meydana gelen kesintileri bilimsel verilere dayandırarak EPDK gibi regülatör kurumlara raporlama yapabilmekte ve cezai yaptırımlardan korunabilmektedir.

Kestirimci bakım sistemlerinin başarısı büyük ölçüde kullanılan veri kaynakları ve uygulanan analiz yöntemlerine bağlıdır. Bu bağlamda veri temelli istatistiksel analizlerin yanı sıra makine öğrenmesi ve derin öğrenme yöntemleri de sıklıkla tercih edilmektedir.

Sonuç olarak kestirimci bakım, elektrik dağıtım sektöründe teknik ve stratejik açıdan çok boyutlu faydalar sağlamaktadır. Arızaların önlenmesi, hizmet sürekliliğinin sağlanması, bakım maliyetlerinin düşürülmesi, iş gücü verimliliğinin artırılması, müşteri memnuniyetinin yükseltilmesi ve regülasyonlara uyumun güçlendirilmesi gibi alanlarda etkinliğini kanıtlayan bu yaklaşımın, gelecekte daha geniş çapta uygulanacağı öngörülmektedir. Kestirimci bakım sistemlerinin yerel koşullara uygun hale getirilmesi, yapay zekâ ve ileri analitik yöntemlerle desteklenmesi ve operasyonel sistemlerle entegre edilmesi sayesinde enerji sektörünün dijital dönüşümünde vazgeçilmez bir unsur haline gelmesi beklenmektedir.



2. LİTERATÜR TARAMASI

Elektrik dağıtım şebekelerinde süreklilik ve güvenilirliği sağlamak açısından kritik öneme sahip olan güç transformatörlerinde meydana gelen arızalar, geniş çaplı enerji kesintilerine ve ekonomik kayıplara neden olabilmektedir. Enerji talebinin her geçen gün artması ve iklim değişikliğinin yol açtığı aşırı hava koşulları, altyapı elemanlarının güvenilirliğini yeniden değerlendirmeyi zorunlu kılmıştır. Bu çerçevede, akıllı şebeke uygulamaları üzerine yapılan ulusal ve uluslararası çalışmalar incelenmiştir. Türkiye’de ve diğer ülkelerde akıllı şebeke kapsamında yapılması gereken çalışmalar, iklim değişikliği vasıtasıyla meydana gelen plansız kesintilerin tespiti, hava durumu ve diğer olaylar ile korelasyonu incelenmiş, SAIDI ve SAIFI değerleri incelenmiş olup bazı politika önerileri ve gereksinimler ortaya çıkartılmıştır [1], [6], [20], [21].

Geleneksel önleyici bakım stratejileri, periyodik kontrollerle arıza ortaya çıktıktan sonra müdahale etmeye odaklanmakta ve bu durum plansız elektrik kesinti sürelerinin ve bakım maliyetlerinin artmasına, vardiya planlamalarında eksikliklere yol açmaktadır [8]. Bu yetersizliği gidermek amacıyla literatürde kestirimci bakım yaklaşımı ön plana çıkararak, makine öğrenmesi ve derin öğrenme yöntemleriyle çok boyutlu veri analizine dayalı öngörü modelleri geliştirilmiştir [22], [23], [24]. Bu kapsamda yürütülen kestirimci bakım çalışmaları, özellikle büyük veri analitiği ve yapay zekâ uygulamalarının gelişmesiyle birlikte önemli bir dönüşüm geçirmektedir. Bu süreçte, geçmiş dönemlerde sadece arıza sonrası müdahale ve periyodik bakım stratejileriyle sınırlı olan bakım faaliyetleri, artık proaktif bir yapıya kavuşmuştur. Böylece, şebeke güvenilirliği ve enerji sürekliliği büyük ölçüde iyileştirilmektedir.

Elektrik kesintileri, şebekede yürütülen planlı bakım faaliyetlerinden veya şebekede meydana gelen beklenmedik durumlar sonucu oluşan plansız durumlardan kaynaklanabilir. Planlı kesintiler genellikle şebeke iyileştirmeleri, ekipman yenileme veya rutin bakım faaliyetlerini içermekte olup, genellikle önceden kullanıcıya bildirilen programlı müdahalelerdir. Plansız kesintiler ise genellikle öngörülemeyen teknik arızalar, hava koşullarına bağlı dış etkenler veya insan kaynaklı yönetimsel hatalardan kaynaklanmaktadır [6], [25]. Bu noktada, plansız kesintileri önceden tahmin etmek ve müdahale etmek üzere geliştirilen kestirimci bakım yaklaşımları literatürde yoğun şekilde incelenmektedir.

PACME Çalışma Grubu tarafından yapılan çalışmada, bağımlı ve ortak mod kesintilerin elektrik sistemlerindeki güvenilirlik analizlerinde ihmal edilen ancak kritik öneme sahip bileşenler olduğunu ortaya koyarak, zaman serisi tabanlı kestirimci bakım ve anomali tahmin modeller bağlamında çevresel ve sistemik risklerin daha doğru değerlendirilmesi için bir kavramsal çerçeve sağlamaktadır [26].

Kestirimci bakım uygulamalarında yapay zekâ teknolojilerinin kullanımına ilişkin temel bileşenler, güvenilirlik gereksinimleri ve geleceğe dönük eğilimler sistematik biçimde ele alınan bir başka çalışmada; sensör verilerinin ön işlenmesi, ML ve DL algoritmaları, karar verme modülleri, iletişim ve entegrasyon altyapısı ile kullanıcı arayüzü ve raporlama katmanlarını tanımlayarak kapsamlı bir bakış sunulmasını sağlamaktadır [27]. Literatürde yer alan LSTM, CNN, XGBoost gibi modellerin karşılaştırılarak avantajları ve maliyet-etkinlik dengelerinin analiz edildiği bir çalışmada; veri kalitesi, etik hususlar, gerçek dünyada geçerlilik ve insan-makine etkileşimi alanlarındaki kısıtlar da vurgulanmıştır.

[27]'e ek olarak, yapılan çalışmada arıza verilerinin LS-SVM ile zaman serisi tahmini ve iki katmanlı optimizasyonla maliyet ile plansız kesinti kayıplarının birlikte minimize edilmesini sağlamıştır [28]. Özellikle arıza-kesinti-maliyet fonksiyonunun kullanımı, bakım frekansının ekonomik ve operasyonel etkilerini eş zamanlı değerlendirme imkânı sunarak, kestirimci bakım süreçlerinin gerçek dünya uygulamalarında optimize edilmesine güçlü bir metodolojik katkı sağlamaktadır.

Bir başka çalışmada ise izolatörlerin durum indekslerine dayalı öngörücü bakım metodolojisiyle bakım kaynaklarının etkin dağıtımına imkân tanımıştır. Yapılan çalışmada, polimer izolatörlerin çok boyutlu fiziksel ve kimyasal testlerinden elde edilen verileri birden fazla durum indeksine dönüştürerek, kalan ömür tahmininde esnek ve kapsamlı bir yaklaşım sunmaktadır. Bu indeks tabanlı bakım stratejisi, saha uygulamalarında bakım kaynaklarının etkin dağıtımını sağlayarak operasyonel fayda yaratmakta ve bakım zamanlamasını optimize ederek ekipman sağlığı izlenimi açısından oldukça önemlidir [29].

Bir başka çalışmada ise trafo buşing izolasyon durumu üzerinden arıza eğilimini tahmin etmeyi amaçlamıştır. Bilgi entropisi ile belirlenen göstergelerin ağırlıklandırıldığı “*Entropy-Weight TOPSIS*” yöntemi kullanılarak güncel test verilerinin ideal izolasyon durumlarına olan uzaklığı niceliksel olarak hesaplanmış ve elde edilen yakınlık derecelerinden oluşan zaman serisi verisi, az sayıda örnekle başa çıkabilen gri model ileriye dönük olarak öngörülmüştür [30]. Çalışma, göstergelerin ağırlıklandırılmasında insanî öznelliklerden arınmak için entropi temelli nesnel ağırlık belirlemesi yaparken, TOPSIS ile mevcut izolasyon durumunu dörtlü standardizasyon içinde sınıflandırmış; ardından gri modelle örneklem sayısının az olduğu durumlarda dahi güçlü trend tahmini sağlamıştır. Ancak çalışma; sadece buşing arızalarına odaklandığı ve veri büyüklüğü olarak sınırlı üç test kaydıyla yürütüldüğü için genelleştirilebilirlik açısından kısıtlıdır.

Transformatörlerdeki izolatörler ve burçların izolasyon durumu üzerinden yapılan çalışmalardan farklı olarak, güç trafolarındaki arızaların önceden tespiti amacıyla, trafo yağındaki çözünmüş gaz (Dissolved Gas Analysis, DGA) verilerinin zaman serisi özelliklerini kullanarak LSTM sinir ağı tabanlı bir tahmin modeli geliştirmeyi hedefleyen araştırmada, bir aylık DGA izleme verisinden hareketli pencere yöntemiyle oluşturulan eğitim ve test kümeleri üzerinde, pencere uzunlukları (3, 6, 12, 20) ile LSTM katman sayısı ve düğüm sayısı (tek katman–6/12/20 düğüm, iki katman–12 düğüm) deneysel olarak karşılaştırılmıştır. Model performansı, beş gaz için hesaplanan MAPE ve RMSE metrikleri temel alınarak değerlendirilmiş; en düşük bileşik hata skorunu veren tek katmanlı, 12 düğümlü yapı tercih edilmiştir. Çalışma, veri setinin nispeten küçük olması, eksik veri tamamlama yöntemlerinin basit interpolasyonla sınırlı kalması ve modelin gerçek zamanlı uygulamalarda genellebilirliğinin henüz test edilmemiş olması gibi kısıtlamalara işaret etmektedir [22].

[22] ‘ye ek olarak güç transformatörlerinin fabrika çıkışı DGA verilerine dayalı arıza tanı tahminlerinde dengesiz sınıf problemini aşmak amacıyla, SMOTE ile oluşturulan sentetik örnekler ve derin sinir ağları (DNN) mimarisini hibrit bir yaklaşımla birleştiren yeni bir yöntem önermektedir. Araştırmada, transformatörlerden elde edilen veri seti SMOTE uygulamasıyla dengeye kavuşturulmuş; ardından çok katmanlı DNN modeli, yüksek seviyeli özellik çıkarımı yaparak %98,22 eğitim ve %94,6 test doğruluğu elde edecek şekilde yapılandırılmıştır [31].

Belçika elektrik iletim şebekesinde farklı zaman dilimlerinde yapılması planlanan bakım faaliyetlerinin güvenli bir şekilde gerçekleştirilebilirliğini tahmin etmek amacıyla; karar ağaçları, SVM ve Naive Bayes gibi sınıflandırma algoritmaları karşılaştırılmıştır. En yüksek başarıyı %98,4 doğruluk oranı ile Random Forest algoritması göstermiştir. Yöntem; sistemdeki yük, hava durumu ve üretim koşulları gibi parametrelere dayalı olarak bakım faaliyetlerinin riskli olup olmadığını ikili sınıflandırma problemi olarak modellemiştir. Çalışmada geliştirilen yaklaşım, bakım planlamasının yalnızca arıza tahmini değil, aynı zamanda şebeke bütünlüğü üzerindeki etkisini de göz önünde bulundurarak sistem güvenliğini optimize etmektedir [32].

Fang Xu tarafından yapılan çalışmada, elektrik kesinti olaylarının gerçek zamanlı izlenmesi ve kök neden analizine yönelik kapalı döngü bir platform tasarlanmıştır. MySQL tabanlı veri depolama yapılan bu sistemde; SCADA, akıllı sayaçlar ve saha terminallerinden toplanan çok kaynaklı heterojen verileri dinamik olarak birleştirerek veri ön işleme esnasında kesinti sıcak noktalarını tanımlayıp Attention-LSTM ile kesinti sürelerini öngörürken korelasyon ile olası neden-sonuç ilişkilerini keşfetmeyi amaçlamaktadır [33].

Kestirimci bakım çalışmalarını hava durumu verilerine, ekipman indeks verilerine veya DGA aracılığıyla da incelenmeleri ele alınmıştır. Ancak, kestirimci bakım çalışmalarında yer alan ML ve DL yaklaşımlar için verilerin anlamlandırılabilmesi de gerekmektedir. Bu yaklaşım kapsamında, veri ön işleme yapılması gereken birden fazla çalışma ya da metodoloji bulunmaktadır. “Label Encoding”, “One Hot Encode”, “Aykırı Değer Tespiti (OD)”, “Korelasyon”, İstatiksel testler, “Time-Series Split (TSS)” vb. birçok çalışma bulunmaktadır [33], [34], [35], [36], [37].

[38] kapsamında; iklim kaynaklı aşırı hava olaylarının eğilimini hem VAE tabanlı veri arttırımı ile, hem de Stacked - LSTM tabanlı çok adımlı tahmin modeliyle ele alarak, granüler ve uzun vadeli kesinti öngörüsü için metodolojik bir çerçeve sunulmaktadır. LSTM-VAE, verinin dengesiz dağılımını giderirken, “sliding-window” ile elde edilen kısa sekanslar Stacked mimariyle işlenerek 120 saatlik lead time’da yüksek doğrulukta tahminler gerçekleştirilmiştir.

[38]'dekinden farklı bir yaklaşım ile sentetik veri SMOTE algoritmasıyla veri dengesini sağlayıp DNN tabanlı modelle tanı doğruluğunu %94,6'ya çıkararak trafolarında dengesiz veri sorununa etkili bir çözüm sunulmuştur [39].

[40]'de ise, sınırlı sayıda kule örneğiyle "Few-Shot Class-Incremental Learning" tabanlı risk tahmin modeli geliştirilerek tayfun kaynaklı iletim hattı arıza olasılıklarını yüksek doğrulukla tahmin edilmiştir. Bu olasılıkları çok aşamalı stokastik optimizasyon çerçevesinde kullanarak jeneratörler, batarya enerji depolama sistemleri ve yük kesintisi planlamasında proaktif karar destek sistemi sunmuşlardır.

Farklı bir optimizasyon çalışmasında ise; önerilen optimize edilmiş tekrarlı rastgele alt örnekleme yöntemi, özellikle dengesiz veri setlerinde negatif örnek seçimi için klasik tek seferlik rastgele alt örneklemenin ötesine geçerek, genetik algoritmayla parametrelendirilmiş SVM tabanlı modelin AUC (Area Under Curve) tabanlı en iyi sürümünü seçmektedir [41].

Kulczycki ve Franus tarafından geliştirilen koşula bağlı yoğunluk kestirimi temelli aykırı değer tespit çalışmaları, kümeleme ve sınıflandırma yöntemleri, trafo kesinti veri setinde hem çevresel hem de zaman serisi tetikleyicilerle ilişkilendirilmiş uç noktaların hassas seçimini mümkün kılmaktadır. Bu yöntem, veri setinin mevsimsel ve anomali tetikleyiciler açısından daha homojen ve güvenilir hale getirilmesine katkıda bulunarak model eğitimi için uygun altyapı hazırlamaktadır [42].

Yao ve ark. geleneksel DL tabanlı transformer arıza tanı yöntemlerinin sınırlı ve dengesiz veri sorunlarıyla karşılaştığı bağlamda, iki yenilikçi bileşen birleştirilerek gerçek transformatör arızalarının yüksek doğrulukla tespit edilmesini incelemiştir. 1D sinyal verileri üzerinde Dilate Konvolüsyon Birimleri (DCU) ve Leaky, ReLU entegrasyonu ile geliştirdikleri Deep Residual Shrinkage Network (DRSN) mimarisiyle sinyal işleme ve özellik çıkarım verimliliğini artırmış; optimize edilmiş Residual Variational Autoencoder tabanlı veri çoğaltma ile az örnekle arıza veri setlerini genişleterek gerçek trafo arıza verisinde %97,14 doğruluk ve %96,96 F1 skoruna ulaşmıştır [43].

[44] , hava koşullarına bağlı güç kesintilerini yerel düzeyde çok adımlı olarak öngörebilmek amacıyla, veri dengesizliği sorununu VAE tabanlı çoğaltma (augmentasyon) ile gidermeyi ve ardından Stacked LSTM mimarisiyle uzun dönem tahmin performansını iyileştirmeyi hedeflemiştir. Öncelikle, orijinal zaman serisi verisi VAE aracılığıyla genişletilmiş; ardından Sliding Window tekniğiyle oluşturulan çok adımlı giriş seti, üç katmanlı Stacked LSTM modeline beslenmiştir. Başarı metriği olarak RMSE ve MAE kullanılarak, artırılmış veri seti üzerinde %15–20’ye varan iyileşme sağlandığı raporlanmıştır.

Bu ön işleme teknikleri, sonraki ML adımlarının doğruluğunu doğrudan etkilediğinden, izlenecek sistematik sıralama ve parametre optimizasyonu kritik önem taşımaktadır. Bu nedenle tez kapsamında yapılan çalışmada verilerin matematiksel değerlere dönüştürülmesi, Label Encoding, One- Hot Encoding, Aykırı Hata Tespiti, MinMax Scaler ve Robust Scaler, sonrasında ise SMOTE, Undersampling ve classweight çalışmaları gerçekleştirilerek sürecin ilerletilmesine karar verilmiştir.

Cruz ve Munasinghe, ABD’nin güneydoğu bölgesindeki ilçelerde kasırga kaynaklı elektrik kesintisi oranını bir sonraki gün için tahmin edebilmek üzere meteorolojik, elektrik tüketim profili ve sosyo-ekonomik verileri bir araya getirerek dört farklı ML modelin performansını değerlendirmektedir. Her ilçede bir günlük aralıklarla ortalama kesinti oranını hedef değişken olarak belirlemiş; veri ön işleme aşamasında üç günlük hareketli ortalamalar, korelasyon matrisi ile değişken seçimi ve beş katlı çapraz doğrulama tabanlı GridSearchCV ile hiperparametre optimizasyonu yapılmıştır. Modellerin başarısı MSE, RMSE ve MAE metrikleriyle ölçülmüş ve RF regresyonunun $5,93 \times 10^{-5}$ ’lik en düşük MSE değeriyle üstün performans sergilediği gözlemlenmiştir [45].

Güç trafolarının yük altında kademe deęiřtirici bileřenlerinde meydana gelen arızaları yüksek doęrulukla tespit etmek ve sınıflandırarak trafo güvenilirliğini arttırılması hedeflenen alıřmada; Tasmanian Devil Optimization (TDO) meta-sezgisel algoritması ile Spike Neural Network (SNN)'ü hibritleřtiren TDO-SNN yöntemini önererek; ilk ařamada normal SNN ile trafo durumunu saęlam veya arızalı olarak ayırt etmiř, ikinci ařamada ise TDO ile SNN parametrelerini optimize ederek arıza tiplerini sınıflandırmıřlardır. alıřma, MATLAB ortamında simülasyon verileriyle doęrulama yapılmıř olup, tek bir trafo tipi ve sadece yük altında kademe deęiřtirici arızaları üzerinde test edilmesi; gerek saha verilerine ve farklı arıza senaryolarına uygulanmamıř olması alıřmanın önemli sınırlılıkları arasındadır [46].

Lu Shua ve ark. tarafından yapılan alıřmada, kapsamlı sensör aęlarıyla toplanan ok boyutlu iřletim verilerini Transformer tabanlı DL modeliyle iřleyerek gü sistemlerinde yer alan ekipman arızalarını önceden tahmin etmeyi amalamıř; veri ön iřleme adımlarında gürültü giderme, eksik deęer tamamlama ve standartlařtırma gibi yöntemler uygulanmıř, ardından özenle seilmiř öznitelik seti Transformer modeline beslenerek zengin gemiř arıza verileriyle eęitilmiřtir. Modelin gerek zamanlı izlemeye uyarlanmasıyla potansiyel arıza riski tespit edildięinde otomatik erken uyarı mekanizması devreye girmiř; deneysel sonuçlara göre önerilen Transformer modeli, ortalama %97,19 doęruluk oranı ve arıza türü tahmininde %90,2'nin üzerinde başarı elde ederek SVM temelli karřılařtırmalı yöntemlere kıyasla belirgin bir üstünlük göstermiřtir [47].

Yuan ve ark. tarafından yapılan alıřma kapsamında in'in Guangdong eyaletine baęlı Xuwen bölgesinde meydana müřteri bazlı elektrik kesintileri ve direk hasarlarını tahmin eden yüksek özünürlüklü (1 km²) bir model alıřtırılmıřtır. Arařtırmada, ařırı hava olaylarının elektrik řebekelerine olan etkilerini önceden öngörmeye yönelik olarak iki ařamalı bir rastgele orman modeli geliřtirilmiřtir. İlk ařama, kesintinin olup olmayacaęını sınıflandırmakta; ikinci ařama ise tahmin edilen bölgelerdeki kesinti ve hasar miktarlarını regresyon ile tahmin etmektedir. alıřmada, müřteri kesinti modeli %5,1 ortalama görelî hata oranı ile alıřırken, direk hasarı modeli %9,6 oranında hata payı göstermiřtir. Önceki modellere kıyasla mekânsal özünürlüğü önemli ölçüde artıran bu yaklařım, özellikle personel ve malzeme konumlandırma stratejileri aısından karar destek sistemlerine deęerli katkılar sunmaktadır [48].

Aizpurua ve ark., rüzgâr enerjisi tabanlı üretim sistemlerinde görev yapan güç transformatörlerinin yaşam süresinin öngörülmesine yönelik olarak, olasılıksal ML temelli bir tahminleme çerçevesi önermektedir. Model Hırvatistan'daki Jelinak rüzgâr çiftliğinden elde edilen gerçek saha verileriyle doğrulanmıştır. Yapılan duyarlılık analizlerinde; ortam sıcaklığı, rüzgâr hızı, yük profili ve coğrafi koşullar gibi çevresel ve operasyonel değişkenlerin yaşlanma süreci üzerindeki etkileri analiz edilmiştir. Önerilen model, %0.47 medyan hata oranı ve %6–7 bandında değişen %80 tahmin aralığı hatası ile yüksek doğruluk sergilemiştir. Bu yönüyle çalışma, transformatörlerin arıza riski oluşturmadan önceki kritik dönemlerinin öngörülmesine olanak sağlayan veri odaklı yaklaşımların uygulanabilirliğini ortaya koymaktadır [49].

IEEE 14-bus sisteminde oluşturulan senaryolarla hem de gerçek saha verisi olan *VSB Power Line Fault Detection* veri setiyle gerçekleştirilen bu çalışma, elektrik dağıtım sistemlerinde meydana gelen arızaların doğru şekilde tespiti ve yerinin belirlenmesi amacıyla, Transformer modelinin yapısal parametrelerinin otomatik olarak optimize edilmesine odaklanmaktadır. Transformer modeli, doğal dil işleme uygulamalarında, LSTM ve RNN ile kurulmuş Seq2Seq modellerinde karşılaşılan bağlam kaybı problemine etkili bir çözüm sunabilmektedir. Çalışmada, klasik derin öğrenme yaklaşımlarının (özellikle CNN ve RNN tabanlı mimarilerin) yüksek empedanslı arızalar ve geçici anomalilerin tespitinde sınırlı kaldığı, ayrıca hiper parametre ayarlamalarının manuel yapılmasının performans üzerinde olumsuz etkileri olabileceği ifade edilmiştir[50].

Model, arıza türü sınıflandırmasında %99,3 ve arıza lokalizasyonunda %97,6 MCC skorları ile yüksek performans sergilemiş; AUC değerlerinin ise her iki görev için de 0.999 seviyesinde olduğu raporlanmıştır. Ayrıca önerilen model, Bi-LSTM, Bi-GRU-Attention ve CNN-Attention gibi yaygın DL mimarileriyle karşılaştırmalı olarak test edilmiş ve bu modellere kıyasla daha düşük hesaplama maliyeti ve daha kısa çıkarım süresi (~1.44 ms) ile üstünlük sağlamıştır [50].

Finlandiya'nın merkezi bölgelerindeki hava koşullarına bağlı dağıtım şebekesi arızalarını yapay sinir ağları kullanarak tahmin etmek ve farklı zaman pencerelerinin modele etkisinin değerlendirildiği çalışmada, meteorolojik gözlemleri ağ seviyesinde toplayıp sıralı modeller üzerinde deneyler gerçekleştirmiş, 6 saatlik gecikme penceresinin en yüksek öngörüsül başarımı sağladığını; sentetik veri üretimi ile nadir yüksek arıza zirvelerinin tahmin doğruluğunun yükseldiğini göstermişlerdir[51].

Çalışmanın sınırlılıkları arasında yalnızca tek bir bölge ve şebeke seviyesi veri kullanımı, simülasyon dışı gerçek zamanlı uygulama altyapılarının değerlendirilmemesi ve karmaşık RNN/LSTM modellerinin kısa dönem bağımlılık gereksinimlerinde gereksiz kalması sayılmaktadır. Buna karşın, meteorolojik veri ön işleme adımları, uygun zaman penceresi seçiminin önemi, sentetik veri üretimi stratejisinin nadir olay tahminine katkısı literatürdeki bir boşluğu doldurarak pratik yöntem önerileri sunmaktadır [51].

[52], 2015–2022 yılları arasında ABD genelinde 15 dakikalık aralıklarla toplanan EAGLE-I kesinti verilerini kullanarak, güç kesintilerinin uzamsal ve zamansal eğilimlerini ortaya koymayı ve Bayesian Gaussian Karışım Modelleri ile anomali kesinti olaylarını tespit etmeyi amaçlamaktadır. 161,5 milyon kayıt içeren veri setini önce ardışık zaman damgalarındaki tutarlılığa dayalı bir algoritma ile anlamlı kesinti olaylarına dönüştürmüş; ardından tüm kesintileri üç süre kategorisine (≤ 2 sa, 2–8 sa, >8 sa) ayırarak frekans ve müşteri etki analizleri yapmış; küme merkezlerinden üç standart sapma uzaklıkta kalan olayları anomali olarak işaretlenmiştir.

[53] kapsamında yarı denetimli bir anomali tespit çerçevesi sunularak, dağıtım şebekesi kesinti öngörülerinin doğruluk ve dayanıklılığının artırılması hedeflenmiştir; VAE, veri kalitesi yönetimi için uçsal veri anormalliklerini latent uzayda tanımlarken, LSTM katmanı zaman serisi bağımlılıklarını yakalayıp anomali sinyallerini güçlendirmiştir. Modelin yarı denetimli öğrenme aşamasında, veriler birlikte kullanılarak dinamik eşikleme yöntemiyle anomali sınırı belirlenmiş ve çevrimiçi–çevrimdışı döngüsel eğitim protokolü ile gerçek zamanlı adaptasyon sağlanmıştır. Modelin kesinti öngörü performansını eğitimde $R^2=0,770$ 'e ve testte $R^2=0,570$ 'ye yükselttiği, RMSE'de sırasıyla %24 ve %27 oranında iyileşme sağladığı belirtilmiştir [53].

Doğu ABD'nin 17 eyaletinde 15.872 benzersiz fırtına olayı üzerinde elektrik kesintisi tahmini gerçekleştirilen çalışmada; model, resmi altyapı verileri yerine yol uzunluğu, konut ve iş yeri sayısı gibi altyapı proxy'leri ile ağaç örtüsü, eko-bölge dağılımları ve ERA5 meteorolojik verilerini birleştirerek CatBoost gradyan artırma algoritmasıyla hem tüm fırtına türlerini hem de rüzgâr temelli fırtına olaylarını ayrı ayrı regrese etmiştir. Model, kronolojik 10 katmanlı çapraz doğrulama ile test edilerek gerçekçi zaman serisi performans ölçümü sağlamış; genel fırtına türlerinde $r^2=0,61$, rüzgâr fırtınalarında ise $r^2=0,31$ elde edilmiştir. [54].

[16], dağıtım şebekesinde yıldırım kaynaklı kesintileri günlük bazda sıfır, bir veya iki ve üzeri kesinti olmak üzere üç kategoriye ayırarak kısa vadeli müdahaleleri kolaylaştırmayı hedeflemiştir. Coğrafi yakınlık temelli K-Means kümeleme yöntemiyle büyük alanlara gruplandırma yapılması ardından ANOVA testleri kullanılarak kesinti ve fırtına olaylarının karakteristikleri doğrultusunda üç istatistiksel küme oluşturulmuştur. Her küme için binom dağılımı bazlı kesinti olasılık parametresi (p) hesaplanarak, günlük fırtına sayısına bağlı olarak farklı olasılık değerleri elde edilmiştir. Son olarak, dengesiz sınıf dağılımını yönetmek üzere ağırlıklı Lojistik Regresyon modeli ve Optuna optimizasyon aracıyla kullanıcı tanımlı minimum algılama oranları sağlanacak şekilde ağırlıklandırılmıştır.

[55] 'de, ABD'nde yer alan bir elektrik dağıtım şebekesinde rüzgâr, ıslak kar ve buzlanma kaynaklı kesintilerin sınıflandırılması amacıyla üç aşamalı bir çerçeve önerilmiştir. İlk olarak tek aşamalı çoklu sınıflandırma, ardından ikili-çoklu aşamalı sınıflandırma ve son olarak Dempster–Shafer teorisine dayalı toplulaştırma yaklaşımları olmak üzere üç farklı modelleme mimarisi tasarlanmıştır. Her mimaride Random Forest, MLP, SVM, KNN, AdaBoost ve Quadratic Discriminant Analysis (QDA) modelleri beş katlı çapraz doğrulama ve hiper parametre optimizasyonu ile değerlendirilmiştir. Elde edilen makro F_1 skorları karşılaştırmalı olarak sunulmuş; son aşamada kullanılan Dempster–Shafer toplulaştırma yöntemi, farklı sınıflandırıcıların çıktılarından pignistik olasılıklar (belirsizlik içeren olasılık kütlelerini, karar vermeye uygun tekil olasılık değerlerine dönüştüren yöntem) üreterek karar belirsizliğini %5–7 oranında azaltmış ve üç sınıf için ortalama F_1 skorunu 0,76 seviyesine yükseltmiştir. Ancak çalışmanın yalnızca tek bir bölgesel veri setiyle sınırlı olması, sonuçların transformatör düzeyinde genellenebilirliğini kısıtlamaktadır.

Morris ve ark. Scottish Power Energy Networks dağıtım bölgesine ait 1985–2011 yılları arasındaki iletim hattı kesinti verilerini iki saatlik duraksama aralığına göre kümelendirerek, her hava olayı için kesinti sayısı, ortalama kesinti hızı (örneğin rüzgâr olaylarında 18,6 kesinti/saat) ve coğrafi yayılım (360 km çap) gibi nicel özellikler tanımlamışlardır [56].

Song ve ark., güç sistemlerindeki plansız kesinti süresi verilerinin karmaşık Non-Gaussian özelliklerini LMLSM modeliyle başarılı şekilde yakalamaktadır. Bu stokastik model, kesinti süresi tahmininde RMSE'nin 10–17 saat aralığına indirilmesini sağlamış olup, restorasyon ve operasyonel planlama süreçlerine doğrudan katkı sunmaktadır. Çalışma, lag-rolling tabanlı zaman serisi modellerin uzun dönem bağımlılıklarını daha derin biçimde anlamlandırmak ve restorasyon sürelerinin öngörüsünü anomali tahmin modellerinin tamamlayıcısı olarak entegre etmek açısından literatüre katkıda bulunmaktadır. [17].

Shengyi Wang ve ark. tarafından yürütülen çalışmada, hava koşullarına bağlı arıza durumlarının gerçek zamanlı bildirimler doğrultusunda çevrimiçi biçimde tahmin edilmesi amaçlanmıştır. Bu kapsamda, rüzgâr hızı ve yıldırım akımına dayalı olarak tanımlanan ikili Bernoulli problemleri ele alınmış; model parametreleri, bildirilen arıza verileri geldikçe güncellenebilecek şekilde yapılandırılmıştır. Geliştirilen iki çevrimiçi kestirimci mekanizmadan ilki, maksimum olabilirlik (Maximum Likelihood) yaklaşımını esas alırken; diğeri ise maksimum a posteriori (MAP) tahminine dayanmaktadır. MAP yaklaşımı, yalnızca gözlemlenen verilere değil, aynı zamanda öncül bilgiye de dayanarak model parametrelerinin en olası değerini belirlemeye olanak tanımakta ve bu sayede karar verme sürecinde belirsizliğin azaltılmasına katkı sağlamaktadır. Her iki yöntem, yapay olarak üretilen 10.000 örnekten oluşan bir veri seti üzerinde doğruluk ve F_1 -skoru temelinde performans analizine tabi tutulmuştur. Bununla birlikte, çalışmanın sınırlılıkları arasında yalnızca iki meteorolojik değişkene (rüzgâr hızı ve yıldırım şiddeti) odaklanılması ve gerçek dünya verileri yerine sentetik veri kullanılması nedeniyle elde edilen sonuçların genellenebilirliğinin belirsiz kalması yer almaktadır[57].

İklim verileri, elektrik talep profilleri ve geçmiş arıza kayıtlarını kullanarak tek tek müşteri devresine yönelik elektrik kesintisi olasılığını tahmin etmeyi hedeflemiş bu çalışmada; yedi özellikten oluşan (yük, sıcaklık, rüzgâr hızı, rüzgâr yönü, nem, basınç ve hava açıklaması) girdi setiyle iki LSTM tabanlı model geliştirilmiştir. Birinci modelde yalnızca hava ve yük verileriyle saatlik talep tahmini yapılarak kesintiler, gerçek ve tahmin edilen yük farkı ve belirli hava koşulları eşiklerine göre tespit edilirken; ikinci modelde buna ek olarak geçmiş kesinti verisi de öz nitelik olarak dahil edilmiş ve dengesiz kesinti örneklerini dengelemek için SMOTE uygulanmıştır. Her iki model farklı görünüm (1 saat/24 saat look-back) ve katman konfigürasyonlarıyla eğitilip test edilmiş; model 1-3 için RMSE=139.9 ve %99.559 güvenilirlik, model 2-1 için ise RMSE=0.003 ve %99.14 güvenilirlik değerleri elde edilmiştir [58].

Mankov ve ark., hava koşulları ve geçmiş acil arıza verilerini kullanarak iletim hattı kesintilerinin sayısal tahminini amaçlamış; bu kapsamda istatistiksel DL modelleri ile ML yaklaşımları karşılaştırmalı olarak değerlendirilmiş ve model parametreleri GridSearchCV ile optimize edilmiştir. 2020 yılı St. Petersburg–Leningrad hattı arızalarını içeren çalışma için yazarlar; artan eğitim verisi hacminin tahmin doğruluğunu önemli ölçüde iyileştireceğine işaret ederek büyük veri entegrasyonu ve bölgesel iklim karakteristiklerinin modele dahil edilmesinin gerekliliğine vurgu yapmışlardır [59].

Olağanüstü hava koşullarının incelendiği başka bir çalışmada; kasırga, aşırı yağış ve buzlanma gibi ekstrem hava koşullarının dağıtım şebekesi ekipmanları üzerindeki etkisini dikkate alarak, öncelikle ekipman arızalanma olasılığı ile rüzgâr ve buz yüklenme basıncı arasındaki korelasyonu geliştirilmiş bir yük basıncı modeli ve hava tahmini verileri kullanılmıştır. Ardından iletim hattı, direk ve trafo gibi cihazlar için ayrı ayrı arıza olasılık modelleri oluşturup, Monte Carlo temelli algoritmayla mevcut işletme koşullarında en az yük kesintisi hedefiyle optimal kesinti öncesi uyarı analitiği gerçekleştirmektedir. Model, IEEE 33-baralı bir şebeke üzerinde simülasyonla test edilmiş; hatların hortum hareket yönüne dik konumlarda daha yüksek arıza riski taşıdığı, bağlayıcı anahtarlarla yük aktarımının risk haritasında fark yarattığı gibi şebeke topolojisi etkilerini de ortaya koymuştur [60].

Kasırğa kaynaklı kesintilerin öngörüsünde uygulanan ML tekniklerinin karşılaştırmalı bir değerlendirmesini sunmak ve dağıtım şebekesi direncini artırmak için izlenebilecek model seçme adımlarını barındıran derleme çalışmasında; literatürde kullanılan coğrafi, şebeke altyapısı ve meteorolojik veri türlerini toplayıp ön işleme tâbi tutarak; karar ağaçlarından DL' ye kadar on bir sınıflandırma algoritmasını ve bunların performans değerlendirme metrikleri incelenmiştir. Çalışmada; veri eksikliği, kalite sorunları, karmaşık modellerin yorumlanabilirlik güçlükleri ve gerçek zamanlı hesaplama yükü gibi sınırlılıkları da ele alarak ileride sentetik veri, uydu ve sosyal medya kaynakları kullanımı ile hesaplama verimliliğinin geliştirilmesi gereğine işaret etmektedir [36].

Tayfun felaketleri sırasında meydana gelen elektrik kesintilerinin sürelerini tahmin etmek amacıyla, çoklu algoritma temelli yığınlanmış topluluk öğrenmesi yaklaşımını kullanarak yeni bir tahmin modeli geliştirmeyi hedefleyen araştırmacılar, meteorolojik veriler ile coğrafi ve elektrik sistemi verilerini birleştirerek kapsamlı ve yüksek çözünürlüklü veri setleri oluşturmuşlardır. Karşılaşılan dengesiz veri seti dağılım problemini çözmek amacıyla CNN ile alt-örnekleme ve SMOTE ile üst-örnekleme yöntemlerini birlikte uygulamışlardır. Bu çalışmada temel öğrenciler olarak Extra Trees, XGBoost, LGBM, Random Forest, Gradient Boosting kullanılmıştır. Meta öğrenci olarak ise Gradient Boosting algoritması tekrardan kullanılmıştır. Önerilen modelin performansı Guangdong Eyaleti'nde yaşanan "Chapaka" tayfunu sırasında test edilerek doğrulanmış ve tahmin doğruluğunun, tek algoritmalarla karşılaştırıldığında önemli ölçüde daha iyi olduğu gözlemlenmiştir. Çalışmanın temel sınırlılığı, kullanılan yöntemlerin yalnızca spesifik afet türlerine yönelik olmasıdır. Buna rağmen, modelin özellikle kesinti sürelerinin yüksek çözünürlükte tahmini ve veri dengesizliği problemlerini etkili bir biçimde çözmesi açısından literatürdeki önemli bir boşluğu doldurduğu görülmüştür [61].

Benzer bir çalışmada Hou ve ark. tarafından gerçekleştirilmiştir. Tayfun gibi aşırı hava olayları altında meydana gelen kullanıcı kaynaklı elektrik kesintilerinin önceden tahminine yönelik olarak, çok algoritmali bir *Stacking* entegrasyon modeli geliştirmektedir. Model, Random Forest, AdaBoost, Extra Trees, SVM, Decision Tree ve Lojistik Regresyon gibi temel öğrenicilerle oluşturulmuş ilk katmanın çıktılarını, meta öğrenici olarak seçilen XGBoost algoritmasına besleyerek yüksek düzeyde sınıflandırma doğruluğu elde etmeyi amaçlamaktadır. Çin'in Guangdong eyaletine bağlı ve tayfunların sıklıkla karaya ulaştığı Xuwen bölgesi örnek alan olarak seçilmiş; 2014 ve 2015 yıllarında yaşanan Rammasun, Kalmaegi ve Mujigae tayfunlarına ilişkin gerçek kesinti verileriyle model test edilmiştir. 1 km²'lik hücresel yapıda oluşturulan veri seti; coğrafi veriler, güç şebekesi verileri ve meteorolojik değişkenler gibi birçok kategoriden oluşmaktadır. Model, senaryo bazlı testlerde %90,59 recall ve %76,78 doğruluk değerlerine ulaşarak, özellikle afet sonrası müdahale planlamasında önem taşıyan “gerçekleşen kesintileri yakalama” kapasitesini artırdığını göstermiştir [62].

Bir başka doğal afetler sonucu meydana gelen elektrik kesintilerini tahmin etmek amacıyla yapılan çalışmada, sentetik elektrik dağıtım sistemleri oluşturmayı ve bu sistemler üzerinden bina seviyesinde elektrik kesinti olasılıklarını belirlenmesi hedeflemiştir. Araştırmacılar, ABD'de gerçek güç dağıtım şebekelerinin güvenlik nedeniyle erişime kapalı olması sorununu aşmak için, kamuya açık verilerden sentetik şebeke tasarımları geliştirmiş ve kırılabilirlik eğrileri kullanarak afet yüklemeleri altındaki arıza senaryolarını simüle etmişlerdir. Çalışmada, müşteri konumlarının ve trafoların belirlenmesi, yol güzergahlarının ve direk konumlarının optimizasyonu gibi aşamalar içeren sentetik ağ oluşturma yöntemi kullanılmıştır. Bu yöntem, gerçek sistemlerle karşılaştırılarak doğrulanmış ve Ohio eyaletinde meydana gelen 2012 Derecho ve Teksas'taki 2017 Harvey Kasırgası gibi büyük afet senaryolarında test edilmiştir. Çalışmanın temel sınırlılıkları arasında, rüzgâr hızı gibi afet yüklemeleri için detaylı mekânsal verilerin eksikliği ve ağaç kaynaklı kesintilerin model dışı bırakılması yer almaktadır. Başarı metriği olarak, gerçek afet senaryolarındaki elektrik kesintilerinin dağılımıyla sentetik modelin simülasyon sonuçları karşılaştırılmıştır. Önemli bulgular arasında, farklı direk sınıfları ve yaşlarının kesinti tahminlerinde belirgin farklılıklar yaratması öne çıkmıştır [37].

Tianqiao Zhao ve ark. çalışması, iklim değişikliği nedeniyle artan sıklıkta ve şiddette yaşanan ekstrem hava olaylarının neden olduğu elektrik kesintilerini tahmin etmeye yönelik tahmin modeli geliştirmeyi amaçlamıştır. Çalışmada, hava durumu verilerinin zaman serisi yapısını etkin kullanmak ve sınırlı sayıdaki kritik hava olayları nedeniyle oluşan veri dengesizliğini gidermek için LSTM tabanlı VAE ve sliding-window tekniği birleştirilmiş yenilikçi bir yöntem önerilmiştir. Bu yöntem, az rastlanan ekstrem olaylara ait verileri sentetik olarak çoğaltarak veri setinin dengelenmesini sağlamaktadır. Geliştirilen model, Kuzey Amerika Mesoscale ve Global Forecast System hava durumu tahminlerini kullanmıştır. Çalışmanın temel sınırlılığı, ekstrem olayların seyrekliği nedeniyle yeterince temsil edici veri setleri oluşturmadaki zorluk olarak ifade edilmiştir. Başarı ölçütü olarak ortalama karesel hata (MSE) skoru kullanılmıştır. Araştırmanın dikkat çekici bulguları arasında, farklı hava olayları arasındaki dengesiz verilerin sentetik olarak dengelenmesi yoluyla tahmin doğruluğunun belirgin bir şekilde artırılabilirdiği vurgulanmaktadır [63].

Güney Çin'de yapılan bir çalışmada, 10 kV düzeyindeki kentsel dağıtım sistemlerinde yağış kaynaklı arıza tahminlerini geliştirmek amacıyla, mekânsal ve zamansal verilerin entegrasyonuna dayalı çoklu hedefli bir ML modeli önermektedir. Araştırmada, yüzey su baskınlarının etkisini modellemek için geliştirilen yağmur-su birikintisi modeli (waterlogging model), ayrıntılı topografik yapı ile ERCD-TIN (Element Reconstruction Constrained Delaunay Triangulated Irregular Network) temelli alt-ağlara uygulanmıştır. Meteorolojik veriler, coğrafi bilgiler, altyapı bileşenleri ve su baskını verileri superposition feature yaklaşımıyla birleştirilmiş ve tahminler Multi-Target Random Forest (SF-MTRF) algoritması kullanılarak gerçekleştirilmiştir. Modelin, özellikle hem yeraltı hem de havai ekipmanları kapsayan "Woe" senaryosunda %94,6'lık Recall oranıyla yüksek doğruluk sağladığı belirtilmiştir. Ayrıca RF, SVR, AdaBoost ve KNN gibi diğer yöntemlerle karşılaştırıldığında hem doğruluk hem de hesaplama süresi bakımından üstün performans sergilediği gözlemlenmiştir. Bu çalışma, yalnızca meteorolojik parametrelerle sınırlı kalan geleneksel yaklaşımlardan farklı olarak, coğrafi ve altyapısal faktörlerin mekânsal-zamansal entegrasyonu ile çok boyutlu tahmin imkânı sağlayarak erken uyarı sistemlerine önemli katkılar sunmaktadır [64].

Richard J. Campbell tarafından hazırlanan çalışma, aşırı hava olaylarının elektrik iletim ve dağıtım sistemleri üzerindeki etkilerini analiz ederek, bu olayların uzun süreli kesintilere ve ciddi ekonomik kayıplara yol açtığını ortaya koymaktadır. ABD genelinde toplanan kesinti verileri üzerinden yapılan agresyon analizleri ve Avrupa ülkeleriyle yapılan SAIDI/SAIFI kıyaslamaları, hava koşullarına bağlı kesintilerin artan bir eğilim gösterdiğini ve mevcut altyapının bu olaylara karşı yeterince dayanıklı olmadığını göstermiştir. Çalışma, hava kaynaklı kesintilerin ABD ekonomisine yıllık 20 ila 55 milyar dolar arası zarar verdiğini belirtmiş; bu nedenle sistem dayanıklılığının artırılması adına ağaç budama takvimlerinin iyileştirilmesi, iletim hatlarının yer altına alınması, akıllı şebeke sistemlerinin yaygınlaştırılması ve dağıtık üretim kaynaklarının entegrasyonu gibi önlemleri önermiştir. Ayrıca, veri eksikliği ve raporlama standartlarının olmaması önemli bir sınırlılık olarak vurgulanmış; özellikle 69 kV altı dağıtım sistemlerinde standardize veri toplanmamasının politika üretimini zorlaştırdığı belirtilmiştir [6].

Literatürde hava koşullarına bağlı elektrik kesintilerinin tahmin edilmesine yönelik birçok çalışma yer almakla birlikte, özellikle yapay zekâ temelli yöntemlerin konvansiyonel yöntemlerle karşılaştırıldığı uygulamalı çalışmalar dikkat çekmektedir. Bu bağlamda Onaolapo ve ark. tarafından gerçekleştirilen çalışmada, Güney Afrika'nın Pietermaritzburg bölgesine ait 2014–2017 yıllarını kapsayan gerçek saha verileri kullanılarak, hava olaylarına bağlı olarak meydana gelen elektrik kesintileri tahmin edilmeye çalışılmıştır. Çalışmada ANN ve üstel düzeltme (ES) modelleri karşılaştırmalı olarak değerlendirilmiştir. Mevsimsel bazda sistem arızası sayısını tahmin etmeyi amaçlayan modelleme sürecinde, Model 1'de tüm meteorolojik değişkenler, Model 2'de ise korelasyon analizine dayalı olarak belirlenen alt değişken seti kullanılmıştır. ANN Model 1, test verisinde $R^2 = 0.9999$, $RMSE = 0.0035$ ve $MAPE = \%0.0001$ gibi istisnai performans sonuçları ile öne çıkmıştır. Araştırma sonucunda ANN modellerinin, hava koşullarına bağlı arızaların öngörülmesinde karmaşık ilişkileri daha etkin biçimde modelleyebildiği ve bu nedenle konvansiyonel modellere kıyasla daha güvenilir sonuçlar sunduğu ifade edilmiştir [65].

Hines ve ark. tarafından gerçekleştirilen çalışmada, ABD'nin doğu kıyısındaki topluluklarda elektrik kesintilerinin tahmin edilmesi amacıyla geliştirilmiş ilk sistematik modelleme girişimlerinden birini sunmaktadır. Araştırma kapsamında, 2008 yılında yaşanan buz fırtınası olayına bağlı olarak 10 eyalette toplanan 3.000'den fazla fırtına ve kesinti verisi bulunmaktadır. Çalışmada, meteorolojik veriler, elektrik dağıtım sisteminin yapısal özellikleri ve nüfus yoğunluğu gibi değişkenler birleştirilerek kesinti olasılığına dair tahmin modelleri oluşturulmuştur. Bu modellerde özellikle regresyon tabanlı analizler ve sınıflandırıcı algoritmalar uygulanmış, bölgesel olarak farklı risk profillerine sahip alanlar haritalandırılarak erken uyarı sistemleri için kullanılabilir veri görselleştirme çıktıları üretilmiştir. Modelleme sürecinde, fırtına şiddeti arttıkça nüfus başına düşen kesinti oranının doğrusal olmaktan çıkıp üssel bir yapıya büründüğü, bu nedenle geleneksel doğrusal modellerin ekstrem hava olaylarında yetersiz kalabileceği vurgulanmıştır [66].

Türkiye'de daha önceden yapılan yüksek öğretim tez çalışmaları kapsamında kestirimci bakım ve çevresinde yer alan bazı çalışmalar bulunmaktadır. Bu çalışmalara örnek olarak, Çankırı'ya ait aylık elektrik kesinti sayılarının panel veri yapısı içinde uygun uzunlamasına modellerle analiz edilmesini ve özellikle verinin %83,4'ünün sıfır gözlemlerden oluşması nedeniyle Bayesian GLMM çerçevesinde MCMCglmm paketi kullanılarak Poisson, sıfır-aşırı (ZIP) ve "hurdle" Poisson modellerinin uygulanması hedeflenmiştir. Çalışmanın başlıca sınırlılıkları arasında tek bir bölge verisi kullanılması, yüksek orandaki sıfır gözlemler nedeniyle ZIP modelinin yakınsama sorunları yaşaması ve MCMCglmm paketinin Negatif Binom dağılım desteğinin bulunmaması yer almaktadır. Literatürde, dağıtım şebekesi kesinti verilerinin uzunlamasına panel veri ve Bayesian GLMM yöntemleriyle modellenmesine dair sayılı örnek bulunmakta olup, bu çalışma özellikle yoğun sıfır içerikli uzunlamasına sayım verisinin Poisson temelli bir Bayesian yaklaşım ile başarılı şekilde analiz edilebileceğini göstermesi bakımından literatürdeki boşluğu doldurmaktadır; bununla birlikte, genellenebilirliği artırmak için çok bölgesel veya hiyerarşik Bayes modelleri kullanılması önerilmektedir [67]. İlgili çalışma kesinti sayılarının tahmini, istatistiksel yaklaşım ve uzunlamasına analizi içermekte olup bu tez kapsamında yapılan çalışmaya istinaden farklılıklar göstermektedir.

[68], Türkiye’de kestirimci bakım hakkında yapılan yüksek lisans tez çalışmasıdır. Dicle Elektrik Dağıtım A.Ş. bölgesinde günlük bazda kaydedilen elektrik kesinti sayıları ile meteorolojik verileri kullanarak bölgesel kesinti yoğunluğu ve sayısının önceden tahmin edilmesini amaçlamış; bu doğrultuda 1D CNN, Simple-RNN ve LSTM modelleri önerilmiştir. Model geliştirme aşamasında, girdiler 1, 5 ve 7 özellikli üç farklı veri seti halinde düzenlenmiş; öğrenme sürecinde Adam ve RMSProp optimizasyon algoritmaları kullanılarak her modelin MAE ve MSE üzerinden performans karşılaştırmaları yapılmıştır. Çalışmanın sınırlılıkları arasında yalnızca tek bir bölge verisinin (Diyarbakır) incelenmesi, modellerde gerçek zamanlı güncelleme ve daha zengin altyapı metadatasının kullanılmaması ve dengesiz veri sorununa yönelik özel yöntemler uygulanmaması yer almaktadır.

Yazar, literatürde yoğun sıfır ve düşük frekanslı kesinti sayılarının DL çerçevesinde modellenmesine dair sayılı örnek bulunduğunu belirterek, bir boyutlu CNN ve RNN/LSTM uygulamalarının bu boşluğu doldurduğunu vurgulamıştır. Başarı metrikleri açısından, en iyi sonuçlar 7 özellikli veri seti üzerinde LSTM+RMSProp modelinde elde edilmiş olup, bu kombinasyon $MAE = 0.0202$ ve $MSE = 0.0008$ değerleriyle en düşük hata oranını sergilemiştir. Performans rakamlarının ötesinde, yazarlar ek özelliklerin (planlı/plansız ayrımı ve meteorolojik değişkenler) modele sağladığı katkıyı ve LSTM’in süreklilik içeren zaman serilerini öğrenmedeki üstünlüğünü özellikle vurgulamış; ayrıca çok bölgeli uygulamalar ve ek veri kaynaklarının entegrasyonunun model genellenebilirliğini artıracığına işaret etmiştir [68].

Yürek'in [69] çalışması, öncelikle üç farklı endüstriyel veri seti (C-MAPSS turbo fan motor verisi, AI4I 2020 freze makineleri arıza verisi ve gerçek dünya hidrolik sistem verisi) üzerinde kestirimci bakım amaçlı ML paradigmasını bütünleşik bir biçimde sunmayı; yani makinenin kalan kullanım ömrünü tahmin etme, arıza durumunu sınıflandırma ve hata-arıza bileşen ilişkilerini keşfetme hedeflemiştir. Çalışmada önerilen OPMEB yöntemi, ordinal ikili ayrıştırma teknikleri (OneVsAll, OneVsFollowers, OneVsNext) ile karşılaştırılmış ve doğruluk, kesinlik, duyarlılık ve F-1 metrikleri üzerinden 10 katlı çapraz doğrulama ile değerlendirilmiştir. Elde edilen sonuçlar, OPMEB'in ortalama doğruluk oranını %79,80'den %84,78'e yükselterek mevcut nominal yaklaşımlara kıyasla belirgin bir iyileşme sağladığını göstermiştir. Bununla birlikte, çalışmanın farklılıkları arasında yalnızca havacılık, imalat ve hidrolik sistem verileriyle sınırlanmış olması; transformatörler veya dağıtım şebekeleri gibi farklı altyapılarda genellenebilirliğin sınırlandırılması ve gerçek zamanlı çevrimiçi uygulamaların test edilmemesi yer almaktadır.

İncelenen çalışmalar kapsamında, endüstriyel süreçlerde ve akıllı şebekelerde "Kestirimci Bakım" uygulamalarının kullanımlarına ait örnekler gösterilmiştir. Ayrıca; kestirimci bakım çalışmaları kapsamında kullanılan ML ve zaman serisi modelleri ile çağrı merkezi, sosyo-ekonomik durumlar ve enerji politikaları ile ilgili çalışmaların da yapıldığı gösterilmiştir. Akıllı şebekelerde kestirimci bakım yaklaşımında hava koşullarına bağlı arızaların öngörülmesinde veri dengesizliği, mekânsal detay yetersizliği ve algoritma seçiminde model karmaşıklığı gibi ortak sınırlılıklarla karşılaştıklarını göstermektedir. Literatürde, sentetik veri üretimi, çevrimiçi öğrenme, stacking gibi ileri düzey yaklaşımların giderek yaygınlaştığı, ancak bölgesel altyapı ve gerçek zamanlı saha verileriyle doğrulamanın halen sınırlı kaldığı görülmektedir.

Bu kapsamda, bu tez çalışması; farklı veri kaynaklarının bütünleştirilerek, dengesiz örnek problemini çözmeye yönelik sentetik veri kullanımı, sınıf ağırlıklandırma durumları ve veri Undersampling yöntemleriyle desteklenmiş, ML ve DL tabanlı modellerin, birden fazla başarı parametresi ile karşılaştırılarak performans karşılaştırması aktarılmaktadır.

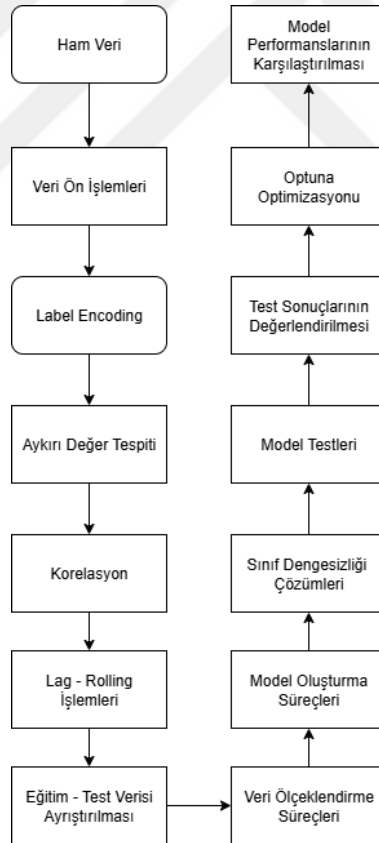
Bu tez çalışması, söz konusu ihtiyaca yönelik İzmir ve Manisa illerinde yer alan 34 adet güç transformatörünün 2021–2024 yıllarına ait kesinti verileri ile hava durumu ve transformatöre ait teknik meta veriler kullanılarak, ML ve DL tabanlı anomali tahminleme modelleri geliştirmeyi ve bu modellerin performanslarını karşılaştırmalı olarak değerlendirmeyi amaçlamaktadır. Çalışmanın nihai hedefi, Türkiye enerji dağıtım sektörü için uygulanabilir, ölçeklenebilir ve yorumlanabilir bir kestirimci bakım sistemi önerisi sunmaktır. Ayrıca veri setinde sınıfların dengesiz olması (%83 oranında kesinti olmayan ve %17 oranında kesinti yaşanan kayıtlar) nedeniyle klasik performans metrikleri (doğruluk gibi) yerine G-Mean gibi dengesiz sınıf problemlerine duyarlı metrikler kullanılmıştır. Bu yaklaşım ile çalışma hem teorik hem metodolojik olarak daha sağlam ve kullanılabilir yapıda oluşmaktadır.

Bu tez çalışmasının bir diğer önemli özelliği, yalnızca akademik bir model önerisi sunmakla kalmayıp, bu önerinin gerçek saha verisiyle test edilerek sonuçlarının değerlendirilmiş olmasıdır. Sıklıkla literatürde yer bulan sentetik veri ile eğitilen modeller, saha uygulamasında yetersiz kalmaktadır. Oysa bu çalışma, saha verileri üzerinden geliştirilmiş, modellenmiş ve yorumlanmış çıktılar sunmaktadır. Bu yönüyle, yalnızca akademik olarak değil; enerji sektöründe faaliyet gösteren ARGE birimleri, proje mühendisleri ve karar vericiler için de doğrudan uygulanabilir çıktılar üretmektedir. Ayrıca bu tez, literatürde ilk kez İzmir ve Manisa bölgesine odaklanan, bölgesel arıza tahminleme açısından bir örnek teşkil etmektedir. Bu tür bölgesel modellemeler, ileride Türkiye geneline yayılacak büyük ölçekli enerji izleme ve kestirimci bakım platformlarının temelini oluşturabileceği öngörülmektedir.

Sonuç olarak, bu tez çalışması; güç trafolarında plansız kesintilerin azaltılmasına yönelik olarak özgün bir yaklaşım sunmakta, çok katmanlı bir veri analizi ile ML ve DL algoritmalarını bütünleştirmekte ve enerji sektöründe kestirimci bakım uygulamaları için temel oluşturması hedeflenmektedir. Bu yönüyle çalışma, akademik literatüre yüksek düzeyde katkı sunmakta ve saha uygulamaları için örnek teşkil edecek düzeyde değer üretmektedir.

3. MATERYAL VE YÖNTEMLER

Bu bölümde, hava koşullarına ve meta verilere bağlı güç trafolarında oluşan kesintilerin kestirimine yönelik geliştirilen modellerin tasarımı, kullanılan veri kümesinin özellikleri, uygulanan ön işleme adımları, öznelik mühendisliği süreçleri, veri dengesizliğinin giderilmesi ve seçilen ML ve DL modellerinin detayları, karşılaştırılması sistematik biçimde sunulmaktadır. Modelleme süreci, gerçek saha verilerinden elde edilen zaman serisi yapısındaki hava, altyapı ve arıza verilerinin analizine dayanmakta olup; farklı algoritmalar karşılaştırmalı olarak değerlendirilmiştir. Ayrıca, modelin başarımını ölçmek üzere kullanılan metrikler ile doğrulama stratejileri açıklanmış ve hiper parametre optimizasyonu yöntemlerine başvurulmuştur. Bu kapsamda izlenen adımlar Şekil 2’de özetlenmekte olup, aşağıda her bir bileşene ilişkin ayrıntılı bilgi sunulmaktadır.



Şekil 2 Veri İşlem Süreçleri

3.1 Veri Hazırlığı Ve Ön İşlem Süreçleri

3.1.1 Verilerin Temini Ve Kullanılan Özellikler

İzmir ve Manisa şehirlerinde bulunan 34 farklı güç transformatörüne ait 01.01.2021 ile 31.12.2024 yılları arasındaki günlük bazda kesinti durumu ile ilgili veri seti temin edilmiştir. Modelleme çalışmaları kapsamında kullanılan özellikler şu şekildedir:

(1) Category

Temin edilen güç transformatörlerinin kategorik olarak numaralandırmasını içermektedir. 0-33 arasında numaralandırma ile 34 transformatör incelemeye alınmıştır.

(2) X_KOORDINAT_BOYLAM

Her bir transformatör için boylam koordinat bilgilerinin yer aldığı özelliktir.

(3) Y_KOORDINAT_ENLEM

Her bir transformatör için enlem koordinat bilgilerinin yer aldığı özelliktir.

(4) Time

Tarih bilgisinin, GG.AA.YYYY formatında tutulan zaman özelliğidir.

(5) day-week

Haftanın hangi gününe denk geldiğinin gösterildiği özelliktir. (1=Pazartesi, 7=Pazar)

(6) **weather_code**

WMO (World Meteorological Organization) tarafından tanımlanan belirli bir gündeki en şiddetli hava durumu koşuludur. 00-99 arasında 100 farklı kod bulunmasına rağmen, veri setinde yer alan kodlar Tablo-1 de gösterilmiştir.

Tablo 1 World Meteorological Organization – Hava Durumu Koşulları

Kod	Tanım
00	No significant weather (hiçbir önemli hava olayı gözlenmedi)
01	Light drizzle (hafif çişe)
02	Moderate drizzle (orta şiddet çişe)
03	Heavy drizzle (şiddetli çişe)
51	Drizzle, not freezing, continuous slight at time of observation (kesintisiz, hafif çişe)
53	Drizzle, not freezing, continuous moderate at time of observation (kesintisiz, orta çişe)
55	Drizzle, not freezing, continuous heavy at time of observation (kesintisiz, şiddetli çişe)
61	Rain, not freezing, continuous slight at time of observation (kesintisiz, hafif yağmur)
63	Rain, not freezing, continuous moderate at time of observation (kesintisiz, orta şiddet yağmur)
65	Rain, not freezing, continuous heavy at time of observation (kesintisiz, şiddetli yağmur)
71	Continuous fall of snowflakes, slight at time of observation (kesintisiz, hafif kar)
73	Continuous fall of snowflakes, moderate at time of observation (kesintisiz, orta şiddet kar)
75	Continuous fall of snowflakes, heavy at time of observation (kesintisiz, şiddetli kar)

(7) temperature_2m_max

Gün içerisinde yerden 2 metre yükseklikteki maksimum sıcaklığı tanımlamaktadır.

(8) temperature_2m_min

Gün içerisinde yerden 2 metre yükseklikteki minimum sıcaklığı tanımlamaktadır.

(9) temperature_2m_avg

Gün içerisinde yerden 2 metre yükseklik gerçekleşen en yüksek sıcaklık ile en düşük sıcaklığın farkı olarak tanımlanmaktadır.

(10) apparent_temperature_max

Gün içerisinde görünen maksimum sıcaklık olarak tanımlanmaktadır.

(11) apparent_temperature_min

Gün içerisinde olarak görünen minimum sıcaklık olarak tanımlanmaktadır.

(12) apparent_temperature_avg

Gün içerisinde görünen en yüksek sıcaklık ile en düşük sıcaklık arasındaki fark olarak tanımlanmaktadır.

(13) Sunrise

İlgili tarihte güneşin doğuş zamanıdır.

(14) Sunset

İlgili tarihte güneşin batış zamanıdır.

(15) sunset-sunrise(minute)

İlgili tarihte, ilgili konumun dakika olarak güneş alma süresidir.

(16) daylight_duration

İlgili tarihte, ilgili konumun saniye cinsinden güneş alma süresini temsil etmektedir.

(17) sunshine_duration

İlgili tarihte ve ilgili konumda gökyüzünün bulutsuz veya güneş ışınlarının yere ulaşabildiği toplam süreyi ifade etmektedir.

(18) precipitation_sum

İlgili tarihte, ilgili konuma düşen tüm yağış miktarını (hem yağmur hem de kar dahil) milimetre (mm) cinsinden ifade etmektedir.

(19) rain_sum

İlgili tarihte ve ilgili konumda sadece sıvı formda düşen yağmurun toplam miktarını mm cinsinden ölçülmesidir. Kar, dolu veya diğer katı yağış formları bu değere dahil edilmemektedir.

(20) snowfall_sum

İlgili tarihte ve ilgili konumda yere düşen karın toplam kalınlığını veya eşdeğer su derinliğini mm veya cm cinsinden gösterilmesidir.

(21) wind_speed_10m_max

Yer seviyesinin yaklaşık 10 metre yukarısındaki en yüksek ortalama rüzgâr hızını metrekare başına saniye (m/s) cinsinden gösterilmesidir.

(22) wind_gusts_10m_max

10 m yükseklikte kaydedilen ve çok kısa süreli ani rüzgâr hız patlamalarının en yüksekini metrekare başına saniye (m/s) cinsinden ifade eder. Türbülans, fırtına değerlendirmeleri, ağaç devrilme risk analizi ve kritik altyapı güvenliği açısından veri setine dahil edilmiştir.

(23) wind_direction_10m_dominant

10 metre yükseklikteki en baskın rüzgâr yönünü derecelerle (0° kuzey, 90° doğu, 180° güney, 270° batı) belirtmektedir.

(24) et0_fao_evapotranspiration

FAO (Food and Agriculture Organization) tarafından önerilen Penman-Monteith yöntemiyle hesaplanan referans bitki örtüsünün potansiyel evapotranspirasyon miktarını mm cinsinden verir. Bitkilerin su tüketimini ve buharlaşma kayıplarını tahmin etmek için standart bir yöntemdir. Sulama planlaması ve su yönetimi kararlarında kullanılmaktadır.

(25) KADEME_SAYISI

İlgili transformatöre ait kademe sayısını temsil etmektedir.

(26) HAT_TOPLAM_UZUNLUK_KM

İlgili transformatöre ait hatların kilometre cinsinden toplam uzunluğunu temsil etmektedir.

(27) YG_DIREK_SAYISI

İlgili transformatörlere ait toplam direk sayısını temsil etmektedir.

(28) MUSTEREK_DIREK_SAYISI

İlgili transformatörlere ait toplam müşterek direk sayısını temsil etmektedir.

(29) Kesinti

İlgili transformatöre ait ilgili tarihte kesinti olup olmadığına dair bilgi içermektedir.

(30) Year

İlgili özelliklerin hangi yıla ait olduğunu temsil etmektedir.

(31) İL_label

İlgili transformatörlerin hangi şehirde bulunduğunu belirtmektedir.

(32) İLÇE_label

İlgili transformatörlerin hangi ilçe bulunduğunu belirtmektedir.

(33) Marka_label

İlgili transformatörlerin markalarının yer aldığı özelliktir.

(34) FAN_DURUMU_label

İlgili transformatörün fan bulundurma durumunun belirtildiği sütundur.

(35) SOGUTMA_TURU_label

İlgili transformatörlerde bulunan soğutma türü hakkında bilgi içeren özelliktir.

(36) TIPI_label

İlgili transformatörlerin tipinin belirtildiği sütundur.

(37) Gücü_label

İlgili transformatörün gücünün belirtildiği alandır.

(38) is_outlier

İlgili tarihte ve ilgili güç transformatöründe aykırı değer olup olmadığını belirten özelliktir.

(39) apparent_temperature_min_lag1

İlgili transformatörün bir önceki güne ait görünen minimum sıcaklığı belirtmek için kullanılır.

(40) apparent_temperature_min_rolling7

İlgili transformatörün son yedi güne ait ölçülen minimum sıcaklık değerlerinin ortalamasıdır. (2021 yılı öncesi kullanılmadığı için ilk yedi güne ait değerler medyan ile doldurulmuştur.)

(41) temperature_2m_min_lag1

İlgili transformatörün Bir önceki güne ait 2 metre yüksekliğindeki minimum sıcaklığı göstermektedir.

(42) temperature_2m_min_rolling7

İlgili transformatörün son yedi güne ait yerden 2 metre yükseklikte minimum sıcaklığın ortalamasıdır.

(43) apparent_temperature_max_lag1

İlgili transformatörün bir önceki güne ait görünen maksimum sıcaklıktır.

(44) apparent_temperature_max_rolling7

İlgili transformatörün son yedi güne ait görünen maksimum sıcaklığın ortalamasıdır. (2021 yılı öncesi verileri kullanılmadığı için 2021 yılına ait ilk yedi güne ait değerler medyan ile doldurulmuştur.)

(45) rain_sum_lag1

İlgili transformatöre ait bir gün öncesinin “rain_sum” değeri gösterir.

(46) rain_sum_rolling7

İlgili transformatöre ait son yedi gün içerisindeki “rain_sum” değerlerinin ortalamasıdır. (2021 yılı öncesine ait veriler kullanılmadığı için 2021 yılının ilk haftasındaki değerler medyan ile doldurulmuştur.)

(47) temperature_2m_max_lag1

İlgili transformatöre ait bir önceki günün “temperature_2m” değerini göstermektedir.

(48) temperature_2m_max_rolling7

İlgili transformatöre ait bir önceki yedi günün “temperature_2m” değerinin ortalamasını göstermektedir. (2021 yılı öncesi veriler bulunmadığı için 2021 yılının ilk haftası medyan ile doldurulmuştur.)

(49) Kesinti_lag1

İlgili transformatöre ait bir önceki gün kesinti olup olmadığını belirtir.

(50) Kesinti_rolling7

İlgili transformatöre ait son yedi gün içerisinde oluşan kesinti ortalaması değerini gösterir. (2021 yılı öncesi veriler bulunmadığı için 2021 yılının ilk haftası medyan ile doldurulmuştur.)

(51) Day

İlgili aya ait kaçınıcı gün olduğunu göstermektedir. (0-31 arası)

(52) Month

İlgili satırdaki değerlerin hangi aya ait olduğunu göstermektedir.

(53) Week

İlgili satırdaki değerlerin yılın hangi haftasına ait olduğunu göstermektedir.

(54) Dayofweek

İlgili satırdaki değerlerin haftanın hangi gününe ait olduğunu göstermektedir.

(55) Season

Hangi mevsime ait olduğunu göstermektedir.

(56) dayofweek_sin

“Dayofweek” değerinin sinüzoidal olarak döngüsünü göstermektedir.

(57) dayofweek_cos

“Dayofweek” değerinin kosinüs üzerinden döngüsünü göstermektedir.

(58) month_sin

“month” değerleri arasındaki sinüzoidal olarak döngüsünü göstermektedir.

(59) month_cos

“month” deęerleri arasındaki kosinüs üzerinden döngüsünü göstermektedir.

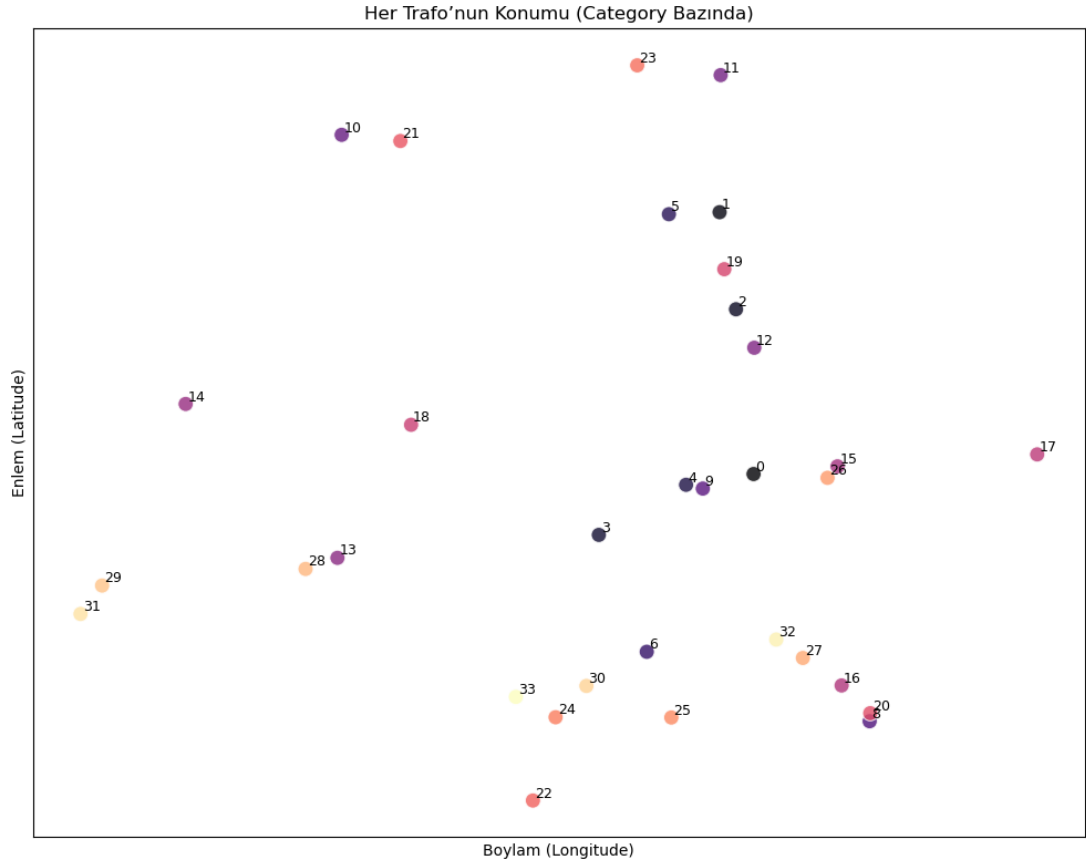
İzmir ve Manisa şehirlerinde yer alan elektrik dağıtım firması tarafından alınan veriler ile <https://open-meteo.com/> üzerinden güç transformatörlerinin enlem ve boylam koordinatları vasıtasıyla, ilgili tarihlerdeki hava durumu verileri ve güç transformatörlerine ait temin edilen duran varlık verileri ile veri seti hazırlanmıştır. Hava durumu verileri Open Meteo tarafından temin edilen hazır API kodları üzerinden deęişiklik yapılarak türetilmiştir. Oluşturulan API vasıtasıyla enlem ve boylam koordinatlarına göre ilgili konumlar için günlük bazda ilgili sütunlar temin edilmiştir.

Modelleme kapsamında hedef özellik (Kesinti Durumu) haricinde 59 farklı özellik kullanmış olup hava durumu ve meta veriler haricinde kalan veriler modelle süreci öncesinde kullanılan metodolojiler ile oluşturulmuştur.

3.1.2 Veri Dağılımı

Model çalışmalarına ve veri ön işleme sürecine başlamadan önce okunabilirlik kapsamında düzenlemeler gerçekleştirilmiştir. Bu kapsamda; enlem ve boylam koordinatlarında kesirli alan için yedi basamakta sınırlandırılacak şekilde yuvarlama işlemi gerçekleştirilmiştir. Güncellenen koordinatlar sonrasında yapılan incelemelerde eksik veri tespit edilmemiştir.

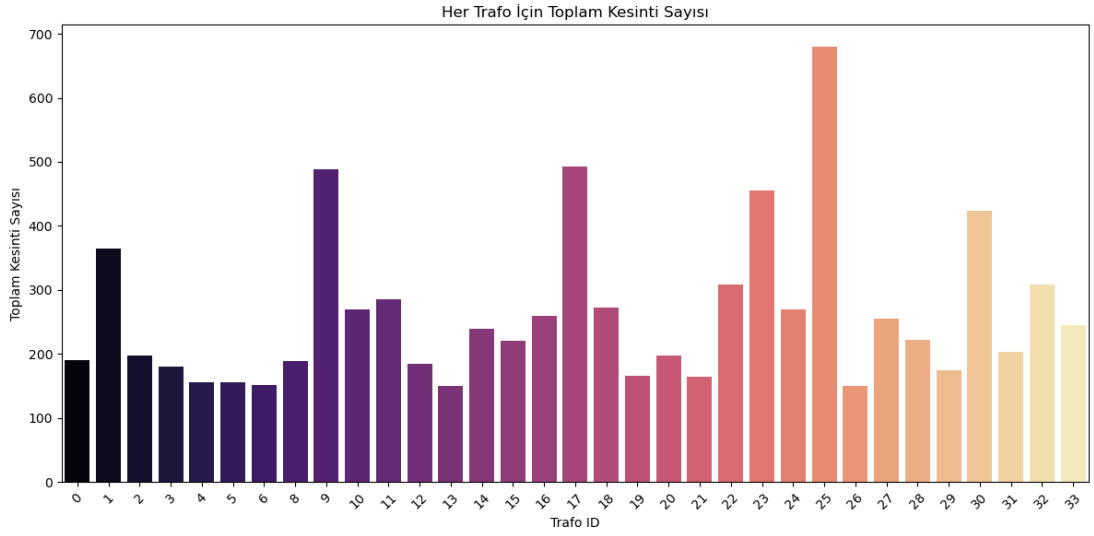
Hedef sütun olan “Kesinti” bilgisini anlamlandırmak için incelenen veri setinde%82.013 “Kesinti Yok” verisi ile %17.987” “Kesinti Var” verisi bulunmaktadır. Gerçek saha verilerinde gözlemlenen dengesiz veri dağılımı kapsamında, modelleme çalışmalarında sınıf dengesizliğinin deęerlendirilmeye alınmasına karar verilmiştir.



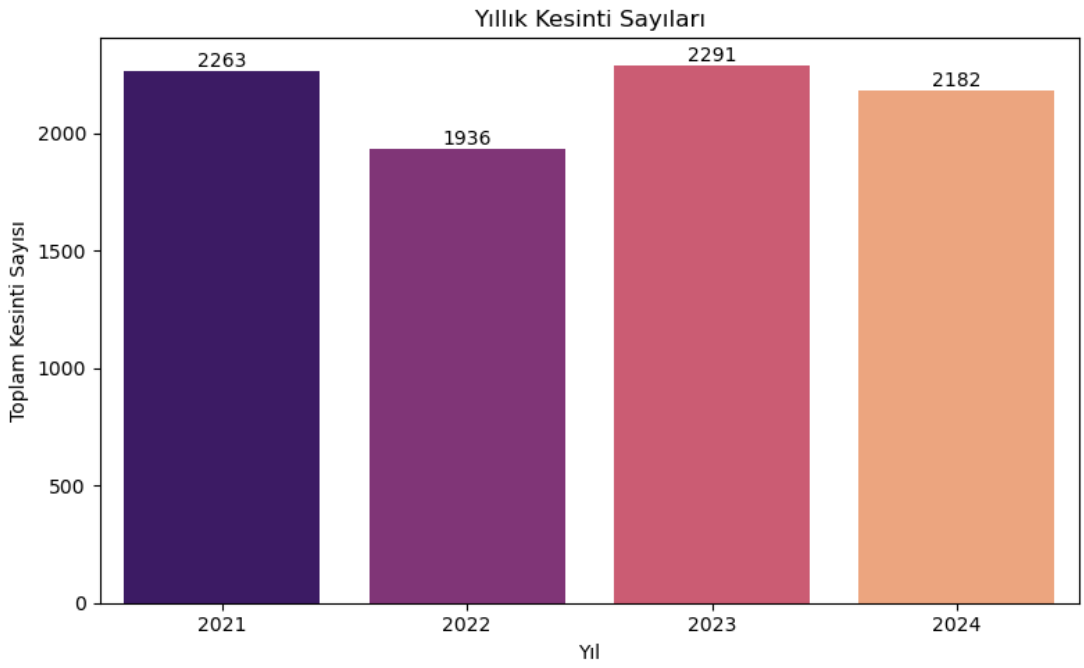
Şekil 3 Transformatörlerin Haritalandırılması

Oluşan her bir kesinti hava durumu ile ilişkili olmak zorunda değildir. Mevcut şebekede bakım eksikliği, anlık yük değişimi, harmonikler, teknik olmayan kayıp enerjiler gibi birçok sebepten dolayı enerji kesintisi de yaşanabilmektedir. Şekil 3'e göre konumsal olarak transformatöre mesafeleri incelendiğinde 8 numaralı ve 20 numaralı transformatörlerin konum olarak yakın olduğu tespit edilmiştir. 4,9,15 ve 26 numaralı transformatörler için de benzer bir durum bulunmaktadır.

Şekil 4'te görüldüğü üzere 8 ve 20 numaralı transformatörlerde kesinti sayısı yakın olarak görülmesine rağmen 4 ve 9 numaralı transformatörlerde yaklaşık olarak 2,5 kat oranında fark bulunmaktadır. Bu durumda bütün kesintilerin hava durumu ile ilişkilendirilemeyeceğinin örneklerinden biridir. Ancak birçok kesintinin sebebi hava durumu ile ilişkili olabileceği düşünülmektedir. Özellikle havai hatlardan dolayı ilişkiselliğin yüksek olması mevcut durum için örnek oluşturmaktadır.

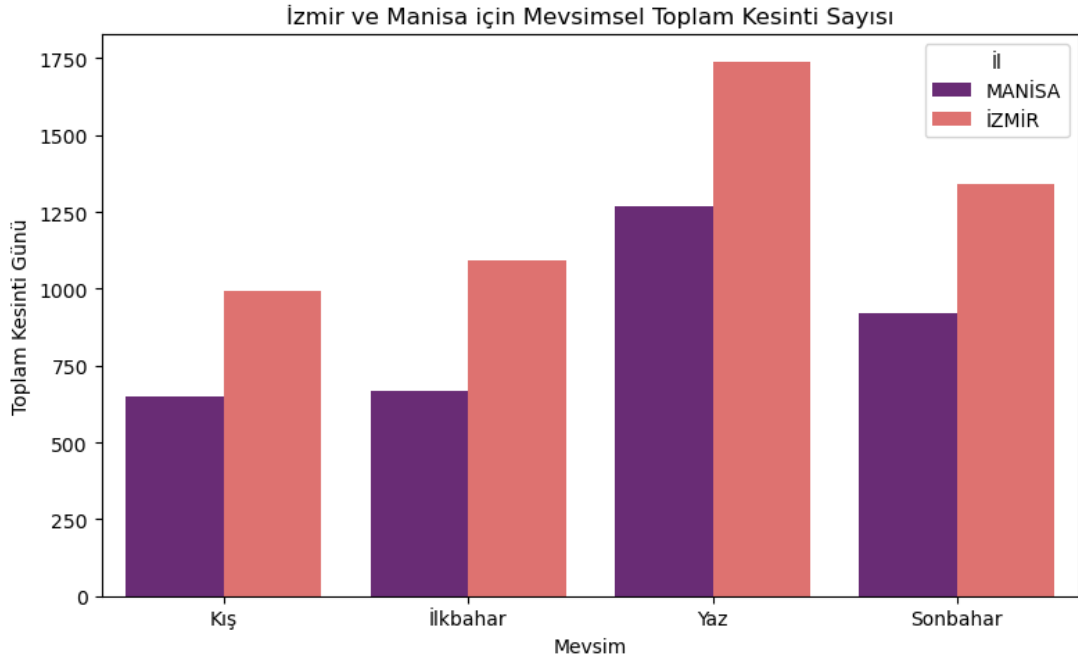


Şekil 4 Transformör Özelinde Kesinti Sayıları



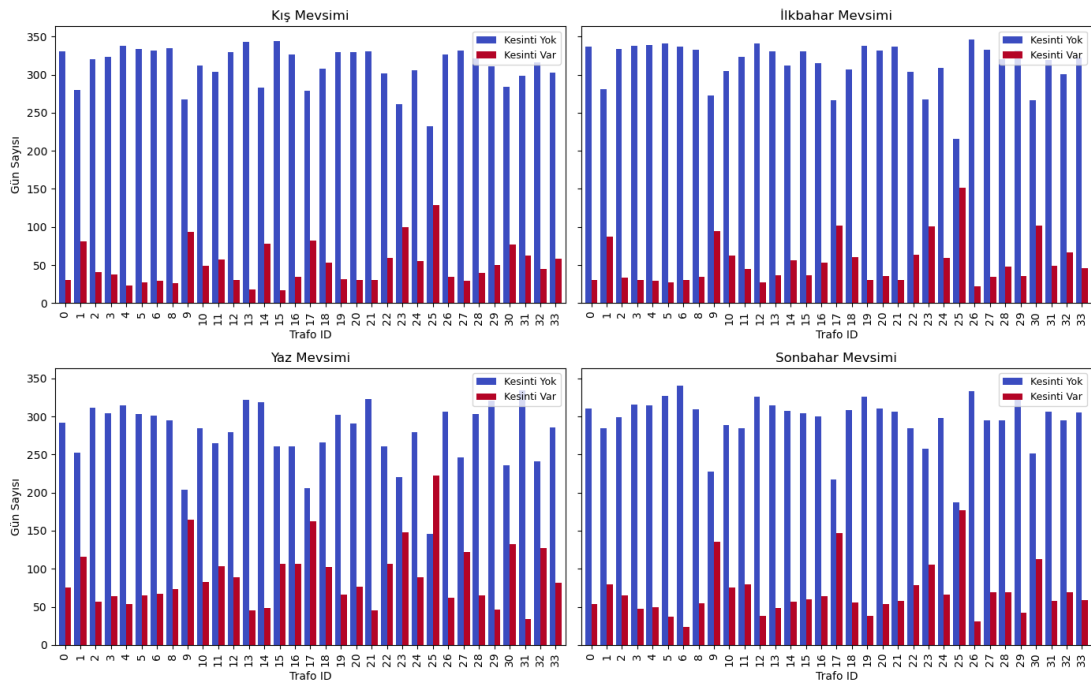
Şekil 5 Yıllara Göre Kesinti Dağılımı

Farklı bir perspektifler ile kesinti dağılımının incelenmesi yaklaşımları gerekirse, Şekil 5’de yıllara göre toplam plansız kesinti sayısındaki farklılıklar gözükmemektedir. Birbirine yakın değer gösteren 4 yıl olmasına karşın mevsimsellik, zaman serisi gibi durumların incelenmesinin gerekliliğini göstermiştir.



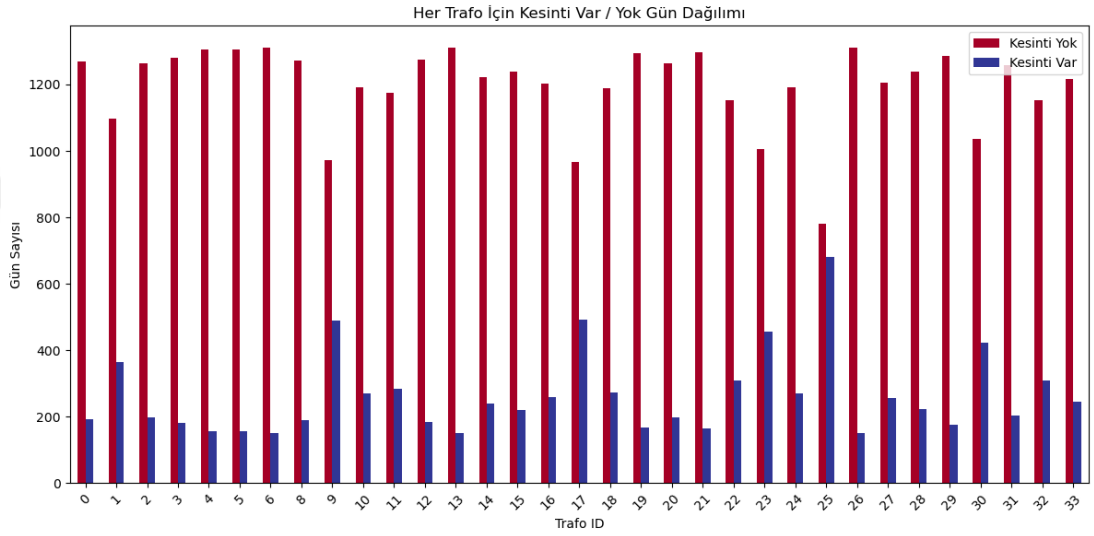
Şekil 6 Şehirlere Göre Mevsimsel Kesinti Dağılımı

Trafo Bazında Kesinti Durumu (Mevsimlere Göre)



Şekil 7 Transformör Bazında Mevsimsel Kesinti Dağılımı

Mevsimsel Kesinti Sayısı ve Haftanın günlerinde yer alan kesinti dağılımını daha detaylı incelenebilmesi için Şekil 7’de her trafoya ait mevsimsel olarak kesinti sınıfı dağılımı gözlemlenmiştir. Genel olarak yaz aylarında artış gözlemlenmekte olup, özellikle 25 numaralı transformatörün yaz aylarında, gün sayısı bakımında “Kesinti Yok” sınıfının geçtiği tespit edilmiştir. Bu durum, her transformatörün hava durumu ile benzer hassasiyeti ve ilişkiye sahip olmadığını da kanıtlamaktadır.



Şekil 8 Ayın Günlerine Göre Transformatör Bazlı Kesinti Dağılımı

Genel olarak inceleme yapıldığında, 4 yıllık veri içerisinde her bir transformatörün toplam gün sayısı içerisinde kesinti sınıf dağılımı Şekil 8’de görselleştirilmiştir.

3.1.3 Veri Ön İşleme Adımları

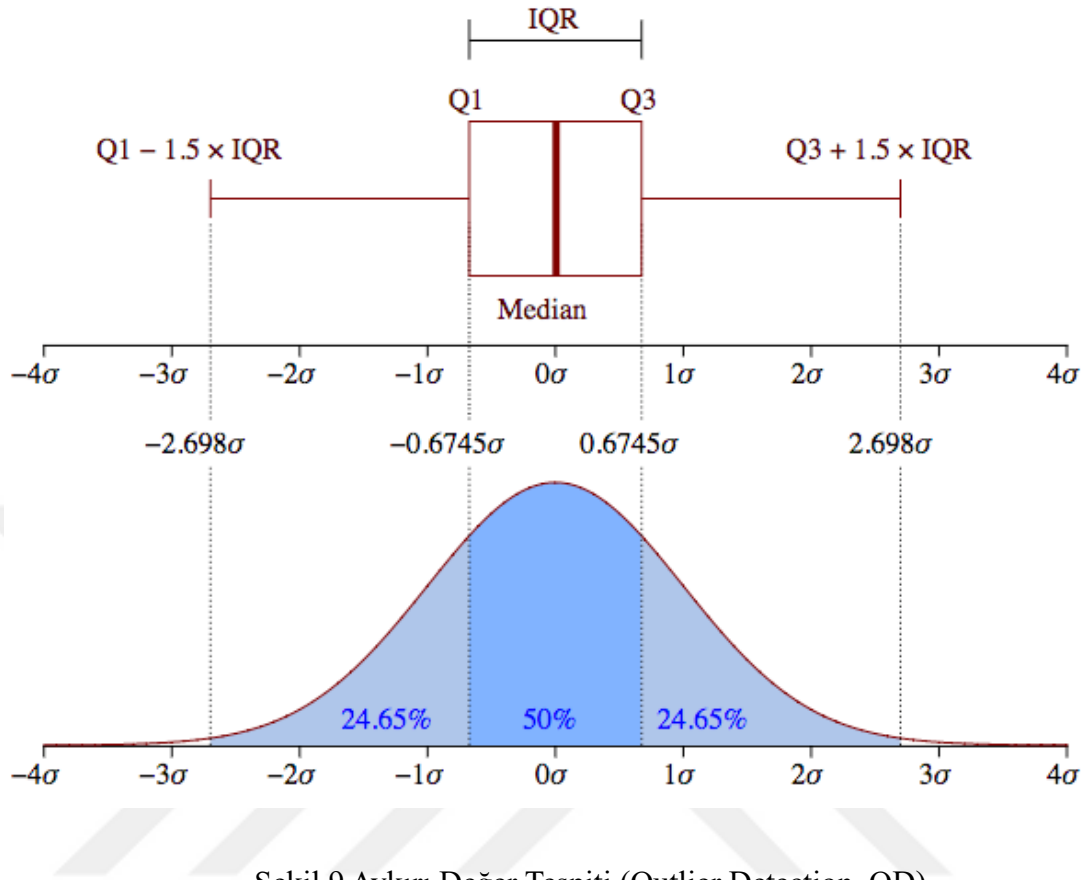
Verilerin mevsimselliği, yıllara göre kategorize olma durumu ve kesinti sınıfının dağılımı gibi gözlemlerin ardından, DL metodolojileri ve ML metodolojileri kapsamında çalışmalara başlanmadan önce, veri seti üzerinde ön işlemler yapılması gerekmektedir. Bu kapsamda, veri setinde yer alan “string” değerlerin “integer” ya da “float” formatlarına çevrilmesi, gerekli olmayan sütunların kaldırılması, verilerin ölçeklendirilmesi vb. işlemlerin yapılması sağlanmıştır [33], [34], [35], [36], [37], [38].

3.1.3.1 Label Encoding Ve One-Hot Encoding

Tahminleme modelleri tasarlanmadan önce veri setinde incelemeler ve düzenlemeler yaparak modelleme için hazır hale gelmesi gerekmektedir. Bu kapsamda, nominal kategorik değişkenlerin ML algoritmaları tarafından işlenebilmesi “Label Encoding” yöntemi ile sağlanmıştır. Bu yaklaşım, her bir kategori değerine ardışık tam sayı atayarak veri setindeki kategorik sütunları sayısal formata dönüştürülmesini sağlamaktadır. İl, İlçe, Marka, Fan Durumu gibi metinsel etiketler 0–N aralığında tam sayılara eşlenerek modele girdi olarak kullanılmıştır. Bu sayede algoritmanın bu değişkenler üzerinde matematiksel işlem yapması mümkün hâle gelmiştir. Label Encoding, özellikle ağaç tabanlı modellerde (Random Forest, CatBoost vb.) performans kaybı olmadan kullanılabilirdiği gibi, sıralı bilgi içermeyen kategorilerde de basit ve hızlı bir ön işleme adımı sunmaktadır. Bununla birlikte, atama sıralamasının yanlış algılanmasını önlemek adına, model eğitiminden önce kategoriler arası gerçek hiyerarşi olup olmadığı dikkatle değerlendirilmiş ve One-Hot Encoding gibi alternatif kodlama yöntemleriyle karşılaştırılmıştır.

One-Hot Encoding, her bir kategori için ayrı bir ikili (binary) sütun oluşturarak, ilgili kategori gözlemi için 0-1 değerleri atanmasını içermektedir. Böylece gerçek sütunda yer alan N farklı sınıf, N adet yeni sütuna dönüştürerek sütunlar üzerinden yalnızca “var/yok” bilgisiyle çalışmasını sağlamaktadır. Bu yaklaşım, özellikle kategoriler arasında doğal bir sıralama ilişkisi bulunmadığında veya algoritmanın kategorik seviyeleri birbirinden bağımsız olarak değerlendirmesi istendiğinde tercih edilmektedir. Özellikle doğrusal modeller gibi Label Encoding’in ardışık sayı atamalarının yanlış regresyon eğilimlerine yol açabileceği durumlarda, One-Hot Encoding hem yorumlanabilirliği arttırmaktadır hem de modelin her kategoriye eşit mesafede yaklaşmasını sağlamaktadır. Bu kapsamda değerlendirme yapıldığında, “İL,” “İLÇE”, “Marka”, “FAN_DURUMU”, “SOGUTMA_TURU”, “TIPI” ve “Gücü” sütunlarında inceleme yapılmıştır. Yapılan incelemede ve test modellerinde alınan sonuçlara istinaden ilgili sütunların her biri “Label Encoding” işlemine tabi tutulmuştur.

3.1.3.2 Aykırı Değer Tespiti



Şekil 9 Aykırı Değer Tespiti (Outlier Detection, OD)

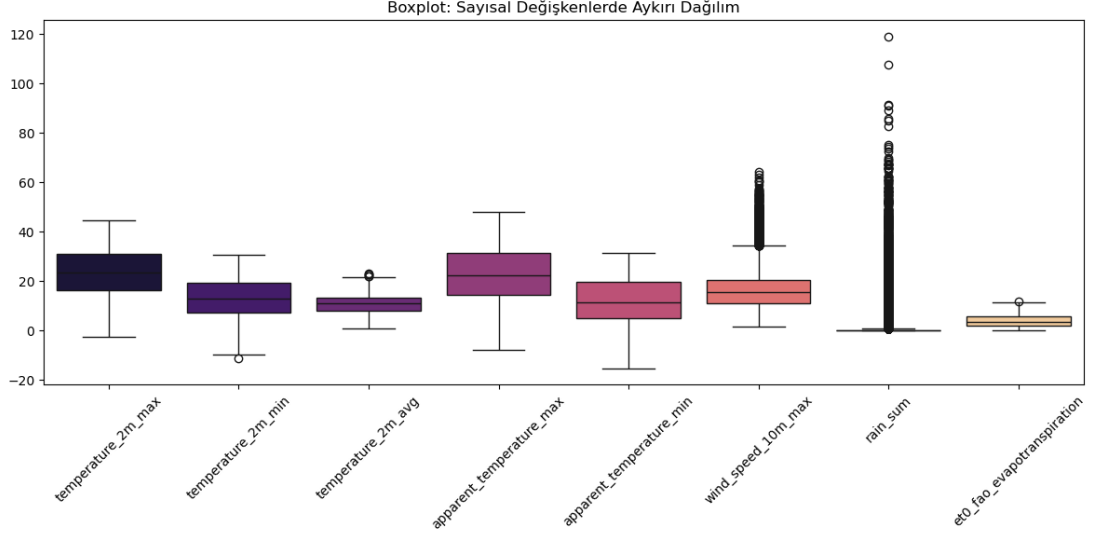
Label Encoding işlemi ardından, Şekil 9’da görüldüğü gibi ilgili sütunlara OD çalışması yapılmıştır. OD, bir veri kümesinde genel veri yapısından anlamlı biçimde sapma gösteren gözlemlerin tespit edilmesi sürecini tanımlamaktadır. Bu gözlemler, genellikle ölçüm hataları, veri giriş hataları, olağan dışı durumlar ya da istatistiksel olarak nadir gerçekleşen olaylar adlandırılmaktadır. Aykırı değerler, veri analizi, ML ve istatistiksel modelleme süreçlerinde sonuçları ciddi şekilde bozabileceği için tespiti ve yönetimi kritik öneme sahiptir [42].

OD tanımlaması kapsamında; Interquartile range (IQR) tanımlanmaktadır. IQR, çeyrek aralığı olarak anlamlandırılmakta olup Q2 değeri medyan iken, 25. çeyrek (Q1) ve 75.çeyrek (Q3) olarak tanımlanmaktadır.

(i) Minimum : $Q1 - 1.5 * IQR$ (1)

(ii) Maksimum : $Q3 + 1.5 * IQR$ (2)

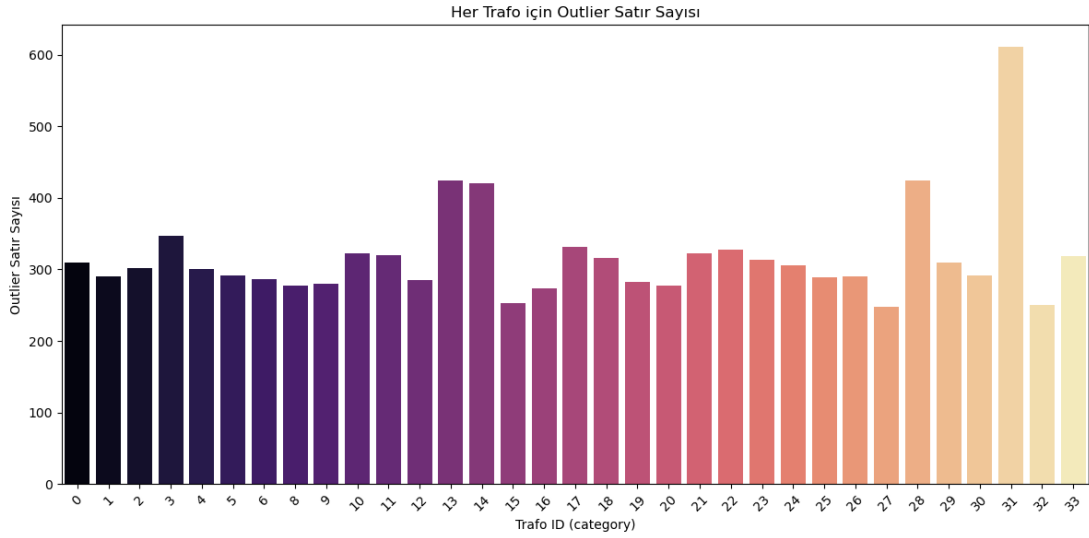
Şekil-9’da ortada kalan alandaki değerler aykırı değilken köşelerde kalan alan aykırı değer olarak tespit edilmektedir.



Şekil 10 Sayısal Değişkenlerde Aykırı dağılımın BoxPlot ile Gösterimi

Bu kapsamda yapılan çalışmada Şekil-10’da gösterilen hava durumu değişken özellikleri için OD çalışması yapılmıştır. OD çalışması; “is_outlier” sütunu oluşturularak Boolean metodu ile “VAR/YOK” olarak doldurulan sütun, daha sonrasında “Label Encoding” işlemine tabi tutularak numerik hale getirilmiştir. Model çalışmaları kapsamında aykırı değerlerin tamamı silinmemiştir. Bunun yerine “is_outlier” sütununun dahil olduğu, “is_outlier” sütununun dahil olmadığı ve aykırı değerlerin (“is_outlier” sütunun da) olmadığı üç farklı veri seti oluşturulmuştur [42].

Veri setleri incelendiğinde OD olan satır sayısı 10492 adet olarak tespit edilmiştir. Her bir transformatör kapsamında aykırı değer dağılımı Şekil-11’de gösterilmektedir. 13, 14, 28 ve 31 numaralı transformatörlerin aykırı değerleri en çok barındıran transformatörler olması ile ön plana çıkmaktadır.

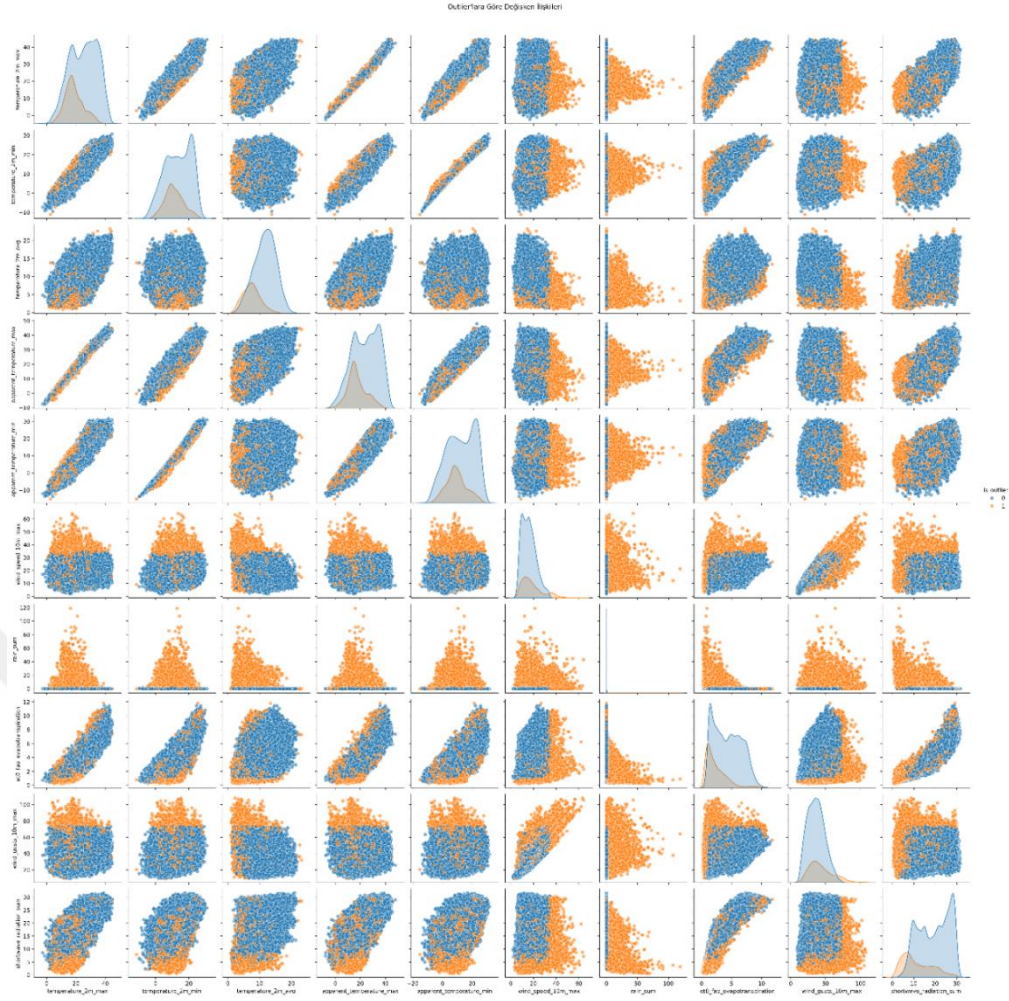


Şekil 11 Her Transformator için Aykırı Değer Dağılımı

3.1.3.3 KORELASYON

Şekil-12’de aykırı değerlere göre değişkenlik görselleştirilmiştir. Ayrıca tüm sütunların kesinti durumuna göre korelasyonu incelendiğinde Tablo-2’de yer alan değerlerin mutlak açıdan en yüksek korelasyon ilişkilerine sahip olduğu tespit edilmiştir. Daha çok sabit verilerin (meta verilerin) korelasyona sebebiyet vermesi ve sıcaklık bilgilerinin korelasyon değerinin yüksek olması, sanılanın aksine rüzgâr, gün ışığı vb. verilerin kesinti ile ilişkisinin daha az olduğunu göstermiştir. Bu durumun ve “Korelasyon İlişkisi” değerlerinin -1 ve 1 değerlerine uzaklığı ilişki kurulmasındaki zayıflığı vurgulamaktadır.

Mevcutta gerçekleşen tüm kesintilerin hava durumu verileri ile ilişkiselliği mümkün olmadığı için korelasyon değerleri de benzer bir sonuç olarak çıktı vermektedir.



Şekil 12 Aykırı Değer Ve Aykırı Değer Olmayan Verilerin İlişkisi

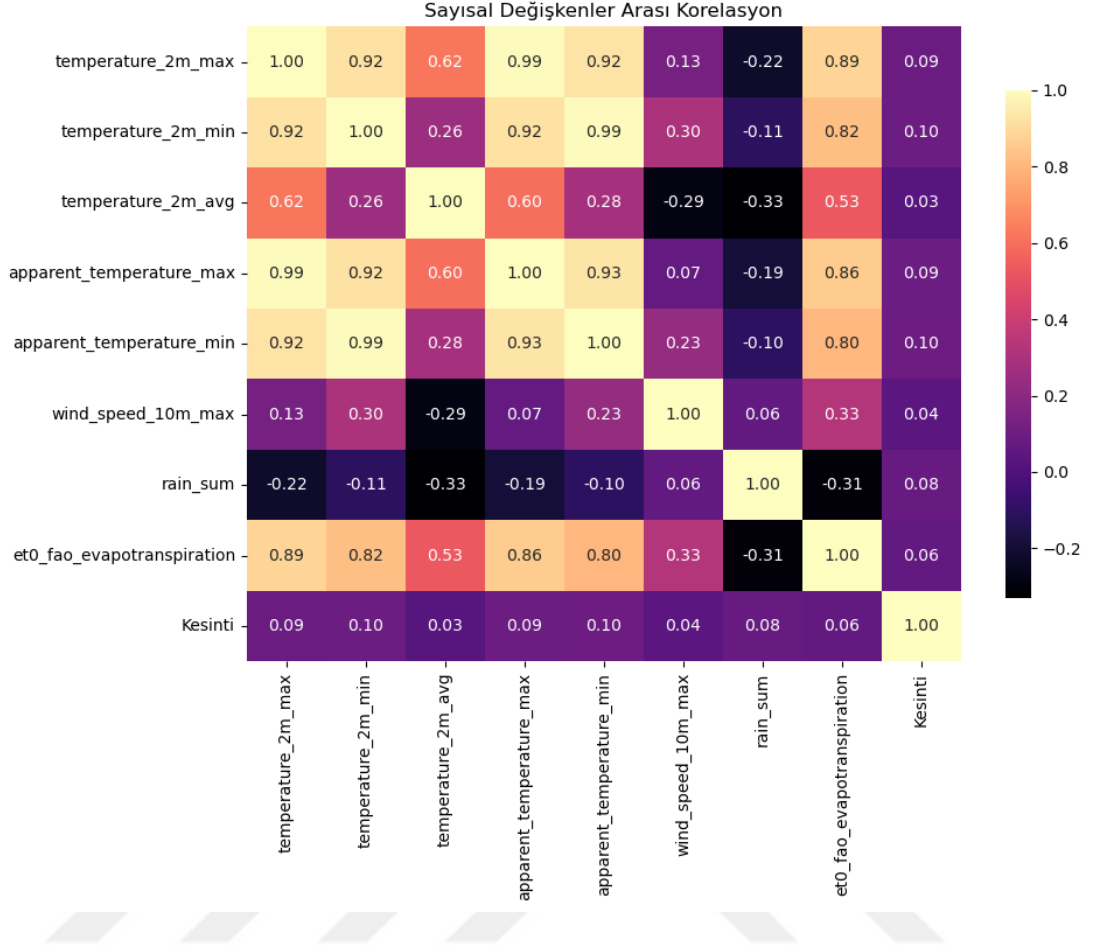
Şekil-12’de aykırı değerlere göre değişkenlik görselleştirilmiştir. Ayrıca tüm sütunların kesinti durumuna göre korelasyonu incelendiğinde Tablo-2’de yer alan değerlerin mutlak açıdan en yüksek korelasyon ilişkilerine sahip olduğu tespit edilmiştir. Daha çok sabit verilerin (meta verilerin) korelasyona sebebiyet vermesi ve sıcaklık bilgilerinin korelasyon değerinin yüksek olması, sanılanın aksine rüzgâr, gün ışığı vb. verilerin kesinti ile ilişkisinin daha az olduğunu göstermiştir. Bu durumun ve “Korelasyon İlişkisi” değerlerinin -1 ve 1 değerlerine uzaklığı ilişki kurulmasındaki zayıflığı vurgulamaktadır.

Mevcutta gerçekleşen tüm kesintilerin hava durumu verileri ile ilişkiselliği mümkün olmadığı için korelasyon değerleri de benzer bir sonuç olarak çıktı vermektedir.

Tablo 2 Sayısal Değerlerin Kesinti Sınıfı ile Korelasyonu

Veriler	Korelasyon İlişkisi
Müşterek Direk Sayısı	0.158989
Hissedilen En Düşük Sıcaklık	0.097917
2 metre yükseklikteki en düşük sıcaklık	0.096012
Hissedilen en yüksek sıcaklık	0.092539
2 metrede yükseklikteki en yüksek sıcaklık	0.089443
Transformatöre Bağlı hatların toplam uzunluğu (Km)	0.084659
YG Direk Sayısı	0.083755
1 Birim Alana Düşen Yağmur Miktarı	0.079104
Toplam Yağış Miktarı	0.079049
Güneşin Doğuşu	-0.075446

Şekil 12’de görselleştirilen aykırı değerler ve aykırı olmayan değerlerin korelasyonu akabinde, Şekil 13 kapsamında aykırı değer için kullanılan sütunların korelasyonu ısı hartiası olarak gösterilmiştir. Tablo-2 ‘deki verilere benzer bir şekilde sıcaklık değerleri haricinden 1 veya 1 değerlerine yakında bir sütun gözlemlenmemektedir.



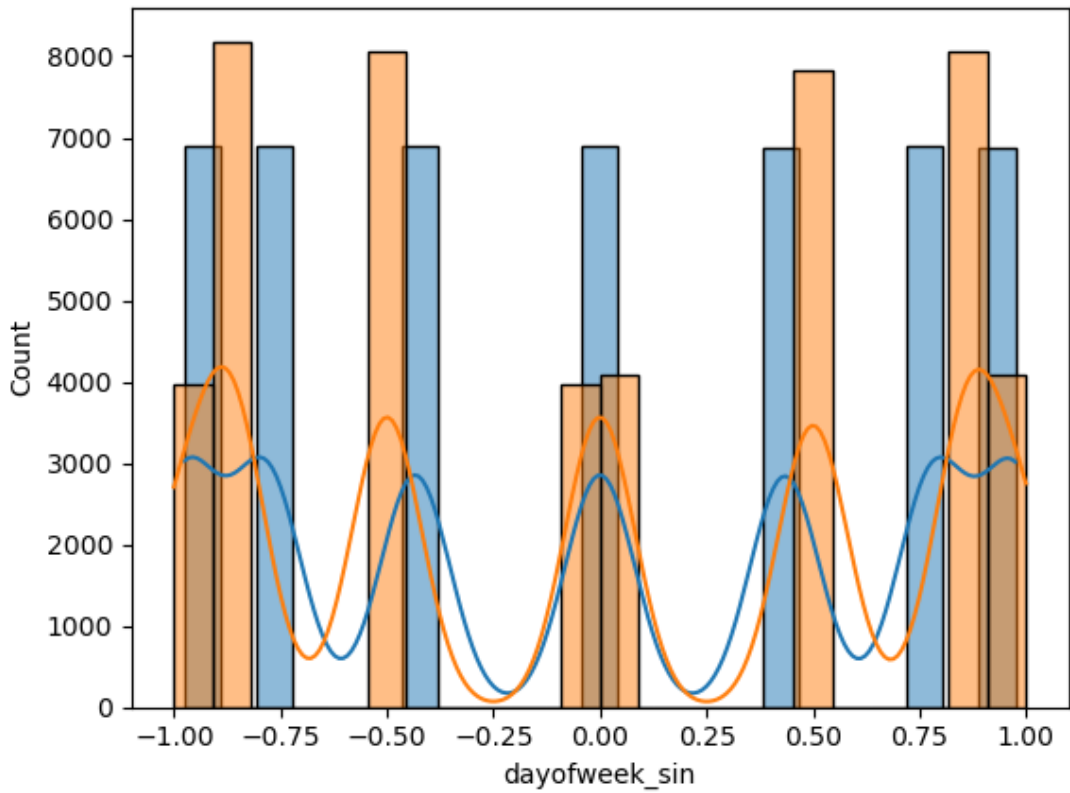
Şekil 13 Aykırı Değer İçin Kullanılan Sütunların Korelasyonu Isı Haritası

3.1.4 Zaman Gecikmesi ve Kayan Pencere Metodolojisi (Lag Rolling)

Yapılan incelemeler sonucunda, kesinti sınıf dağılımının dengesiz olması ve güç transformatörlerinde oluşan kesintilerin hava durumu ile korelasyonunun beklenenden düşük olması yeni özellikler türetilmesini elzem kılmıştır. Bu kapsamda sezonsallık ve mevsimsellik kısmından da ölçümlenebilmesi için; 1 gün içerisinde gerçekleşen; hissedilen en düşük sıcaklık, 2 metre seviyesindeki en düşük sıcaklık, hissedilen en yüksek sıcaklık, 1 birim alana düşen yağmur miktarı ve 2 metre seviyesindeki en yüksek sıcaklık gözlemlerinde “lag-rolling” metodolojisi kullanılmıştır.

“Lag”, gecikmeli olarak ilgili deęerin yansıtılması olup bu kapsamda ilgili sütünların bir gn nceki deęerlerinin yansıtılmasına karar verilmiřtir. Sezonel ve mevsimsel inceleme yapılmasının kolaylařtırılması iin “Rolling” deęeri 7 olarak belirlenmiřtir. “Rolling”, ilgili stundaki deęerin gemiře ynelik yedi gn boyunca gsterdięi deęerlerin ortalamasıdır. Benzer alıřma kesinti durumu iin de yapılmıř olup yeni oluřan stunlar veri setine dahil edilmiřtir. 2021 yılının birinci gn itibariyle bařlayan veri seti iin 1 Ocak 2021 ile 7 Ocak 2021 tarihleri arasında “lag-rolling” stunlarının deęerleri NaN biiminde oluřmaktadır. Eksik veri durumunun ortadan kaldırılması iin oluřturulan stunlardaki NaN deęerleri medyan ile doldurularak veri setinde oluřan eksikliklerin giderilmesi saęlanmıřtır.

ML modellerinde ve DL metodolojilerinde zamanın daha iyi tanımlanması iin “time” stunu referans alınarak; “season”, “year”, “month”, “week”, “day”, “dayofweek” stunları oluřturulmuřtur. Periyodik rntlerin ML tarafından algılanması iin sin/cos dnřm uygulanarak yeni stunlar oluřturulmuřtur.



řekil 14 Sinus Verileri ile Haftanın Gnleri

3.2 Model Oluřturma Sreçleri

3.2.1 Eđitim Verisi Dzenlemeleri

ML ve DL metodolojilerinin kullanılacağı veri seti integer ve float deęerlere dnřtrlerek gerek n iřlem sreçlerinden geçirilmiřtir. Deęerlendirilmeye alınacak birden fazla sreç olduđu iin her bir sreç ve model iin farklı durumlara ait eđitim ve test verisi oluřturulmuřtur.

3.2.1.1 Test – Eđitim Verileri

2024 yılının son eyređine ait verilerin test verisi olarak kullanılması, Ocak 2021 yılı ile 2024 Eyll ayının son gnne kadar olan verilerin eđitim iin kullanılmasına karar verilmiřtir. Test verisi zerinde deęiřiklik (sentetik veri retimi, veri azaltımı vb. durumlar) yapılamayacağından dolayı eđitim verisi zerinde birden fazla senaryo ile inceleme yapılmıřtır.

Oluřturulan veri setinin hibir filtre veya eksiltme yapılmadan kullanıldıđı durum iin eđitim verisinde bir deęiřiklik yapılmamıřtır. OD yapılan verilerin model ierisinde oluřturduđu fayda – zarar iliřkisini gzlemlemek iin aynı eđitim verisi ierisinde “is_outlier”,(OD durumu) stunu “1” deęerine sahip olan satırların ıkartılması ile yeni bir eđitim seti daha oluřturulmuřtur. “is_outlier” stunu; aykırı deęerleri iřaretleme yapmasından kaynaklı filtre yapılmamıř eđitim veri seti iin “**bayraklı eđitim seti**”, filtre yapılan veri seti iin ise “**outliersiz eđitim seti**” olarak isimlendirilmiřtir.

Bayraklı eđitim setinde 45.177 satır veri bulunmasına karřın outliersiz eđitim seti ierisinde 35.458 satır veri bulunmaktadır. Test verisi birebir aynı olmasına karřın model eđitim sreçlerinde veri sayısında oluřan farklılıkların yanında eđitim veri setleri zerinde uygulanması gereken leklendirme iřlemi iin iki farklı metodoloji kullanılmıřtır.

3.2.1.1.1 Minmax Scaler

Veri ön işleme sürecinde, modelin farklı ölçeklerdeki özellikler arasında adil karşılaştırma yapabilmesi için MinMax Ölçekleyici (MinMax Scaler) kullanılmıştır. Bu yöntem, her bir sürekli değişkenin gözlem değerlerini minimum (min) ve maksimum (max) değerler arasındaki aralığa (genellikle 0 ile 1) lineer olarak dönüştürür. Dönüşüm formülü şu şekildedir:

$$X_{Scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

MinMax Ölçekleyici, verinin dağılımını bozmadan tüm gözlemleri belirlenen aralığa taşıırken, algoritmaların öğrenme sürecinde özelliklerin mutlak büyüklüklerinden kaynaklanan ağırlık dengesizliklerini ortadan kaldırmaktadır.

Bununla birlikte, aykırı değerlerin dönüştürme aralığını aşırı derecede genişletebilme riski bulunduğundan, önce aykırı değer analizi ve kırpma uygulamalarına istinaden eğitim setleri yapılarak MinMax ölçeklemenin etkisi kontrol altına alınmıştır. Bu sayede, transformatör arıza verilerindeki sıcaklık, nüfus yoğunluğu ve diğer fiziksel özellikler homojen bir ölçekte modellenerek, tahmin performansında iyileşme sağlanmıştır.

3.2.1.1.2 Robustscaler

Veri normalizasyonu aşamasında uç değerlerin (aykırı değer) etkisini en aza indirmek amacıyla Robust Scaler yöntemi uygulanmıştır. Bu yöntem, her bir sürekli değişkenin merkezi eğilimini medyan ve yayılımını ise çeyrekler arası mesafe (IQR—Interquartile Range) temel alarak ölçeklendirmektedir. Dönüşüm işlemi şu formülle gerçekleştirilir:

$$X_{Scaled} = \frac{x - \text{medyan}(x)}{IQR(x)} \quad (4)$$

Medyan(x), ilgili deęiřkendeki ortanca deęeri, IQR(X) ise $Q_3 - Q_1$ olarak üçüncü çeyrek ile birinci çeyrek farkını temsil etmektedir.

Böylece gözlemler, mutlak deęerlerden ziyade daęılımın orta bölgesine göre normalize edilmektedir ve aşırı sapmaların ölçekleme sonucunu çarpıtması engellenmektedir. Uygulama öncesinde medyan ve IQR deęerleri eğitim setinden türetilir, test veri setine bu istatistikler sabitlenerek dönüřtürme işlemi tekrarlanır; böylece veri sızıntısı (data leakage) önlenerek genellenebilirlik korunmaktadır. Bayraklı eğitim seti ve outliersiz eğitim seti için 2 farklı ölçeklendirme metodolojisi olan MinMax Scaler ve Robust Scaler uygulanarak 4 farklı eğitim seti oluşturulmuřtur. Ölçeklendirme yapılan deęerlerin kategorik bir deęer olmaması gerektięi için hava durumu verileri ve Lag – Rolling metodu ile üretilen parametrelere ve gözlemlere uygulanmıřtır.

ML ve DL algoritmaları ile oluşturulacak modellerin performansının ölçümlenmesi gerekmektedir. “Literatür Taraması” kapsamında yapılan incelemelerde; doęruluk, kesinlik veya F-1 skoru üzerinden ölçüm yapılmaktadır. Oluřturulan eğitim veri setleri ve test verilerinin başarısının ölçümlenmesi ve deęerlendirilmesi için birden fazla başarı parametresinin incelenmesi gerekmektedir. Bu kapsamda model eğitimi sırasında başarı parametrelerinin çıktılarının oluşturulması için tanımlanan metrikler řu şekildedir:

(i) Doęruluk (Accuracy)

Doęruluk, sınıflandırma modelinin tüm tahminleri içinde doęru sınıflandırma oranını ifade etmektedir. Özellikle dengesiz olmayan veri kümelerinde modelin genel başarı seviyesini göstermektedir. Bu nedenle dengesiz veri setlerinde yanıltıcı sonuçlar üretebilir. Matematiksel olarak;

$$Doęruluk = \frac{Doęru Pozitif (TP) + Doęru Negatif (TN)}{Doęru Pozitif (TP) + Doęru Negatif (TN) + Yanlıř Pozitif (FP) + Yanlıř Negatif (FN)} \quad (5)$$

şeklinde tanımlanır. Burada TP (True Positive) doğru pozitif, TN (True Negative) doğru negatif, FP (False Positive) yanlış pozitif ve FN (False Negative) yanlış negatif örnek sayılarıdır.

(ii) F₁ Skoru (F-1 Score)

F₁ skoru, hassasiyet (precision) ve duyarlılık (recall) değerlerinin harmonik ortalamasını sunmaktadır. 6 da belirtilen formüle göre hesaplaması yapılmaktadır. ;

$$F_1 = 2 \times \frac{\text{Hassasiyet} \times \text{Duyarluluk}}{\text{Hassasiyet} + \text{Duyarluluk}} \quad (6)$$

(iii) F₂ Skoru (F-2 Score)

F₂ skoru, F₁ skoruna benzer biçimde Precision ve Recall değerlerini birleştirir ancak Recall değerini iki kat ağırlık vermektedir. Kritik hataların (FN) maliyetinin operasyonel olarak daha yüksek olduğu senaryolarda tercih edilir. Matematiksel tanımı 7’de belirtilmiştir.

$$F_2 = (1 + 2^2) \times \frac{\text{Hassasiyet} \times \text{Duyarluluk}}{(2^2 \times \text{Hassasiyet}) + \text{Duyarluluk}} \quad (7)$$

(iv) Hassasiyet (Precision)

Hassasiyet, pozitif sınıf olarak etiketlenen örneklerin ne kadarının gerçekten pozitif olduğunu ölçmektedir. Yanlış uyarı oranını minimize etmeye odaklanmaktadır.

$$\text{Hassasiyet} = \frac{TP}{TP+FP} \quad (8)$$

(v) Duyarlılık (Recall) -- sensitivity

Duyarlılık, pozitif olan örneklerin model tarafından ne oranda doğru tespit edildiğini gösterir. Kaçırılmaması gereken olgu veya hastalık taramalarında kritik metriklerden biridir.

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \quad (9)$$

(vi) ROC–AUC (Receiver Operating Characteristic – Area Under Curve)

ROC eğrisi, modelin farklı eşik değerlerinde FP oranı (1 – specificity) ile TP oranı (sensitivity) arasındaki ilişkiyi gösterir. Eğri altında kalan alan AUC değeri 0.5 ile 1 arasında olup, 1'e yaklaştıkça modelin ayırt edici gücü artar. Matematiksel kapalı formülü yoktur; eğri altındaki integral yoluyla hesaplanır.

Specificity, özgülük olarak Türkçe'ye çevrilmektedir. Negatif sınıfa ait gerçek örneklerin model tarafından ne orada doğru tespit edildiğini temsil etmektedir. Matematiksel formülü ise şu şekildedir:

$$\text{Özgüllük} = \frac{TN}{TN+FP} \quad (10)$$

Özgüllük 0 – 1 arası değerleri almaktadır. 1 değerine yaklaştıkça negatif örnekler üzerinde yanlış alarm oranının azaldığı anlamına gelmektedir.

(vii) PR–AUC (Precision–Recall Curve – Area Under Curve)

PR eğrisi ise farklı eşiklerde precision ve recall değerlerinin grafiğini sunmaktadır. Özellikle az görülen pozitif sınıfın performansını değerlendirmede ROC–AUC'dan daha duyarlıdır. Eğri altındaki alan ne kadar büyükse modelin az sayıda pozitif örneği doğru ayırt etme başarısı o kadar yüksektir.

(viii) MAE (Mean Absolute Error)

MAE, gerçek değerlerle model tahminleri arasındaki mutlak farkların ortalamasıdır ve regresyon hatalarının ortalama büyüklüğünü verir.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (11)$$

(ix) RMSE (Root Mean Squared Error)

RMSE, kare hata ortalamasının kareköküdür ve büyük hataları daha yüksek ağırlıklayarak ortalama hatayı ölçer.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (12)$$

Formülü nedeniyle uç hatalar karşı MAE'ye göre daha hassastır.

(x) MCC (Matthews Correlation Coefficient)

MCC, dört hücreli karışıklık matrisinin tüm bileşenlerini dikkate alan bir korelasyon katsayısıdır ve -1 ile $+1$ arasında değer alır. $+1$ tam mutabakat, 0 rastgele tahmin, -1 tam ters ilişkiyi gösterir.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}} \quad (13)$$

(xi) Karışıklık Matrisi (Confusion Matrix, CM)

CM, sınıflandırma sonuçlarını dört hücrede özetler:

- TN (True Negative): Negatif örneklerin doğru tahmini
- TP (True Positive): Pozitif örneklerin doğru tahmini
- FN (False Negative): Pozitif örneklerin negatif sınıfa yanlış atanması
- FP (False Positive): Negatif örneklerin pozitif sınıfa yanlış atanması

Karışıklık matrisi, diğer tüm sınıflandırma metriklerinin hesaplanmasına temel oluşturur.

(xii) G-Oran (G-Mean)

G-Mean, pozitif ve negatif sınıfların başarı oranlarının geometrik ortalamasıdır; sınıflar arası dengeyi gözetir.

$$G\ MEAN = \sqrt{\text{Duyarlılık} \times \text{Özgüllük}} = \sqrt{\frac{TP}{TP+FN}} \times \sqrt{\frac{TN}{TN+FP}} \quad (14)$$

Dengesiz sınıflandırmalarda her iki sınıf için de dengeli ölçüm yapılmasını sağlamaktadır.

(xiii) Dengelenmiş Doğruluk (Balanced Accuracy)

Balanced Accuracy, recall (duyarlılık) ve specificity (özgüllük) ortalamasıdır ve dengesiz veri setlerinde doğruluğun yanılmasını engeller.

$$\text{Dengelenmiş Doğruluk} = \frac{1}{2}(\text{Duyarlılık} + \text{Özgüllük}) = \frac{1}{2} \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right) \quad (15)$$

Her iki sınıf performansı eşit ağırlıkta değerlendirilmesi için kullanılmaktadır. Bu çalışma kapsamında değerlendirmeye alınan ML ve DL algoritmaları ile formüller yardımıyla amacı açıklanan başarı parametreleri üzerinden yorumlamalar yapılarak süreç ilerletilmektedir. Bu perspektifte inceleme yapıldığında çalışma kapsamında değerlendirmeye alınan algoritmalar şu şekildedir:

(i) MLP

Çok Katmanlı Algılayıcı (Multi-Layer Perceptron, MLP), temel yapay sinir ağlarından biri olarak DL metodolojisi ile verideki karmaşık örüntülerin öğrenilebilirliği amacıyla birden fazla katmanda lineer olmayan dönüşümler gerçekleştirmektedir.

MLP'nin çekirdek yapısını; giriş katmanı, en az bir gizli katman ve çıkış katmanından oluşan ardışık hücreler (nöronlar) meydana getirir; her hücre, kendisine gelen ağırlıklı girdileri toplayarak önyargı ekler ve bunu genellikle sigmoid veya ReLU gibi bir aktivasyon fonksiyonundan geçirerek sonraki katmana iletir.

Model eğitilirken, “ileri besleme” yoluyla her katman çıkışı hesaplanmaktadır, daha sonra tahmin ve gerçek değer arasındaki hata “geri yayılım” ile katmanlara doğru iletilerek ağırlıklar stokastik gradyan iniş veya türevli optimizasyon algoritmaları (Adam, RMSProp vb.) ile güncellenmektedir. Bu süreç, modelin doğrusal olmayan karar sınırları oluşturmasına imkân tanıyarak sınıflandırma veya regresyon problemlerine çözüm sunmaktadır.

(ii) LSTM

Uzun Kısa Süreli Bellek (Long Short-Term Memory, LSTM), ardışık veri ve zaman serisi problemlerinde uzun vadeli bağımlılıkları etkili bir biçimde yakalamayı amaçlayan özel bir tekrarlayan sinir ağı (RNN) mimarisidir. LSTM hücreleri, içerdikleri hücre durumu ve üç ana kapı (giriş kapısı, unutma kapısı ve çıkış kapısı) aracılığıyla bilgiyi saklar, günceller veya unuttur. Giriş kapısı, ağı hangi bilgilere odaklanacağını; unutma kapısı, geçmişteki hangi bilgilerin geçerliliğini yitirdiğini; çıkış kapısı ise hücredeki hangi verilerin bir sonraki zamana veya çıktı katmanına aktarılacağını belirlemektedir.

Her kapı, kendisine gelen girdiye ve bir önceki gizli duruma ait ağırlıklı toplamı, bir sigmoid aktivasyon fonksiyonundan geçirerek 0 ile 1 arasında bir unutmama katsayısı üretmektedir. Model eğitimi sırasında, hata geri yayılımı ile kapı parametreleri optimize edilirken, dropout ve L2 düzenlemesi gibi tekniklerle aşırı uyum riski kontrol altında tutulur.

(iii)GRU

Kapılı Tekrarlayan Birim (Gated Recurrent Unit, GRU), ardışık verilerde uzun ve kısa vadeli bağımlılıkları öğrenmek amacıyla geliştirilmiş bir RNN varyantıdır. GRU hücresi, LSTM'deki üç kapı yerine iki kapı mekanizması (güncelleme kapısı ve sıfırlama kapısı) kullanarak daha sade bir yapı sunmaktadır.

Güncelleme kapısı z_t , mevcut girdi x_t ve önceki gizli durum h_{t-1} arasındaki ilişki değerlendirerek, hücre durumunun korunma yüzdesini incelemektedir. Bu sayede ne kadar korunması gerektiğini ve ne kadar güncellenmesi gerektiğini hesaplar:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (16)$$

Sıfırlama kapısı r_t , geçmiş bilginin ne ölçüde unutulacağını ayarlayarak, gizli durumun hesaplanmasında hangi oranda önceki duruma başvurulacağını kontrol etmektedir:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (17)$$

Bu sadeleştirilmiş kapı düzeni, hem daha az parametreyle hesaplama maliyetini düşürür hem de gradyan kaybolması sorununun önüne geçmektedir.

(iv)CNN

Evrişimli Sinir Ağları (Convolutional Neural Networks, CNN), özellikle iki boyutlu veri yapıları (görüntü, spektral haritalar vb.) üzerinde yerel örüntüleri otomatik olarak öğrenmek üzere tasarlanmış DL mimarisidir. Evrişim Katmanı, girişten küçük boyutlu bir çekirdek (kernel veya filtre) üzerinden kayan pencere ile örtüşen alt bölgelere doğrusal işlemler uygulamaktadır. Aktivasyon fonksiyonu; her evrişim sonucu üzerine uygulanan lineer olmayan dönüşümde modelin kompleks veri ilişkilerini öğrenmeye sağlamaktadır. Bu kapsamda en yaygın kullanıma sahip olan aktivasyon fonksiyonu ReLU 'dur.

Havuzlama Katmanı ise yüksek düzeydeki korelasyonu vurgulamak için kullanılmaktadır. CNN, ağırlık paylaşımı ve yerel bağlantı yapısı sayesinde parametrik verimlilik sağlamaktadır; böylece büyük boyutlu girişler için bile öğrenilebilecek filtre sayısı makul düzeyde kalmaktadır. Bu sayede hem hesaplama hem de bellek açısından daha etkin, yüksek doğruluklu modeller oluşturulabilmektedir.

(v) Transformer

Transformer mimarisi, ardışık verilerde hem uzun vadeli hem de kısa vadeli ilişkilendirmeleri eş zamanlı olarak öğrenebilmek amacıyla tasarlanmış bir DL modelidir. Temel yenilik, geleneksel RNN temelli yapılar yerine “self-attention” mekanizmasını kullanarak her giriş ögesinin tüm diğer öğelerle etkileşimini doğrudan hesaplamasıdır.

Eğitim sürecinde, Transformer’lar genellikle “maskeli dil modelleme” (masked language modeling) veya “otomatik kodlayıcı-çözücü” (encoder-decoder) paradigmasıyla optimize edilir. Encoder blokları girdinin tüm bağlamını temsil vektörlerine dönüştürürken, decoder blokları bu temsilleri belirli bir çıktı dizisine dönüştürür.

(vi) Random Forest

Rastgele Orman, birden fazla karar ağacının birlikte kullanılması yoluyla tek bir ağacın aşırı uyum riskini azaltmayı amaçlamaktadır. Her bir ağaç, eğitim verisinin bootstrap yöntemiyle rastgele örneklenmiş bir alt kümesi üzerinde büyütülmektedir. Ayrıca her düğüm bölünmesinde, değerlendirilecek özellikler arasından rasgele seçilen bir alt küme üzerinden en iyi bölünme kriteri uygulanmaktadır. Böylece ağaçlar hem veri hem de özellik boyutunda çeşitlendirilerek yüksek varyanslı hataların ortalamasını almak mümkün hâle gelmektedir. Son tahmin, sınıflandırma problemlerinde ağaçların oy çokluğuna göre, regresyonda ise öngörü değerlerinin aritmetik ortalamasına göre belirlenir:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (18)$$

Burada T toplam ağaç sayısını belirtirken h_t ise t. ağacın tahmin fonksiyonunu göstermektedir. Rastgele Orman, yüksek boyutlu veri setlerin özneliklerin rasgele seçimi sayesinde korelasyonlu ağaç kümelerini önlemektedir ve güçlü bir genelleme sağlamasını hedefleyen bir makine öğrenme algoritmasıdır. Ayrıca her ağaç için out-of-bag (OOB) tahminleriyle dahili bir hata tahmini imkanı sunmaktadır ve değişken önem skorları aracılığıyla modelde en etkili öznelikleri belirlemeye olanak tanımaktadır.

(vii) ExtraTrees

Aşırı Rastgele Ağaçlar (Extremely Randomized Trees, ExtraTrees), Rastgele Orman yaklaşımının varyantı olarak geliştirilmiş bir topluluk yöntemidir. Temel farklılık, her bir karar ağacında hem veri örneklerinin tamamının (bootstrap yerine) hem de düğüm bölünmelerinde kullanılacak özelliklerin yanı sıra bölünme eşiklerinin de rastgele seçilmesiyle yüksek çeşitlilik yaratmaktır.

(viii) Gradyan Artırma (Gradient Boosting)

Gradyan Artırma, zayıf öğrencilerden (genellikle sığ karar ağaçları) oluşan ardışık bir topluluk (ensemble) yöntemi olup, her yeni ağaç eklemesinde önceki modelin hatalarını düzeltmeye odaklanarak güçlü bir tahmin modeli oluşturmayı hedeflemektedir. Temel fikir, hedef fonksiyonun kümülatif kaybını minimize edecek yönde her adımda bir ağaç tahminini “gradyan” (kaybın türevi) bilgisi kullanarak eklemektir.

Gradyan Artırma'nın avantajı, her ek adımda kaybı doğrudan hedefleyerek güçlü bir optimizasyon sağlaması, hiper parametrelerle (derinlik, yaprak sayısı, öğrenme hızı) ince ayar yapıldığında aşırı uyum kontrolü ve genelleme performansında yüksek başarı sunmasıdır. Dezavantajı ise hesaplama maliyetinin, özellikle çok sayıda ağaç ve derin ağaç yapılarıyla artmasıdır. Bu yöntemin yaygın uygulamaları arasında kredi riski tahmini, müşteri segmentasyonu ve çeşitli regresyon problemlerinin yanında sınıflandırma görevleri de yer alır.

(ix) Histogram Tabanlı Gradyan Artırma (HistGradientBoosting)

Histogram tabanlı gradyan artırma, klasik yönteminin veri yoğunluğu ve hesaplama verimliliği açısından optimize edilmiş şeklidir. Bu yaklaşımda sürekli değişkenler, belirli aralıklara ayrılarak histogram temelli özetler oluşturularak her ağaç bölünme kararında kullanılır. Böylece her düğümde tüm gözlemler yerine histogram istatistikleri üzerinden en uygun bölünme eşiği hızla belirlenebilir. Histogram tabanlı bu strateji, veri setlerinde hem bellek kullanımını hem de işlem süresini kritik seviyelerde azaltırken, bölünme kalitesini korumaktadır. Ayrıca bu yöntem mimarisi, XGBoost ve LGBM gibi kütüphanelerin temelinde olan hız optimizasyonlarını oluşturmaktadır.

(x) LightGBM (Light Gradient Boosting Machine, LGBM)

LGBM, Microsoft tarafından geliştirilen ve gradyan artırma çerçevesi temelli bir topluluk yöntemidir. Öncelikle LGBM, sürekli özellikleri ön işleme adımı olarak histogram tabanlı iki önemli iyileştirme ile discretize eder: bucket (bin) oluşturma ve histogram özetleme. Bu sayede her bölünme kararında tüm bireysel örnek yerine her bin içindeki örnek sayısı ile gradyan ve Hessian toplamalarına bakılır; bu sayede hem hesaplama hem de bellek maliyetini kritik biçimde düşürür. İkinci olarak, LGBM yaprak-öncelikli büyüme stratejisi kullanmaktadır. Geleneksel derinlik-öncelikli yöntemlerde her ağaç seviyesi eşit olarak genişletilirken, LGBM her adımda en yüksek kazanç getirecek yaprağı seçip derinleştirmektedir. Bu sayede daha düşük ağaç sayısı ve daha sık yapılarla bile güçlü öğrenme yeteneği elde edilir; ancak dengelemek için ağaç derinliği veya yaprak başına minimum örnek sayısı gibi düzenleme parametrelerine ihtiyaç duyulmaktadır.

(xi) XGBoost (Extreme Gradient Boosting)

XGBoost, gradyan artırma yöntemini ölçeklenebilirlik ve hız odaklı optimizasyonlarla güçlendiren açık kaynaklı bir kütüphanedir. Temel olarak ağaç temelli zayıf öğrenicileri ardışık olarak inşa ederken her adımda kayıp fonksiyonunun negatif gradyanını hedefleyerek klasik yöntemlere göre üç ana yenilik getirmektedir.

Düzenleme yöntemi ile ağaç yapısına eklenen L1 (Lasso) ve L2 (Ridge) ceza terimleri sayesinde modelin karmaşıklığını doğrudan kontrol etmektedir. Ağacın yaprak sayısı ve ağırlık büyüklükleri düzenleme terimlerine dahil edilerek, aşırı uyum (overfitting) riski minimize edilmektedir.

Yaklaşık bölünme optimizasyonu ile büyük veri setlerinde tüm olası bölünme eşliğini denemek yerine, histogram tabanlı gruplayıcı (quantile sketch) algoritması kullanarak sürekli özellikler belirli sayıda bucket'a ayrılır. Böylece her düğümde en iyi bölünme hızlıca tahmin edilir, eğitim hızı önemli ölçüde artmaktadır.

Sistem mimarisi ve dağıtık eğitim ile bellek ve hesaplama kaynaklarını verimli kullanmak için veriyi bloklara bölerek ve satır- ve sütun- bazlı iş parçacığı dağıtımını yapmaktadır.

Eğitim süreci, önce sabit bir başlangıç tahminiyle (örneğin hedefin log-odds'ı) başlar; ardından her iterasyonda mevcut model hatalarına karşılık gelen ikinci dereceden (Newton) yaklaşımlar içeren hedef fonksiyonun gradyan ve Hessian değerleri kullanılarak yeni bir ağaç inşa edilir. Her ağacın çıktıları, belirlenen öğrenme oranı (learning rate) ile çarpılıp mevcut modele eklenir:

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x) \quad (19)$$

Burada η değeri öğrenme oranını, h_m ise m'inci ağacın tahmin fonksiyonunu temsil etmektedir. Model içerisinde yer alan erken durdurma (early stopping), sütun ve satır örnekleme oranları (subsample, colsample_bytree) gibi çok sayıda hiperparametre seçeneği hem genelleme hem de hesaplama verimliliği açısından hassas ayarlamalar yapılabilmesini sağlamaktadır.

(xii) CatBoost

CatBoost, Gradient Boosting çerçevesi içinde özel olarak kategorik değişkenlerin etkili bir şekilde işlenmesi ve kayıp fonksiyonunun düzenlenmesi amacıyla geliştirilmiştir. Diğer ağaç tabanlı gradyan artırma kütüphanelerinden ayrılarak “Ordered Boosting” adı verilen bir teknik kullanılmaktadır. Kategorik sütunlar için, her kategoriye ait istatistiksel özetleri (ortalama hedef değeri, sıralı permütasyonlar vb.) hesaplayan ve bu temsilleri sayısal ağaç bölünmelerine besleyen algoritmalar hem bellek kullanımı hem de bilgi kaybı açısından optimize edilmiştir.

CatBoost ağaçları simetrik yapıdadır. Her düğüm derinlik adımında aynı sayıda bölünme gerçekleştirerek aşırı uyum riskini azaltır ve hesaplama verimliliğini iyileştirir. Modelin öğrenme süreci, ikinci türev bilgisi (Hessian) kullanılarak Newton optimizasyonuna yakınsarken, L2 düzenleme ve öğrenme oranı parametreleri, aşırı öğrenmeyi kontrol altına alacak şekilde yapılandırılabilir. Ayrıca, CatBoost varsayılan hiper parametrelerle bile yüksek performans sergileyecek biçimde tasarlandığından, hiper parametre optimizasyonu sürecini kısaltır. Sonuç olarak hem sürekli hem de kategorik değişkenlerin yoğun olduğu veri setlerinde yüksek doğruluk, dengeli genelleme ve hızlı eğitim sağlayan CatBoost, sınıflandırma ve regresyon problemlerinin yanı sıra sıralama (ranking) ve çok sınıflı görevlerde de yaygın olarak tercih edilmektedir.

(xiii) Lojistik Regresyon (Logistic Regression)

Lojistik Regresyon, ikili (binary) sınıflandırma problemlerinde kullanılan temel ve interpretasyonu kolay bir ML algoritmasıdır. Model, doğrusal regresyonda olduğu gibi giriş özelliklerinin ağırlıklı toplamını hesaplar ancak çıktıyı bir eşikleme işlemine tabi tutmak yerine, bu toplamı lojistik (sigmoid) fonksiyondan geçirerek $[0, 1]$ aralığında bir olasılık değeri üretir:

$$P(y = 1|x) = \sigma(w^T x + b) = \frac{1}{1 + e^{-(w^T x + b)}} \quad (20)$$

Burada x gözlem vektörü, w ağırlık vektörü ve b sapma terimidir. Modelin parametreleri, maksimum olabilirlik ilkesiyle, eğitim verisindeki gerçek sınıf etiketlerine karşılık gelen olasılıkların logaritmik toplamını maksimize edecek şekilde optimize edilmektedir.

(xiv) KNN (En Yakın Komşu)

KNN, denetimli öğrenme kapsamında hem sınıflandırma hem de regresyon amaçlı kullanılabilen basit ve sezgisel bir algoritmadır. Model, eğitim verisindeki her bir örneği çok boyutlu bir uzayda nokta olarak ele alır ve yeni bir gözlem geldiğinde, önceden tanımlı komşu sayısı K kadar en yakın noktayı belirlemektedir. KNN'nin en büyük avantajı eğitim aşamasında hiçbir model parametresi öğrenmeye gerek kalmadan sadece veri kümesini saklamasıdır. Böylece bütün yük çıkarım (inference) sürecine kaymaktadır. Bu durum, küçük veri setlerinde çok başarılı sonuçlar verebilmektedir.

(xv) Naive Bayes

Bayes teoremi temelinde çalışan ve her bir özelliğin hedef sınıftan bağımsız olduğu kabulüne dayanan basit ancak etkili bir denetimli öğrenme yöntemidir. Modelin temel amacı, bir örneğin hangi sınıfa ait olma olasılığının maksimum olduğunu hesaplamaktır.

Bu yöntemin trafo arızası tahmininde kullanılması durumunda, her bir meteorolojik ve işletme değişkeninin (ör. sıcaklık, yağış, yük değeri) arıza var/yok sınıfı bağlamında koşullu olasılık dağılımları tahmin edilerek, yeni gözlemler için arıza riski olasılığı hızlıca hesaplanabilmektedir.

Modellerin normalizasyonu sonrasında, sınıf dengesizliğini çözmek için temel durumda `class_weight` fonksiyonu, SMOTE ve Undersampling metodolojileri çalıştırılmıştır.

- SMOTE (Synthetic Minority Over-sampling Technique)

SMOTE fonksiyonu, dengesiz sınıflı veri kümelerinde azınlık sınıfının örnek sayısını artırmak için geliştirilmiş bir yeniden örnekleme yöntemidir. Temel fikir, azınlık sınıfındaki gözlemler arasındaki lineer interpolasyonla yeni, sentetik örnekler oluşturmaktır. Komşu seçimi esnasında Öklid mesafesi temel alınarak seçilen rastgele komşulardan rastgele bir tanesi ile orijinal örnek arasındaki fark vektörünün hesaplanmasını sağlamaktadır.

Bu vektör, 0 ile 1 arasında rastgele bir katsayı ile çarpılarak orijinal örneğe eklenir:

$$x_{new} = x_i + \delta (x_{NN} - x_i), \delta \sim U(0,1) \quad (21)$$

Burada x_i , orijinal azınlık örneği, x_{NN} ise seçilen komşu ve δ rastgele ağırlık katsayısıdır.

- Alt Örnekleme (Undersampling)

Undersampling fonksiyonu, dengesiz sınıflı veri kümelerinde çoğunluk sınıfının örnek sayısını azaltarak her iki sınıf arası dengeyi sağlamayı amaçlayan bir yeniden örnekleme yöntemidir. Bu teknik, azınlık sınıfının bilgisini korurken çoğunluk sınıfından rastgele veya stratejik olarak gözlemleri çıkarmaktadır; böylece model, baskın sınıfın aşırı temsilinin neden olduğu yanlış tahminlerden kaçınmaktadır. Bu yaklaşımın en olumsuz etkisi, değerli örneklerin kaybolma riski taşımasıdır ve modelin genelleme gücünü olumsuz etkileyebilmesidir.

- Sınıf Dengelemesi (class_weight fonksiyonu)

class_weight fonksiyonu, denetimli öğrenme algoritmalarında sınıf dengesizliğinin yol açtığı önyargıyı gidermek için kullanılan bir parametredir. Modelin kayıp fonksiyonuna her bir örneğin katkısını, ait olduğu sınıfın frekansıyla ters orantılı ağırlıklandırarak verir.

$$w_c = \frac{N}{K \times N_c} \quad (22)$$

Burada w_c , c sınıfının ağırlığı; N, toplam örnek sayısını; K, sınıf sayısını; N_c ise c sınıfındaki örnek sayısını temsil etmektedir. Eğitim sırasında kayıp fonksiyonu her örnek için bu w_c değeri ile çarpılarak sınıf dengeseizliklerinin etkisi azaltılır.

Tanımlanan sınıf dengesizliği fonksiyonları, çalışma kapsamında değerlendirilmeye alınan yapay zeka modellerinde yer alması planlanan farklı eğitim setlerinde uygulamaya tabi tutulmuştur.

ML ve DL modeller arasında yer alan durumlar için dört farklı eğitim setinde dört farklı strateji ile çalışmalar yapılmıştır. Eğitim setleri; “Bayraklı MinMax Eğitim”, “Bayraklı Robust Eğitim”, “Outliersiz MinMax Eğitim”, “Outliersiz Robust Eğitim” ve stratejiler ise; “SMOTE”, “Undersampling”, “Classweight” ve “Ham veri” olarak adlandırılmaktadır.

Başlangıç olarak DL modellerde “sequence” yapısı çalıştırılmıştır. “Sequence”, Keras’ın en temel model “container”ıdır. İngilizce “sequence” kelimesinin Türkçesi “ardışık dizi” ya da “sıralı” anlamına gelmektedir. Model mimarisindeki katmanların birbiri ardına, birbirinden bağımsız dallanma olmadan, lineer bir akış izlemesidir. Bir veri setindeki gözlemlerin veya özellik vektörlerinin zamana veya belirli bir düzen ilişkisine göre sıralanmış hâlidir; her bir adım, önceki adımlarla doğal bir ardılık bağı içerir. Bu nedenle, LSTM veya Transformer gibi ardışık veriye özel mimarilerde “Sequence” yapısı hem uzun süreli bağıntıların hem de kısa dönemli dinamiklerin eş zamanlı öğrenilmesine imkân tanır. Ardışık verinin modele doğrudan beslenmesi, klasik düzleştirilmiş yaklaşımların kaybettiği bilgi bütünlüğünü koruyarak tahmin doğruluğunu artırır ve zamansal tutarlılığı gözetir. “Sequence” kurgusunda döngü, mini-küme boyutu ve pencere uzunluğu tanımlamalarının yapılması gerekmektedir.

Bir eğitim setinin tamamının modele bir kez sunulup ağırlıkların güncellendiği eğitim adımına “epoch” denir. Epoch sayısının artması modelin veriyi tekrar tekrar görerek öğrenmesini derinleştirir; ancak aşırı uyum (overfitting) riskini de beraberinde getirebilir.

Modelin her ağırlık güncelleme adımında kaç örneği aynı anda işlediğini belirleyen parametreye “batch size” denir. Türkçede “küme boyutu” veya “alt-küme büyüklüğü” olarak adlandırılabilir. Örneğin batch size = 32, her seferinde 32 örnek üzerinden hesaplamalar yaparak gradyan inişi yöntemini (gradient descent) uygular. Akademik olarak, küçük batch’ler daha gürbüz (noisy) güncellemeler getirip genelleme kabiliyetini artırırken, büyük batch’ler daha stabil ancak yerel minima tuzaklarına yatkın olabilir.

Özellikle ardışık (zaman serisi) verilerde geçmiş adımların kaç tanesinin modele girdi olarak verileceğini belirleyen parametredir. Türkçede “pencere” veya “kaydırma penceresi uzunluğu” olarak geçer. Örneğin window = 7, modelin her tahminde son yedi zaman adımını göz önünde bulunduracağı anlamına gelir. Akademik açıdan, uygun pencere uzunluğu seçimi hem kısa dönemli dalgalanmaları hem de uzun dönemli eğilimleri yakalayarak tahmin performansını optimize eder.

Epoch değerinin 50, batch_size değerinin 32 ve window değerinin 7 olduğu (haftalık ve mevsimselliğin belirginleşmesi için 7 seçilmiştir.) senaryoda “sequence” çalışması yapılmıştır.

MLP modeli için, ReLU aktivasyon fonksiyonuna sahip iki gizli katman belirtilmiştir. Birinci gizli katman 64 nöronlu yapıya sahipken ikinci gizli katman 32 nöronlu yapıya sahiptir. Dropout değeri 0.3 verilerek her eğitim adımında %30 birim rastgele “kapatarak” (dropout) aşırı uyum riskini azaltılması hedeflenmiştir. Çıkış katmanında; hedeflenen çalışma için 0-1 tahminlemesi yapılması gerekmektedir. Bu nedenle tek nöronlu “sigmoid” aktivasyon fonksiyonu kullanılmıştır. “Sigmoid” fonksiyonu daha çok ikili sınıflandırma kullanılmaktadır.

LSTM ve GRU modelleri için aynı yapı tasarlanmıştır. Girdi katmanında hedef sütun (“kesinti”) haricindeki 59 özellik alınmış pencere değerinin de 7 olması sebebiyle (7,59) girişi sağlanmıştır. return_sequences= False satırı ile sadece son zaman adımının çıktısının, yani tek bir 64 boyutlu vektör döndürülmesi sağlanmıştır. Dropout değeri 0.3 verilerek her eğitim adımında %30 birim rastgele “kapatarak” (dropout) aşırı uyum riskini azaltılması hedeflenmiştir. Çıkış katmanından “sigmoid” aktivasyon fonksiyonu kullanılmıştır.

CNN modeli için; LSTM ve GRU da yapılan (7,59) girişi sağlanmıştır. CNN modeli içerisinde tanımlanması gereken 1D konvolüsyon katmanı, 32 adet filtre ve çekirdek uzunluğu 3 olarak belirlenmiştir. ReLU fonksiyonu ve çekirdek uzunluğu sayesinde 3'erli pencerelerde de özelliklerin değerlendirilmesini sağlamaktadır. Maksimum Havuzlama Katmanı vasıtasıyla en baskın özellikleri koruyarak özet çıkarılması ve hesap yükünün azaltılması hedeflenmiştir. Düzleştirme (Flatten) Katmanı ile Konvolüsyon ve havuzlama çıktılarını tek boyutlu vektöre çevirir. Böylece tam bağlı katman girişi için düz bir vektör elde edilir. Çıkış katmanı diğer DL modellerde olduğu gibi "sigmoid" aktivasyon fonksiyonu kullanılmıştır.

Transformer Model, temel seviyede oluşturulmuş olup öznitelik gömme (embedding) katmanı için diğer DL modellerde gibi bir giriş oluşturulmuştur. Girişte self-attention mekanizmasıyla farklı başlıklar üzerinden öznitelik etkileşimleri öğrenmektedir. "Besleme İleri (Feed-Forward) Ağı" iki katmanlı küçük bir MLP ile her pozisyon için ayrı ayrı öğrenme yapacağı şekilde kutgulanmıştır. Çıkış katmanı olarak diğer modellerde olduğu gibi tek nöronlu, sigmoid aktivasyonlu katman, ikili sınıflandırma problemlerinde 0–1 arasında olasılık üretir.

20 farklı test arasında değerlendirme yapıldığında G-Mean parametresinin 0.3 üzeri (0.38) olarak çıktısı bulunan yalnızca bir model oluşturulmuştur. "bayrakli_minmax" eğitim seti ile çalıştırılan GRU modeli, en yüksek değerde olmasına rağmen sektörel anlamda yeterli seviyede olmadığı tespit edilmiştir.

Mevcut kapsamda yapılan çalışmada zamana bağlı k-fold çalışması olarak TSS uygulaması da kullanılmıştır.

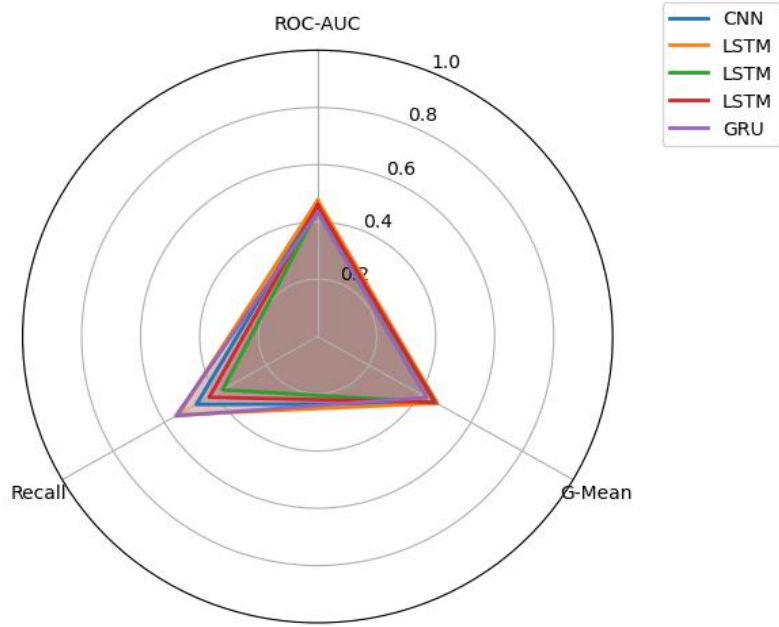
Tablo 3 Fold'ların Eğitim - Test Tarihleri

Fold	Train_Start	Train_End	Test_Start	Test_End
1	1.01.2021	27.03.2021	27.03.2021	21.06.2021
2	1.01.2021	21.06.2021	21.06.2021	15.09.2021
3	1.01.2021	15.09.2021	15.09.2021	10.12.2021
4	1.01.2021	10.12.2021	10.12.2021	6.03.2022
5	1.01.2021	6.03.2022	6.03.2022	31.05.2022
6	1.01.2021	31.05.2022	31.05.2022	25.08.2022
7	1.01.2021	25.08.2022	25.08.2022	19.11.2022
8	1.01.2021	19.11.2022	19.11.2022	13.02.2023
9	1.01.2021	13.02.2023	13.02.2023	10.05.2023
10	1.01.2021	10.05.2023	10.05.2023	4.08.2023
11	1.01.2021	4.08.2023	4.08.2023	29.10.2023
12	1.01.2021	29.10.2023	29.10.2023	23.01.2024
13	1.01.2021	23.01.2024	23.01.2024	18.04.2024
14	1.01.2021	18.04.2024	18.04.2024	13.07.2024
15	1.01.2021	13.07.2024	13.07.2024	7.10.2024
16	1.01.2021	7.10.2024	7.10.2024	31.12.2024

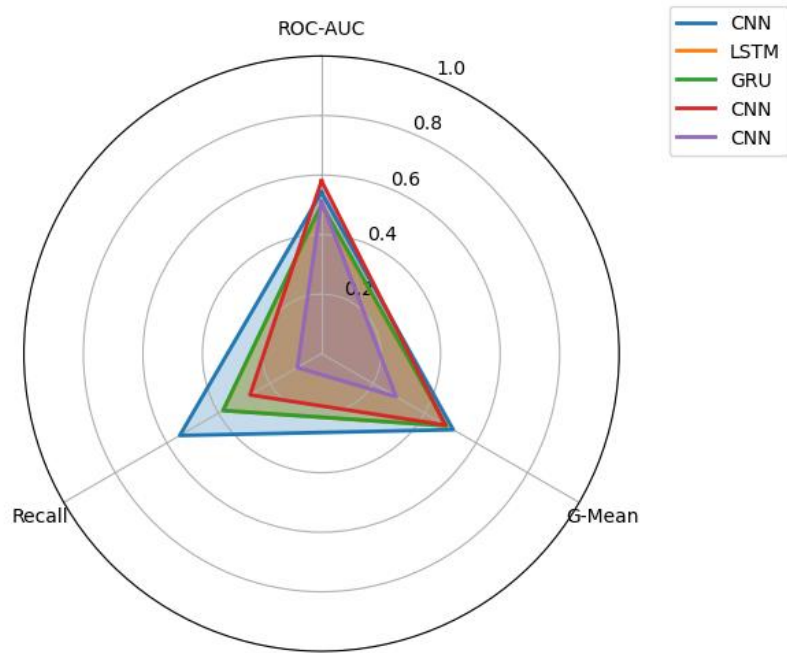
TSS, denetimli öğrenme kapsamında zaman serisi verileri için özel olarak tasarlanmış bir çapraz doğrulama (cross-validation) yöntemidir. Geleneksel k-fold yaklaşımlarının aksine, rastgele karıştırma (shuffling) yerine verinin doğal zaman sırasını koruyarak her adımda eğitim ve test bloklarını artan bir şekilde oluşturur.

- (i) Her iterasyonda eğitim verisi bir önceki adıma kıyasla genişler, model büyüyen tarihsel bağlamı da öğrenme fırsatı yakalar.
- (ii) Geri yayılım veya optimizasyon sırasında geleceğe dair bilgi kullanılmaz; böylece tahmin performansı, gerçek dünya koşullarına yakın bir şekilde validasyon edilir.
- (iii) N_splits , test bloklarının sayısını; max_train_size opsiyonu ile eğitim setinin üst sınırını belirleyerek kaydırmalı pencere (Rolling window) veya genişleyen pencere (expanding window stratejileri arasında esneklik sunar.

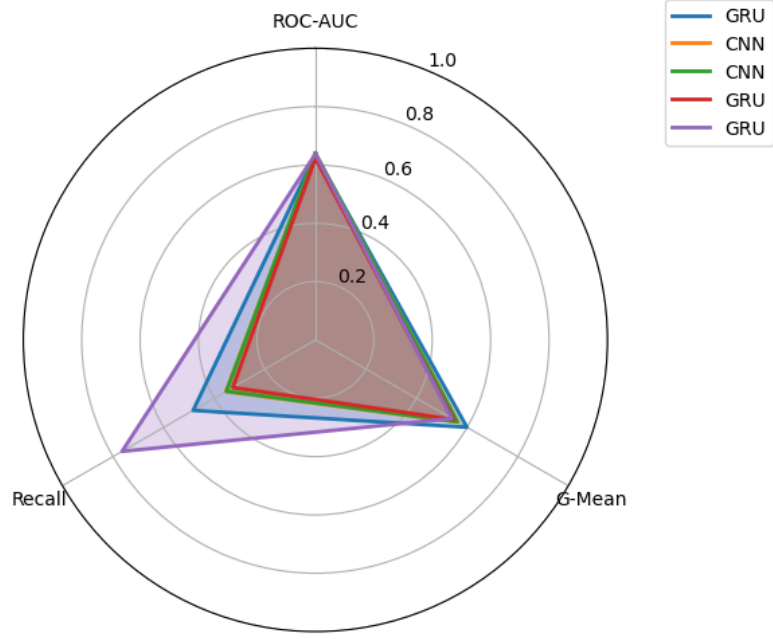
TSS uygulaması ile mevsimselliğin daha iyi ölçülmesi için 16 Fold bölünmesi yapılmıştır (Tablo-3). Bu kapsamda ise her fold kendi için değerlendirilerek en iyi performans veren algoritmaların (Şekil 28- Şekil 44) grafiği oluşturulmuştur.



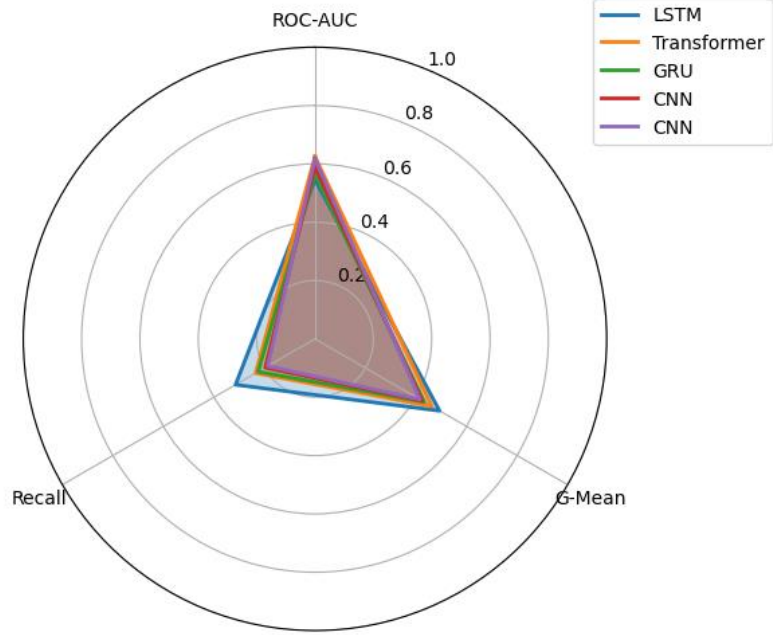
Şekil 15 G-Mean Parametresine Göre En Başarılı Modeller (Fold-1)



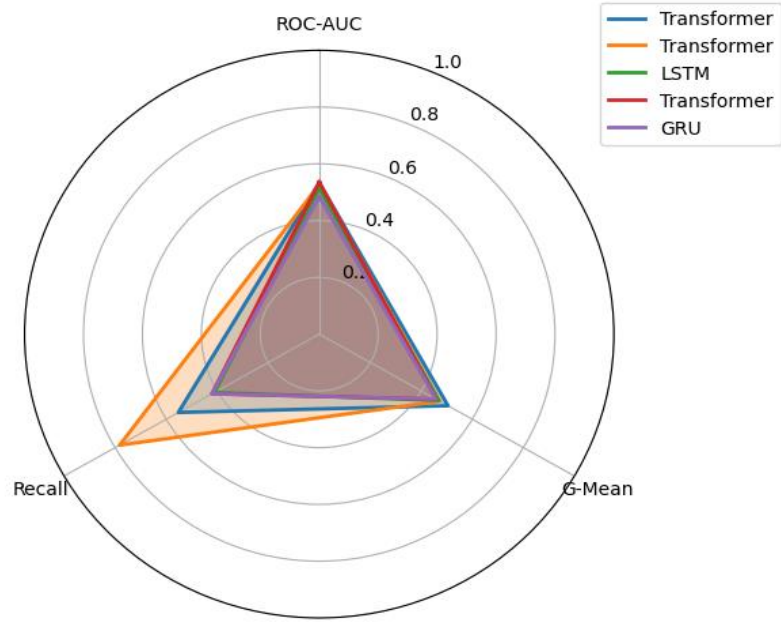
Şekil 16 G-Mean Parametresine Göre En Başarılı Modeller (Fold-2)



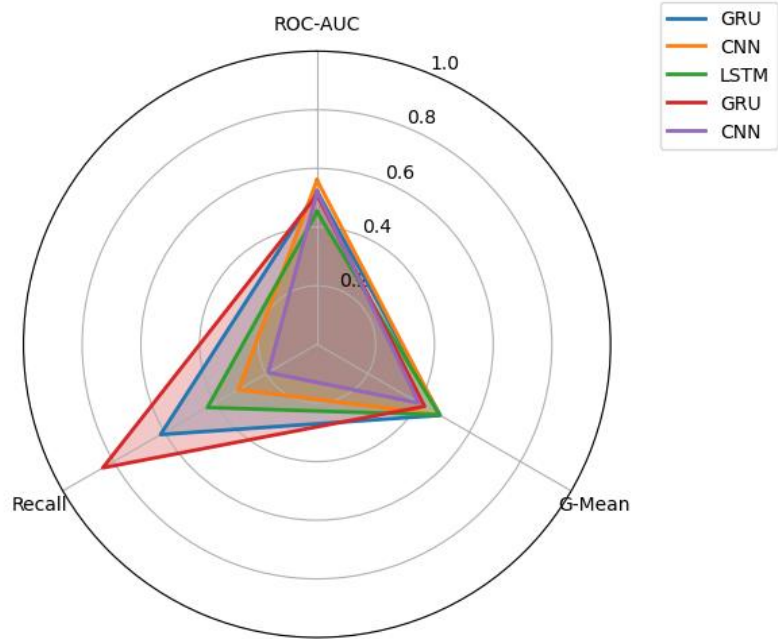
Şekil 17 G-Mean Parametresine Göre En Başarılı Modeller (Fold-3)



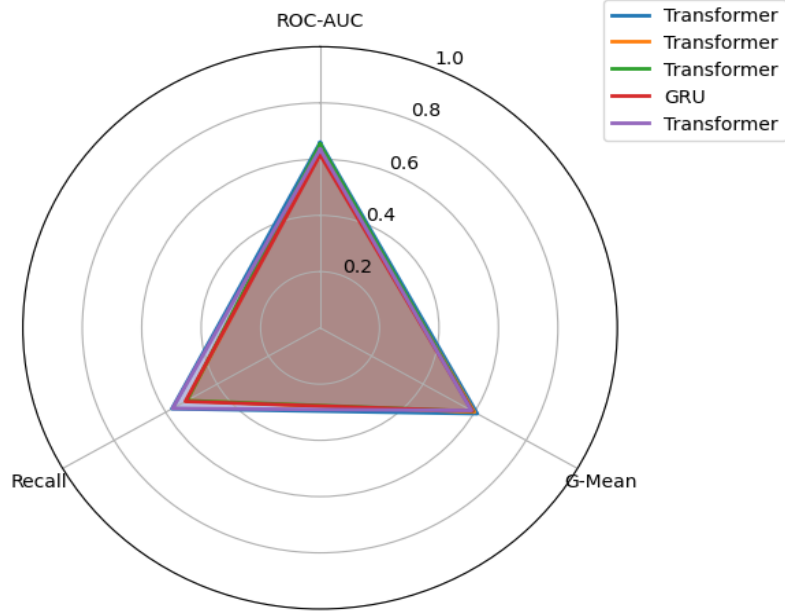
Şekil 18 G-Mean Parametresine Göre En Başarılı Modeller (Fold-4)



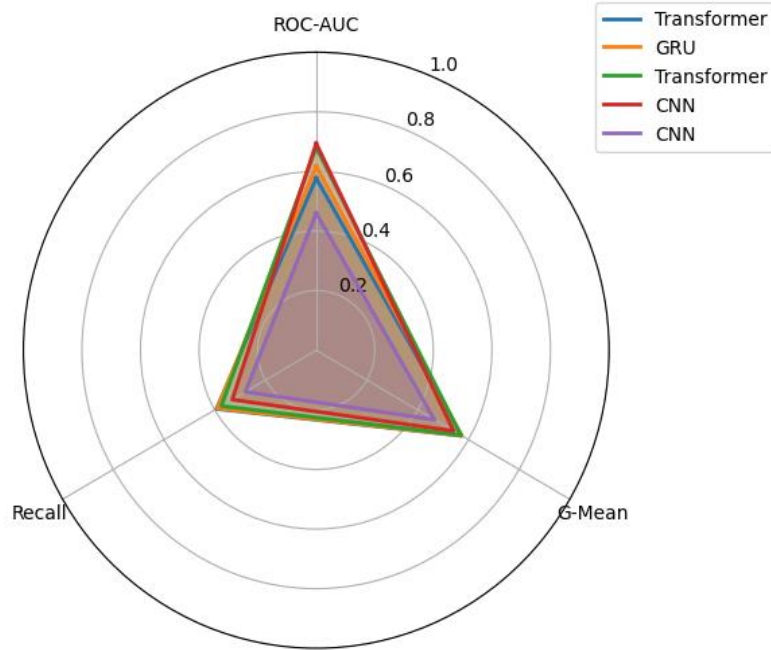
Şekil 19 G-Mean Parametresine Göre En Başarılı Modeller (Fold-5)



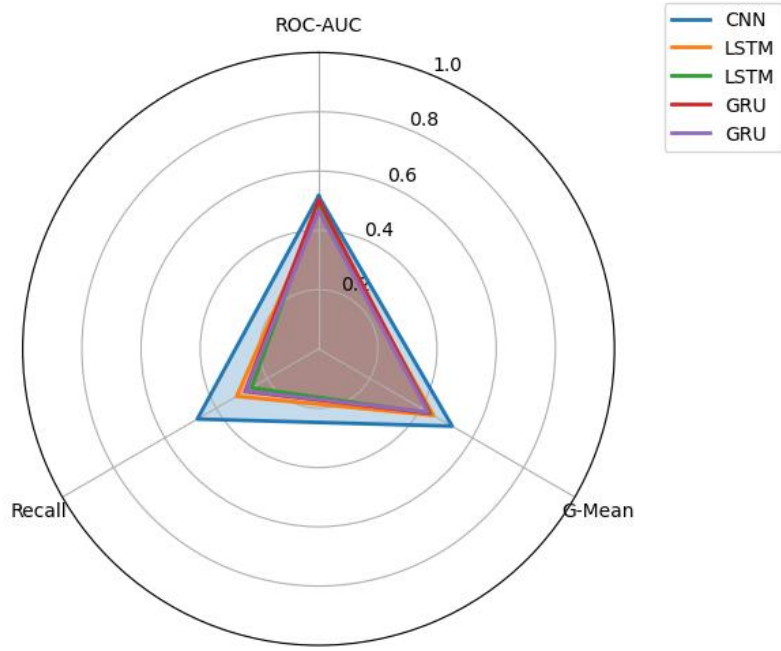
Şekil 20 G-Mean Parametresine Göre En Başarılı Modeller (Fold-6)



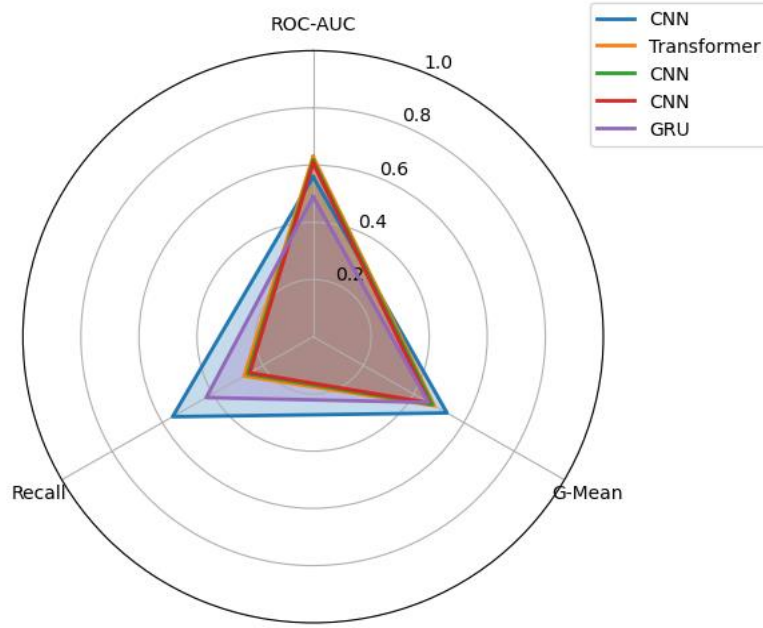
Şekil 21 G-Mean Parametresine Göre En Başarılı Modeller (Fold-7)



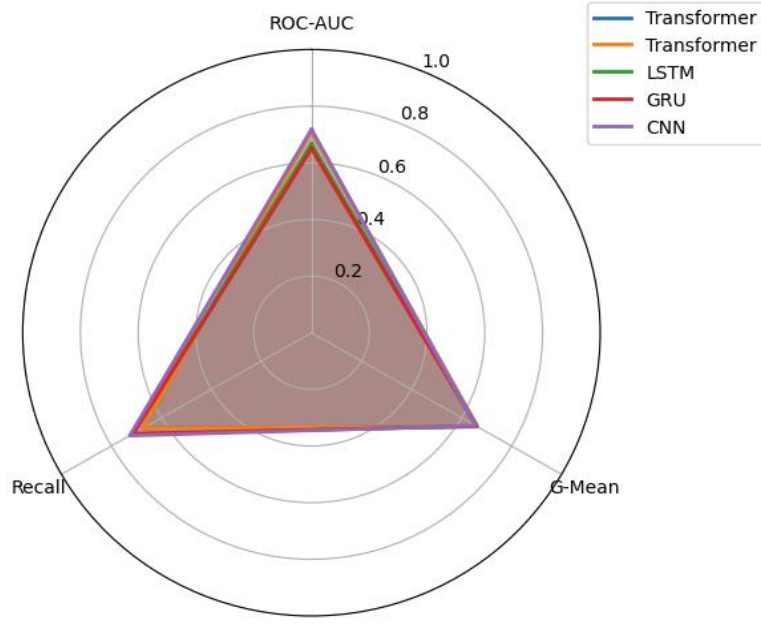
Şekil 22 G-Mean Parametresine Göre En Başarılı Modeller (Fold-8)



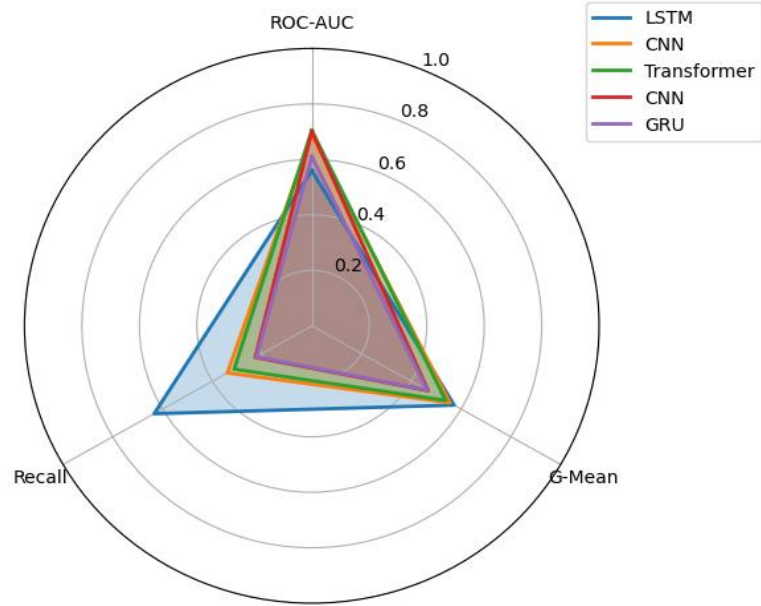
Şekil 23 G-Mean Parametresine Göre En Başarılı Modeller (Fold-9)



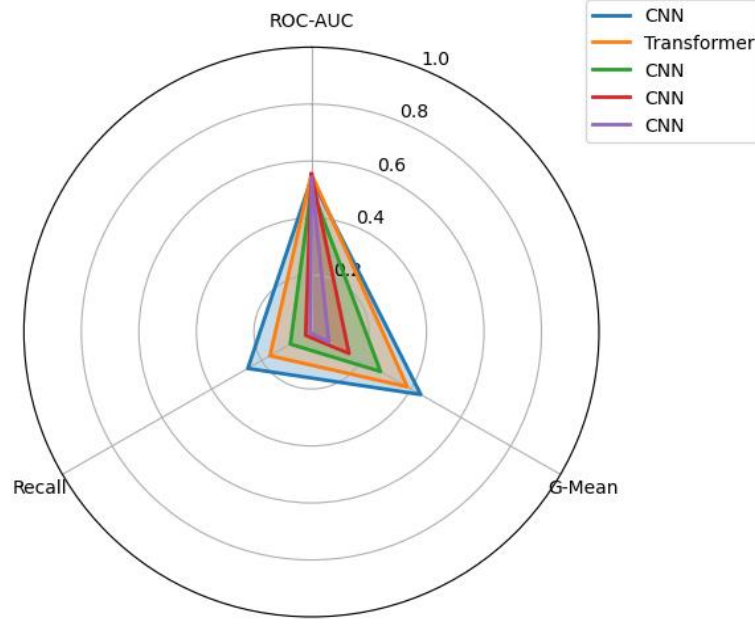
Şekil 24 G-Mean Parametresine Göre En Başarılı Modeller (Fold-10)



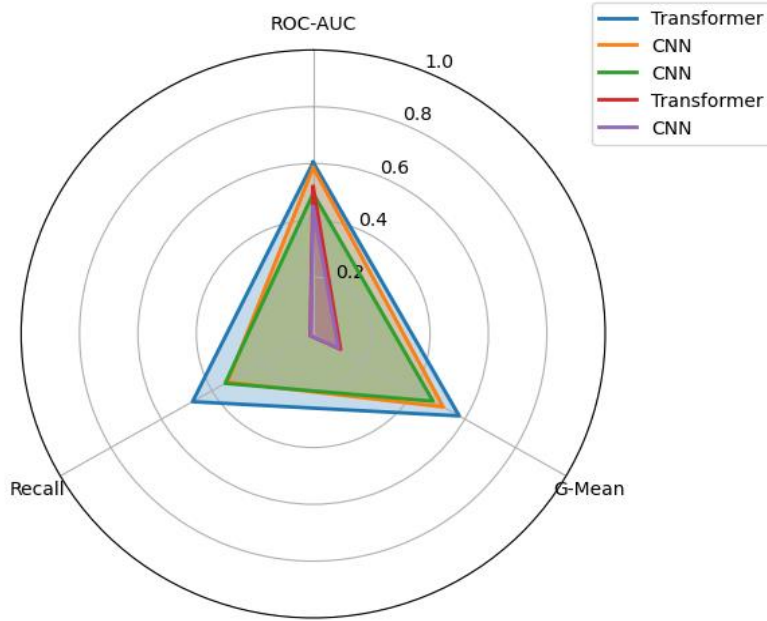
Şekil 25 G-Mean Parametresine Göre En Başarılı Modeller (Fold-11)



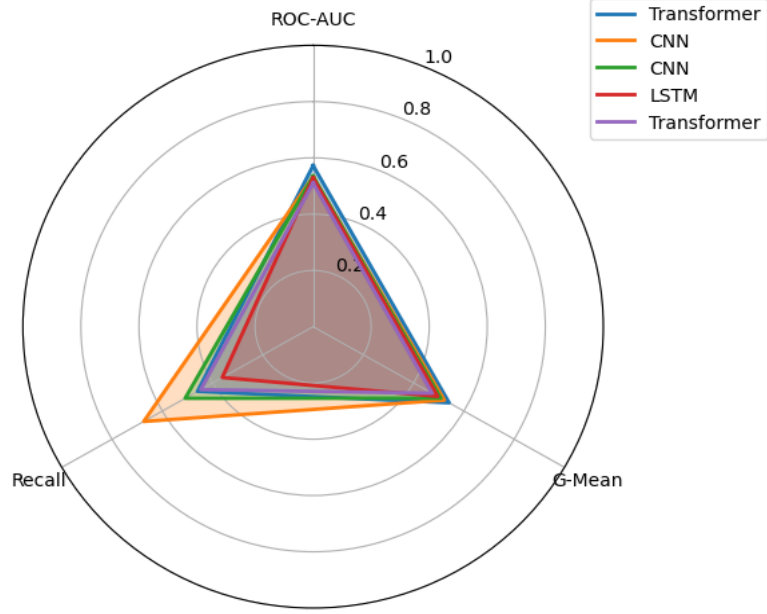
Şekil 26 G-Mean Parametresine Göre En Başarılı Modeller (Fold-12)



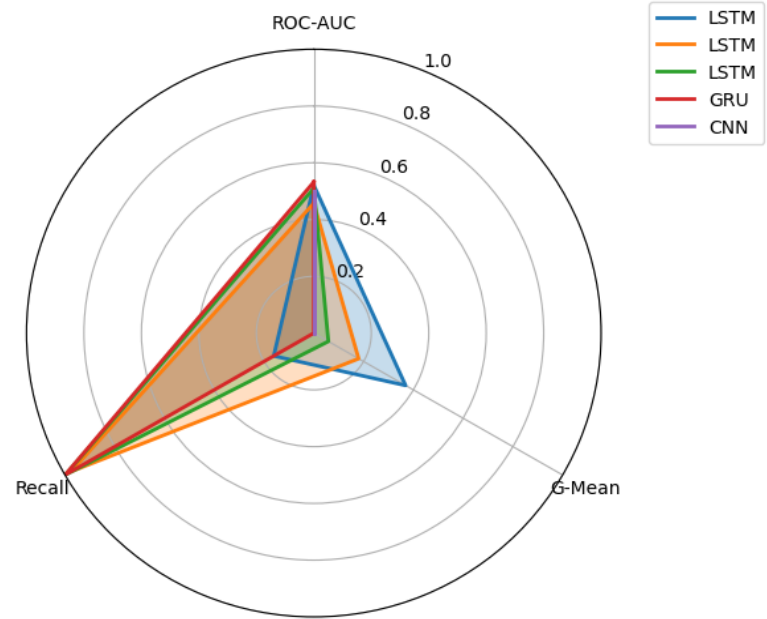
Şekil 27 G-Mean Parametresine Göre En Başarılı Modeller (Fold-13)



Şekil 28 G-Mean Parametresine Göre En Başarılı Modeller (Fold-14)



Şekil 29 G-Mean Parametresine Göre En Başarılı Modeller (Fold-15)



Şekil 30 G-Mean Parametresine Göre En Başarılı Modeller (Fold-16)

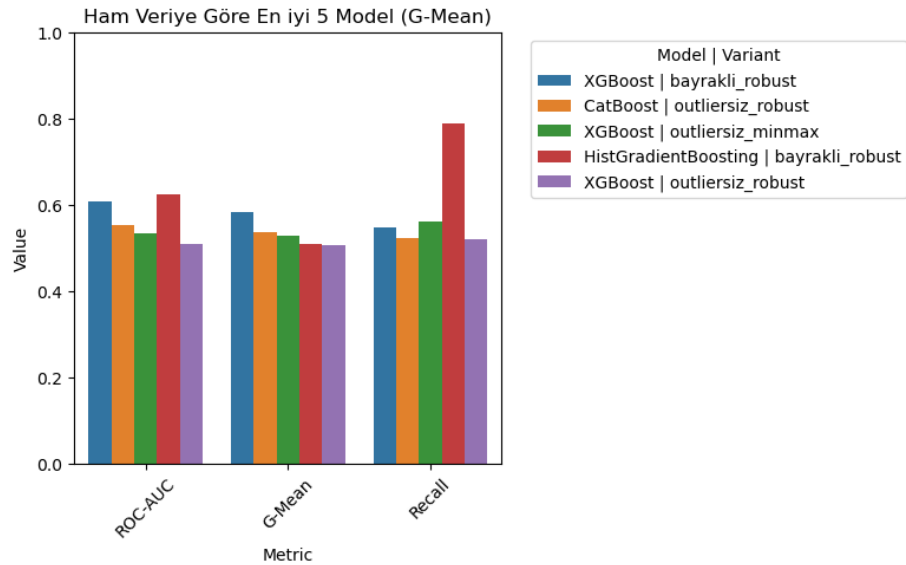
Grafiklerde inceleme yapıldığında; Fold-11'in neredeyse bütün modellerde en iyi performansı gösterdiği tespit edilmiştir (Şekil 25). Ham veri ile sequence yapısında çalışan DL modeller, G-Mean değeri olarak en yüksek 0,38 değerini gösterirken; LSTM, Transformer, GRU ve CNN modelleri Fold 11 de sırasıyla 0.66, 0.663, 0.659, 0.656 gibi G-Mean değerlerine sahip olmasıyla ön plana çıkmıştır. TSS uygulaması ile, DL modellerin zaman ile ilişkisi, tarihsel eğitimleri ve bu bağlamda gerçekleşen performans çıktılarındaki artış; G-Mean parametresine göre ilgili şekillerde belirtilmiştir.

DL modeller için yapılan çalışmalara benzer bir şekilde, ML modeller arasında yer alan durumlar için dört farklı eğitim setinde dört farklı strateji ile çalışmalar yapılmıştır. Eğitim setleri; “Bayraklı Minmax Eğitim”, “Bayraklı Robust Eğitim”, “Outliersiz Minmax Eğitim”, “Outliersiz Robust Eğitim” ve stratejiler ise; “SMOTE”, “Undersampling”, “Classweight” ve “Ham veri” olarak adlandırılmaktadır.

- Random Forest
 - Random_state = 42
- XGBoost
 - use_label_encoder=False (manuel label encoding yapıldığı için bu fonksiyon pasife çekilmiştir) , eval_metric="logloss" (Model eğitimi sırasında izlenecek değerlendirme ölçütünü tanımlar, logloss İkili sınıflandırmada lojistik kaybı (log loss) kullanır. Her iterasyonda bu metriği raporlar ve optimizasyona yardımcı olur.), random_state=42
- CatBoost
 - (verbose=0, Eğitim sırasında ne kadar bilgi basılacağını kontrol eder.), random_state=42),
- LightGBM
 - random_state=42
- KNN

- Logistic Regression
 - random_state=42, maksimum iterasyon sayısı 1000
- Naive Bayes
- Extra Trees Classifier
 - random_state=42
- HistGradientBoosting
 - random_state=42
- GradientBoosting
 - random_state=42

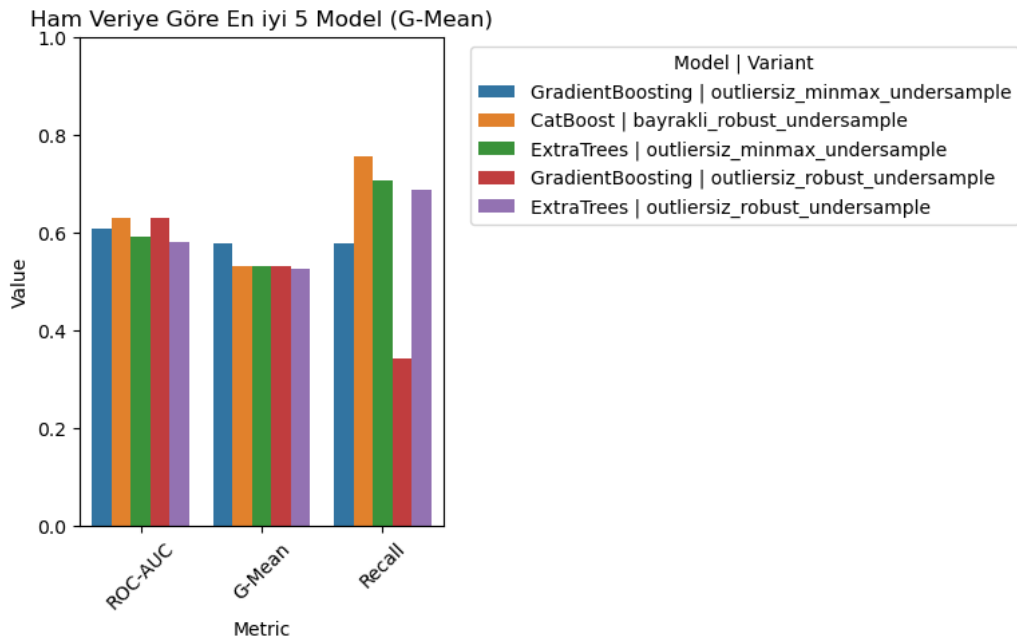
belirtilen ilgili makine öğrenmeleri tanımlı oldukları ham veri ve temel özellikleri ile teste tabi tutulduğunda 2024 yılı son çeyreği test sonuçlarına göre; XGBoost “bayrakli_robust” eğitim seti ile çalıştırılmış modelde, 0.584 G-Mean değeri, 0.61 ROC-AUC değeri ve 0.54 Recall değerine ulaşmıştır (Şekil 31).



Şekil 31 Ham Veri ile G-Mean Parametresine göre En Başarılı Modeller

Outliersiz Robust eğitim seti ile çalıştırılan CatBoost 0.538 G-Mean değeri alırken, outliersiz MinMax eğitim seti ile çalıştırılan XGBoost 0.529, bayraklı Robust eğitim seti ile çalıştırılan Hist Gradient Boosting ise 0.512 değerini almıştır.

İlgili modellerde parametre özellikleri değiştirilmeden “Random Under Sampling” fonksiyonu ile veri dengesizliği problemine odaklanılmıştır. Bu kapsamda değişen değerler ve model başarıları, Şekil 32’de gösterilmektedir.



Şekil 32 Under Sampling ile G-Mean Parametresine göre En Başarılı Modeller

İlgili sonuçlar incelendiğinde; sektörel ve akademik çalışmalarda değerlendirilmeye alınan Recall, Accuracy ve ROC-AUC gibi parametrelerin yanıltıcı olabileceği gözükmemektedir. Çünkü ilgili model eğitimleri sonrasında, model herhangi bir sınıfa ağırlık vermeye eğilim gösterirse ya da aşırı öğrenme durumu yaşanırsa ilgili değerlerde artış gözükmemektedir.

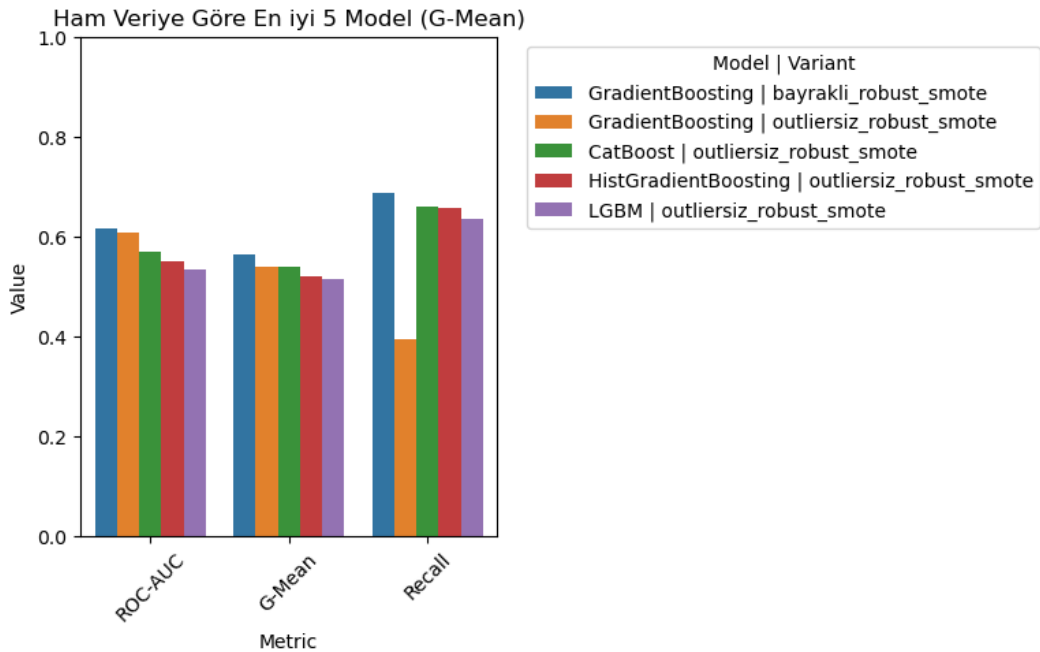
“LogisticRegression” modeli, örnek azaltımı durumundan sonra “Kesinti Var” sınıfına %100 ağırlık vermeye başlamış olup tüm tahminlerini 1 olarak sınıflandırmıştır. Böylece Recall değeri de 1 olmuştur. Farklı bir örnek verilmesi gerekirse “bayrakli_minmax” eğitim seti ile çalıştırılmış ve ham veri üzerinden performansı ölçülmüş olan LGBM modelinde 0.963 Recall değerine ulaşılmıştır.

Bu kapsamda incelemeler yapıldığında G-Mean değeri ise 14 de gösterildiği gibi geometrik ortalama almaya odaklanarak sektörel kullanıma en uygun modelin seçilmesine odaklanılmaktadır. Bu kapsamda değerlendirme yapıldığında; outliersiz_minmax eğitim seti ile çalıştırılan GradientBoosting 0.578 değeri ile ilk sırada yer almıştır. bayrakli_robust eğitim seti ile çalıştırılan CatBoost 0.533, outliersiz_minmax ile çalıştırılan Extra Trees 0.53 değerini almıştır. Tabloda ilgili modellerin Recall ve ROC- AUC değerleri de yer almaktadır.

Tablo 4 Under Sampling ile G-Mean Parametresine göre En Başarılı Modeller

Variant	Model	Recall	ROC-AUC	G-Mean
outliersiz_minmax_undersample	GradientBoosting	0.580	0.609	0.578
bayrakli_robust_undersample	CatBoost	0.758	0.630	0.533
outliersiz_minmax_undersample	ExtraTrees	0.706	0.593	0.533

Sentetik veri üretimi ile süreç ilerletildiğinde, Gradient Boosting ve Hist Gradient Boosting modelleri ön plana çıkmaktadır.



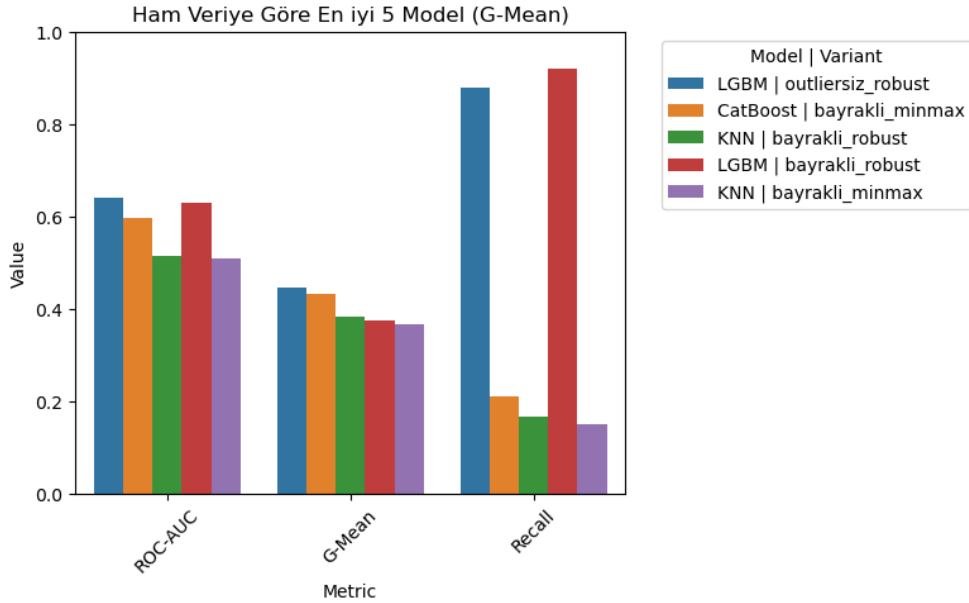
Şekil 33 SMOTE ile G-Mean Parametresine göre En Başarılı Modeller

Variant	Model	Recall	ROC-AUC	G-Mean
bayrakli_robust_smote	GradientBoosting	0.688	0.617	0.564
outliersiz_robust_smote	GradientBoosting	0.396	0.608	0.541
outliersiz_robust_smote	CatBoost	0.660	0.570	0.540
outliersiz_robust_smote	HistGradientBoosting	0.658	0.551	0.522
outliersiz_robust_smote	LGBM	0.638	0.535	0.515
bayrakli_minmax_smote	ExtraTrees	0.509	0.500	0.511

Tablo 5 SMOTE ile G-Mean Parametresine göre En Başarılı Modeller

Tablo 5 incelendiğinde outliersiz_robust eğitim setinin çalışmalara daha fazla katkı verdiği gözükmemektedir. OD sonrasında ilgili aykırı değerlerin çıkartılması işlemi yapılan bu eğitim setine sentetik veri üretimi (SMOTE) çalışması uygulandığında alınan sonuçlar 0.5 değerinin üzerindedir.

Sınıf dengesizliğinin çözümlenmesi için üçüncü senaryo olan class_weight fonksiyonu ilgili ML modellere uygulanarak çalışma yapılmıştır. Tanımlı algoritmalarda class_weight = 'balanced' fonksiyonu tanımlanarak sınıf dengesizliği çözümü sağlanmıştır. XGBoost algoritması içerisinde class_weight fonksiyonu tanımlı olmadığı için; scale_pos_weight fonksiyonu ile manuel biçimde sınıf dengesizliği ortadan kaldırılmıştır.



Şekil 34 Class_Weight ile G-Mean Parametresine göre En Başarılı Modeller

Yapılan incelemelerde G-Mean değeri diğer metodolojilere göre kritik bir şekilde değer kaybetmiştir. Ancak Recall değerinde ve ROC – AUC değerinde yer alan dengesizlik ilgili parametrelere göre optimizasyon yapılmasının fayda göstermeyeceğini net bir şekilde ifade etmektedir.

Variant	Strategy	Model	Recall	ROC-AUC	G-Mean
outliersiz_robust	clasweight	LGBM	0.881	0.643	0.446
bayrakli_minmax	clasweight	CatBoost	0.212	0.598	0.434
bayrakli_robust	clasweight	KNN	0.167	0.515	0.383
bayrakli_robust	clasweight	LGBM	0.920	0.631	0.376

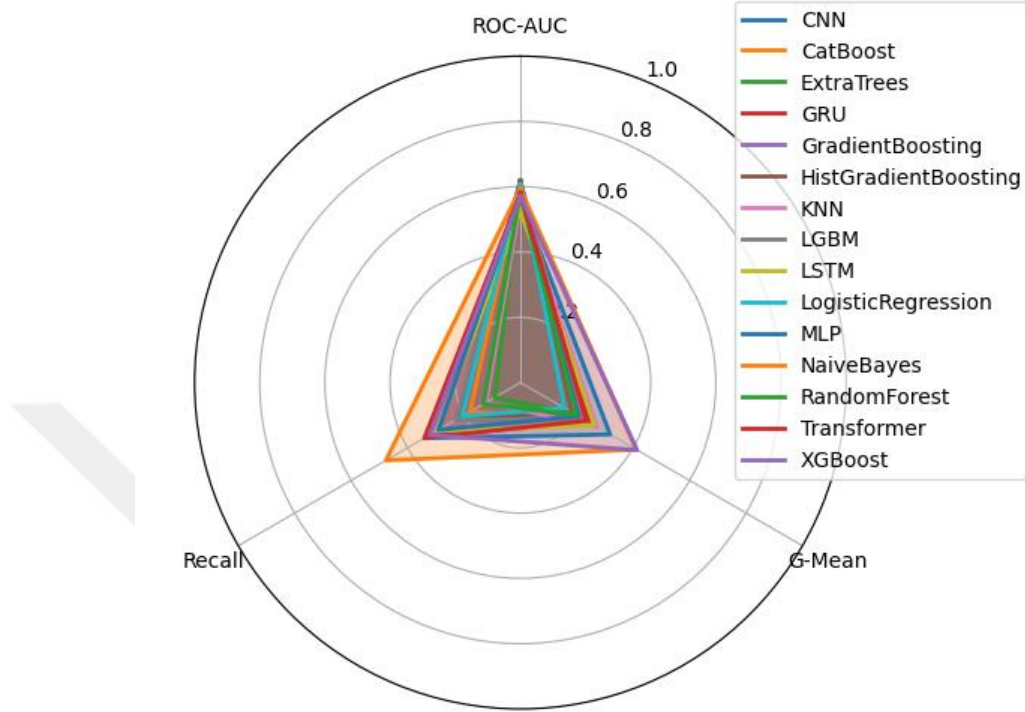
Tablo 6 Class_Weight ile G-Mean Parametresine göre En Başarılı Modeller

İlgili ML modeller ile 16 parçaya ayrılması üzerine kurulu TSS uygulaması değerlendirmeye alınmıştır. En iyi performans veren Fold ve model bulguları Tablo 7’de yer almaktadır.

Tablo 7 TSS ile G-Mean Parametresine göre En Başarılı Modeller

Fold	Model	Accuracy	Recall	ROC-AUC	G-Mean
bayrakli_minmax_fold12	NaiveBayes	0.671	0.641	0.696	0.659
bayrakli_robust_fold12	NaiveBayes	0.672	0.639	0.696	0.659
outliersiz_minmax_fold12	NaiveBayes	0.580	0.727	0.692	0.635
outliersiz_robust_fold12	NaiveBayes	0.580	0.727	0.692	0.635
bayrakli_minmax_fold05	NaiveBayes	0.607	0.642	0.653	0.621
bayrakli_robust_fold05	NaiveBayes	0.607	0.642	0.653	0.621
outliersiz_robust_fold03	MLP	0.647	0.568	0.667	0.612
outliersiz_robust_fold07	NaiveBayes	0.596	0.624	0.627	0.606
outliersiz_robust_fold05	NaiveBayes	0.569	0.660	0.647	0.605
outliersiz_minmax_fold05	NaiveBayes	0.569	0.660	0.647	0.605
outliersiz_minmax_fold07	NaiveBayes	0.595	0.622	0.627	0.605
bayrakli_minmax_fold07	NaiveBayes	0.666	0.515	0.657	0.604
bayrakli_robust_fold07	NaiveBayes	0.666	0.515	0.657	0.604
outliersiz_robust_fold07	MLP	0.640	0.543	0.659	0.602
outliersiz_minmax_fold06	NaiveBayes	0.575	0.637	0.652	0.601
outliersiz_robust_fold06	NaiveBayes	0.575	0.637	0.652	0.601

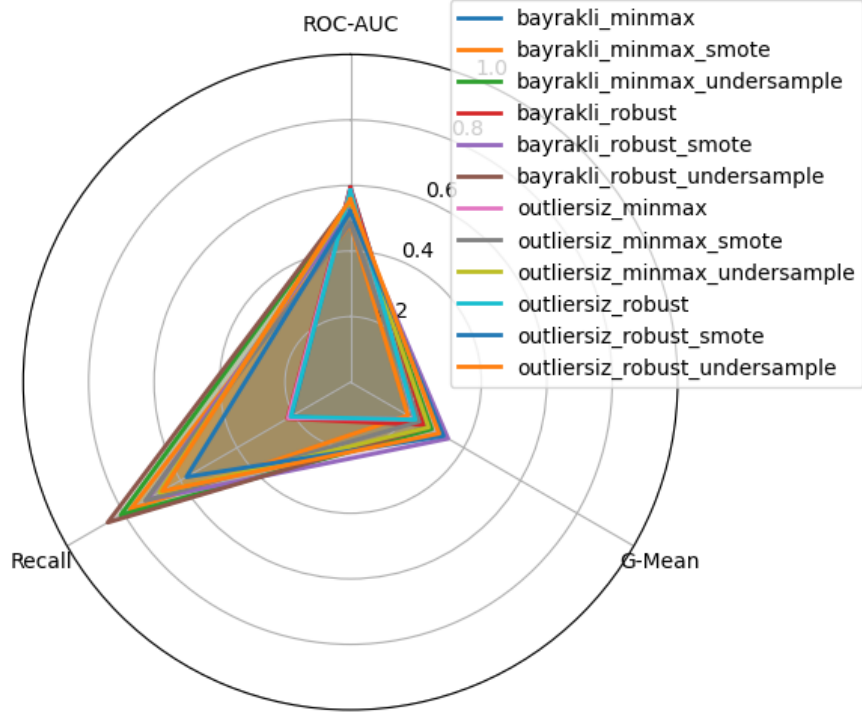
Zamana göre öğrenme konusunda, diğer ML modellerine göre NaiveBayes uygulaması önde olduğunu göstermiştir (Tablo-7). Diğer sistemlere göre daha dengeli bir Recall, G-Mean, ROC-AUC ve Accuarcy değerleri olduğu tabloda gözükmektedir.



Şekil 35 Model Bazında Ortalama Başarı Değerleri

Tablo 8 Model Bazında Ortalama Başarı Değerleri

Satır Etiketleri	Ortalama G-Mean	Ortalama ROC-AUC	Ortalama Recall
CatBoost	0.226	0.617	0.178
CNN	0.318	0.561	0.340
ExtraTrees	0.192	0.604	0.091
GradientBoosting	0.214	0.618	0.228
GRU	0.244	0.533	0.326
HistGradientBoosting	0.183	0.618	0.222
KNN	0.272	0.549	0.122
LGBM	0.226	0.608	0.228
LogisticRegression	0.156	0.606	0.207
LSTM	0.262	0.527	0.297
MLP	0.206	0.580	0.290
NaiveBayes	0.411	0.594	0.477
RandomForest	0.197	0.582	0.131
Transformer	0.234	0.579	0.339
XGBoost	0.414	0.571	0.321



Şekil 36 Varyant Bazında Ortalama Başarı Değerleri

Tablo 9 Model Bazında Ortalama Başarı Değerleri

Satır Etiketleri	Ortalama G-Mean	Ortalama ROC-AUC	Ortalama Recall
bayrakli_minmax	0.261	0.594	0.209
bayrakli_minmax_smote	0.203	0.482	0.777
bayrakli_minmax_undersample	0.287	0.551	0.811
bayrakli_robust	0.261	0.593	0.222
bayrakli_robust_smote	0.345	0.526	0.717
bayrakli_robust_undersample	0.274	0.554	0.857
outliersiz_minmax	0.227	0.585	0.220
outliersiz_minmax_smote	0.242	0.485	0.724
outliersiz_minmax_undersample	0.276	0.552	0.674
outliersiz_robust	0.229	0.583	0.211
outliersiz_robust_smote	0.329	0.521	0.579
outliersiz_robust_undersample	0.313	0.558	0.665

“Random Research” metodolojisinden daha detaylı, manuel veya diğer yöntemlere göre daha performanslı çalışan (Gradient daha detaylı ama çok yavaş) Bayesian Optimizasyon yöntemlerinden Optuna kütüphanesi ile optimizasyon çalışmaları yapılmıştır. Genel olarak dengesiz bir veri setinde model çalışmaları dolayı, en yüksek performans veren modellerde doğruluk, kesinlik ve hassasiyet gibi parametreler yerine “G-Mean” parametresi üzerinden optimizasyon yapılmıştır.

Optuna, hiper parametre optimizasyonu amacıyla geliştirilmekte olan açık kaynaklı bir Python kütüphanesi olup “define-by-run” felsefesini benimsemektedir. Bu felsefe, kullanıcıların dinamik kod yapılarını doğrudan optimizasyon sürecine dahil etmektedir. Bir çalışma kapsamında, deneme olarak adlandırılan farklı parametre konfigürasyonları değerlendirilmektedir. Örnekleyici olarak varsayılan Tree-structured Parzen Estimator (TPE) yöntemi, Bayesian yaklaşımla önceki iyi ve kötü sonuçları modelleyerek yeni öneriler üretmektedir. Erken durdurma mekanizması ise kaynak tüketimini azaltmak üzere yetersiz performans gösteren denemeleri durdurmaktadır. Paralel ve dağıtık çalışmaya uygun mimarisi, farklı optimizasyon stratejileri arasında kolay geçiş imkânı sunmakta, kapsamlı görselleştirme araçları sağlamaktadır. Tüm bu özellikler Optuna’nın hem hızlı prototipleme hem de büyük ölçekli hiper parametre araştırmaları için güçlü bir çözüm sunmakta olduğunu göstermektedir.

XGBoost için tüm eğitim setleri üzerinde optimizasyon çalışması gerçekleştirilmiştir. Bu optimizasyon çalışması kapsamında belirtilen XGBoost parametreleri aşağıda sunulmuştur:

```
params = {  
    'n_estimators': trial.suggest_int('n_estimators', 100, 1000),  
    'learning_rate': trial.suggest_loguniform('learning_rate', 1e-4, 0.5),  
    'max_depth': trial.suggest_int('max_depth', 3, 10),  
    'min_child_weight': trial.suggest_int('min_child_weight', 1, 10),  
    'gamma': trial.suggest_loguniform('gamma', 1e-8, 10.0),  
    'subsample': trial.suggest_uniform('subsample', 0.5, 1.0),  
    'colsample_bytree': trial.suggest_uniform('colsample_bytree', 0.5, 1.0),  
    'reg_alpha': trial.suggest_loguniform('reg_alpha', 1e-8, 10.0),
```

```
'reg_lambda': trial.suggest_loguniform('reg_lambda', 1e-8, 10.0),
'scale_pos_weight': scale_pos_weight,
'use_label_encoder': False,
'eval_metric': 'logloss',
'random_state': 42,
'n_jobs': -1
}
```

Değerlendirmeye alınan parametreler sonucunda tespit edilen en yüksek G-Mean değeri; “bayrakli_minmax” eğitim seti göstermiştir. 0.5917 değeri gösteren optimizasyon çalışmasında kullanılması gereken parametre girdi değerleri aşağıdaki şekilde ayarlanmıştır:

```
params = {
'n_estimators': 368,
'learning_rate': 0.06698699176361983,
'max_depth': 3,
'min_child_weight': 6,
'gamma': 0.007498548632456873,
'subsample': 0.8229916743943375,
'colsample_bytree': 0.8269533547636057,
'reg_alpha': 3.2053185128219836e-08,
'reg_lambda': 0.027726996835007782}
```

LGBM için tüm eğitim setleri üzerinde optimizasyon çalışması gerçekleştirilmiştir. Bu optimizasyon çalışması kapsamında belirtilen LGBM parametreleri belirtilmiştir:

```
params = {
'learning_rate': trial.suggest_loguniform('learning_rate', 1e-4, 0.5),
'n_estimators': trial.suggest_int('n_estimators', 100, 1000),
'num_leaves': trial.suggest_int('num_leaves', 31, 256),
'max_depth': trial.suggest_int('max_depth', 3, 12),
'min_child_weight': trial.suggest_int('min_child_weight', 1, 50),
'subsample': trial.suggest_uniform('subsample', 0.5, 1.0),
'colsample_bytree': trial.suggest_uniform('colsample_bytree', 0.5, 1.0),
```

```
'reg_alpha': trial.suggest_loguniform('reg_alpha', 1e-8, 10.0),
'reg_lambda': trial.suggest_loguniform('reg_lambda', 1e-8, 10.0),
'class_weight': 'balanced',
'random_state': 42,
'n_jobs': -1
}
```

Değerlendirmeye alınan parametreler sonucunda tespit edilen en yüksek G-Mean değeri; “bayrakli_minmax” eğitim seti göstermiştir. 0.5971 değeri gösteren optimizasyon çalışmasında kullanılması gereken parametre girdi değerleri şu şekildedir:

```
params = {
'learning_rate': 0.4651082718536084,
'n_estimators': 180,
'num_leaves': 44,
'max_depth': 3,
'min_child_weight': 50,
'subsample': 0.9975500493571117,
'colsample_bytree': 0.9966890980657296,
'reg_alpha': 0.0648781523132833,
'reg_lambda': 0.43795194972727663}
```

GradientBoostingClassifier algoritması için “Under Sampling” metodolojisi ile optimizasyon çalışması gerçekleştirilmiştir. Bu optimizasyon çalışması kapsamında belirtilen GradientBoostingClassifier parametreleri aşağıda sunulmuştur:

```
params = {
'n_estimators': trial.suggest_int('n_estimators', 100, 1000),
'learning_rate': trial.suggest_loguniform('learning_rate', 1e-3, 0.2),
'max_depth': trial.suggest_int('max_depth', 3, 10),
'min_samples_split': trial.suggest_int('min_samples_split', 2, 50),
'min_samples_leaf': trial.suggest_int('min_samples_leaf', 1, 50),
'subsample': trial.suggest_uniform('subsample', 0.5, 1.0),
'max_features': trial.suggest_uniform('max_features', 0.5, 1.0),
```

```
'random_state': 42
}
```

Değerlendirmeye alınan parametreler sonucunda tespit edilen en yüksek G-Mean değeri; “bayrakli_robust” eğitim seti göstermiştir. 0.6062 değeri gösteren optimizasyon çalışmasında kullanılması gereken parametre girdi değerleri şu şekildedir:

```
params = {
'n_estimators': 291,
'learning_rate': 0.06687125307703855,
'max_depth': 5,
'min_samples_split': 26,
'min_samples_leaf': 9,
'subsample': 0.7382300680560255,
'max_features': 0.8088198237120888}
```

Extra Trees için tüm eğitim setleri üzerinde optimizasyon çalışması gerçekleştirilmiştir. Bu optimizasyon çalışması kapsamında belirtilen Extra Trees parametreleri aşağıda belirtilmiştir:

```
params = {
'n_estimators': trial.suggest_int('n_estimators', 100, 500),
'criterion': trial.suggest_categorical('criterion', ['gini', 'entropy']),
'max_depth': trial.suggest_int('max_depth', 3, 20),
'min_samples_split': trial.suggest_int('min_samples_split', 2, 20),
'min_samples_leaf': trial.suggest_int('min_samples_leaf', 1, 20),
'max_features': trial.suggest_uniform('max_features', 0.1, 1.0),
'bootstrap': trial.suggest_categorical('bootstrap', [True, False]),
'class_weight': 'balanced',
}
```

Değerlendirmeye alınan parametreler sonucunda tespit edilen en yüksek G-Mean değeri; “bayrakli_minmax” eğitim seti göstermiştir. 0.5669 değeri gösteren

optimizasyon çalışmasında kullanılması gereken parametre girdi değerleri şu şekildedir:

```
params = {  
  'n_estimators': 483,  
  'criterion': 'gini',  
  'max_depth': 15,  
  'min_samples_split': 3,  
  'min_samples_leaf': 20,  
  'max_features': 0.20111327584022184,  
  'bootstrap': False}
```

Extra Trees için yapılan optimizasyonda, kullanılan girdi parametreleri “Under Sampling” metodu kullanıldıktan sonra tekrar optimizasyon yapılmıştır. Bu kapsamda G-Mean değeri; en yüksek değeri “bayrakli_minmax” eğitim setinde göstermiştir. 0.5891 değeri gösteren optimizasyon çalışmasında kullanılması gereken parametre girdi değerleri şu şekildedir:

```
Params = {  
  'n_estimators': 112,  
  'criterion': 'gini',  
  'max_depth': 20,  
  'min_samples_split': 5,  
  'min_samples_leaf': 16,  
  'max_features': 0.6139340470553416,  
  'bootstrap': True}
```

Random Forest için tüm eğitim setleri üzerinde optimizasyon çalışması gerçekleştirilmiştir. Bu optimizasyon çalışması kapsamında belirtilen Random Forest parametreleri şu şekildedir:

```
params = {  
  'n_estimators': trial.suggest_int('n_estimators', 100, 500),  
  'max_depth': trial.suggest_int('max_depth', 3, 20),  
  'min_samples_split': trial.suggest_int('min_samples_split', 2, 20),
```

```
'min_samples_leaf': trial.suggest_int('min_samples_leaf', 1, 20),
'max_features': trial.suggest_uniform('max_features', 0.1, 1.0),
'bootstrap': trial.suggest_categorical('bootstrap', [True, False]),
'class_weight': 'balanced',
'random_state': 42,
'n_jobs': -1
}
```

Değerlendirmeye alınan parametreler sonucunda tespit edilen en yüksek G-Mean değeri “bayrakli_minmax” eğitim setinde göstermiştir. 0.5840 değeri gösteren optimizasyon çalışmasında kullanılması gereken parametre girdi değerleri aşağıdaki şekilde belirtilmiştir:

```
params = {
'n_estimators': 280,
'max_depth': 3,
'min_samples_split': 16,
'min_samples_leaf': 13,
'max_features': 0.75195833337004,
'bootstrap': True}
```

Extra Trees algoritmasında yapılan “Under Sampling” metodu ile yeniden optimize etme senaryosu, Random Forest için de yapılmıştır. G-Mean değeri; en yüksek değeri “bayrakli_minmax” eğitim setinde göstermiştir. 0.5830 değeri gösteren optimizasyon çalışmasında kullanılması gereken parametre girdi değerleri sunulmuştur:

```
params = {
'n_estimators': 371,
'max_depth': 4,
'min_samples_split': 23,
'min_samples_leaf': 48,
'max_features': 0.5955450025295629}
```

MLP için yapılan optimizasyonda;

```
'hidden_layer_sizes': hidden_sizes,  
'activation': trial.suggest_categorical('activation', ['relu', 'tanh', 'logistic']),  
'solver': trial.suggest_categorical('solver', ['adam', 'sgd']),  
'alpha': trial.suggest_loguniform('alpha', 1e-6, 1e-1),  
'learning_rate_init': trial.suggest_loguniform('learning_rate_init', 1e-5, 1e-1),  
'batch_size': trial.suggest_categorical('batch_size', ['auto', 32, 64, 128]),  
'learning_rate': 'adaptive',  
'max_iter': trial.suggest_int('max_iter', 200, 1000),  
'tol': trial.suggest_loguniform('tol', 1e-5, 1e-2),  
'early_stopping': True,  
'validation_fraction': 0.1,  
'n_iter_no_change': 10,  
'random_state': 42,  
'verbose': False
```

parametreleri ile değerlendirmeye alınmış olup G-Mean değeri; en yüksek değeri “outliersiz_minmax” eğitim setinde göstermiştir. 0.4825 değeri gösteren optimizasyon çalışmasında kullanılması gereken parametre girdi değerleri şu şekildedir:

```
params = {  
    'n_layers': 1,  
    'n_units_l0': 105,  
    'activation': 'tanh',  
    'solver': 'adam',  
    'alpha': 6.1682020976541694e-06,  
    'learning_rate_init': 0.0032323251869364944,  
    'batch_size': 32,  
    'max_iter': 622,  
    'tol': 1.1212110773841794e-05}
```

Extra Trees ve Random Forest algoritmalarında yapılan “Under Sampling” metodu ile yeniden optimize etme senaryosu MLP için de yapılmıştır. G-Mean değeri; en yüksek değeri “outliersiz_robust” eğitim setinde göstermiştir. 0.6027 değeri gösteren optimizasyon çalışmasında kullanılması gereken parametre girdi değerleri şu şekildedir:

```
'n_layers': 3,  
'n_units_l0': 16,  
'n_units_l1': 102,  
'n_units_l2': 60,  
'activation': 'tanh',  
'solver': 'adam',  
'alpha': 0.006051216663169329,  
'learning_rate_init': 0.0018610417202730345,  
'batch_size': 'auto',  
'max_iter': 889,  
'tol': 0.005546934118633894}
```

LSTM için tüm eğitim setleri üzerinde optimizasyon çalışması gerçekleştirilmiştir. Bu optimizasyon çalışması kapsamında belirtilen LSTM parametreleri aşağıda yer almaktadır:

```
units      = trial.suggest_int('units', 16, 256)  
dropout    = trial.suggest_uniform('dropout', 0.0, 0.5)  
recurrent_dropout= trial.suggest_uniform('recurrent_dropout', 0.0, 0.5)  
optimizer_name = trial.suggest_categorical('optimizer', ['adam', 'rmsprop'])  
lr         = trial.suggest_loguniform('lr', 1e-5, 1e-2)  
batch_size = trial.suggest_categorical('batch_size', [32, 64, 128])  
epochs     = trial.suggest_int('epochs', 20, 100)
```

parametreleri ile değerlendirmeye alınmıştır. Sistem mimarisinde kullanılan yapı ise şu şekildedir:

```

model = Sequential([
    LSTM(units,
         input_shape=(window_size, len(scale_cols)),
         dropout=dropout,
         recurrent_dropout=recurrent_dropout),
    Dense(1, activation='sigmoid')
])
optimizer = (tf.keras.optimizers.Adam(learning_rate=lr)
             if optimizer_name == 'adam'
             else tf.keras.optimizers.RMSprop(learning_rate=lr))
model.compile(loss='binary_crossentropy', optimizer=optimizer)
return model, batch_size, epochs

```

Bu kapsamda; G-Mean değeri; en yüksek değeri “outliersiz_minmax” eğitim setinde göstermiştir. 0.5387 değeri gösteren optimizasyon çalışmasında kullanılması gereken parametre girdi değerleri şu şekildedir:

```

'units': 185,
'dropout': 0.25554389239994013,
'recurrent_dropout': 0.489247820368088,
'optimizer': 'rmsprop',
'lr': 2.5655290812448024e-05,
'batch_size': 64,
'epochs': 85}

```

Hist Gradient Boosting için yapılan optimizasyon çalışmasında;

```

params = {
'learning_rate': trial.suggest_loguniform('learning_rate', 1e-4, 1.0),
'max_iter': trial.suggest_int('max_iter', 50, 500),
'max_leaf_nodes': trial.suggest_int('max_leaf_nodes', 10, 200),
'max_depth': trial.suggest_int('max_depth', 3, 20),
'min_samples_leaf': trial.suggest_int('min_samples_leaf', 5, 100),
'l2_regularization': trial.suggest_loguniform('l2_regularization', 1e-8, 10.0),
'max_bins': trial.suggest_int('max_bins', 64, 255),

```

```
'tol':          trial.suggest_loguniform('tol', 1e-7, 1e-4),
'early_stopping': 'auto',
'validation_fraction': 0.1,
'n_iter_no_change': 10,
'class_weight':   'balanced',
'random_state':  42
}
```

parametreleri ile değerlendirmeye alınmış olup G-Mean değeri; en yüksek değeri “bayrakli_robust” eğitim setinde göstermiştir. 0.5885 değeri gösteren optimizasyon çalışmasında kullanılması gereken parametre girdi değerleri şu şekildedir:

```
params = {
'learning_rate': 0.001121154216335312,
'max_iter': 478,
'max_leaf_nodes': 67,
'max_depth': 20,
'min_samples_leaf': 31,
'l2_regularization': 4.590594914041739,
'max_bins': 243, 'tol': 3.949507491802852e-06}
```

Naive Bayes için yapılan optimizasyon çalışmasında;

```
vs = trial.suggest_loguniform("var_smoothing", 1e-12, 1e-1)
ppos = trial.suggest_uniform("prior_positive", 0.01, 0.99)
```

parametreleri değerlendirilmiş olup G-Mean değeri 0.5401 olarak sonuç vermiştir. Bu sonucu verirken parametre çıktısını şu şekilde vermiştir:

```
'var_smoothing': 0.00016340970330600552,
'prior_positive': 0.9199057793605877}
```

NaiveBayes için yapılan optimizasyon çalışması, Under Sampling metodu ile genişletilerek tekrardan yapılmıştır. 0.5412 değerine ulaşan G-Mean için;

```
'var_smoothing': 2.39724297671834e-10,  
'prior_positive': 0.9326075329989663}  
sonuçları üretilmiştir.
```

KNN modeli için, Under Sampling metodu ve ham veri ile optimizasyon çalışması yapılmış olup;

```
n_neighbors = trial.suggest_int("n_neighbors", 1, 50)  
weights = trial.suggest_categorical("weights", ["uniform", "distance"])  
p = trial.suggest_int("p", 1, 2)
```

parametreleri üzerinde değerlendirme yapılmıştır. G-Mean değeri; ham veri üzerinde 0.4244 ('n_neighbors': 2, 'weights': 'distance', 'p': 2) olarak bulunurken Under Sampling metodu ile 0.5565 ('n_neighbors': 14, 'weights': 'distance', 'p': 2) değeri bulunmuştur.

Lojistik regresyon için yapılan çalışmada;

```
'penalty', [None, 'l1', 'l2', 'elasticnet'])  
solver = 'saga'  
l1_ratio = 'l1_ratio', 0.0, 1.0)  
C = ('C', 1e-4, 1e2)  
tol = ('tol', 1e-5, 1e-1)
```

parametreleri değerlendirmeye alınmış olup outliersiz_robust eğitim setinde 0.5942 G-Mean değerine ulaşmıştır. Parametre sonuçları; 'penalty': None, 'C': 0.055880245548626896, 'tol': 0.0028153545049531474 çıktısını vermiştir.

CatBoost modeli için yapılan optimizasyonda;

```
'iterations': trial.suggest_int('iterations', 100, 1000),  
'depth': trial.suggest_int('depth', 3, 10),  
'learning_rate': trial.suggest_loguniform('learning_rate', 1e-3, 1e-1),  
'l2_leaf_reg': trial.suggest_loguniform('l2_leaf_reg', 1e-3, 10),  
'bagging_temperature': trial.suggest_uniform('bagging_temperature', 0, 1),  
'random_strength': trial.suggest_uniform('random_strength', 0, 1),  
'border_count': trial.suggest_int('border_count', 32, 255),  
'auto_class_weights': 'Balanced',  
'random_state': 42,  
'thread_count': -1,  
'verbose': False
```

parametreleri değerlendirmeye alınmış olup bayrakli_robust eğitim seti ile G-Mean değeri en yüksek sonucu vermiştir. 0.5838 değeri çıktı olarak veren CatBoost için çıktı parametreleri ise şu şekilde belirtilmiştir:

```
'iterations': 112,  
'depth': 4,  
'learning_rate': 0.0024467585151287923,  
'l2_leaf_reg': 0.002434916226546548,  
'bagging_temperature': 0.07735935935093266,  
'random_strength': 0.154401173186064,  
'border_count': 49
```

Optimizasyon yapılmadan önce, TSS değerleri hariç değerlendirme yapılması gerekirse; G-Mean başarı ölçütüne göre bayrakli_robust eğitim seti ile XGBoost (0.584), outliersiz_minmax eğitim seti ile Gradient Boosting (0.578), bayrakli_robust eğitim seti ile Gradient Boosting (0.564), outliersiz_robust eğitim seti ile Gradient Boosting (0.541) ilk sıraları almaktadır.

Optuna algoritması ile yapılan optimizasyon alıřmaları sonrasında Gradient Boosting (Under Sampling) ile 0.6062, MLP (Under Sampling) ile 0.6027, LGBM ile 0.5971, Lojistik Regresyon ile 0.5942 ve XGBoost ile 0.5917 deęerleri alınmıřtır.

TSS ve “Sequence” yapılarında yapılan alıřmalarda is en yksek deęer olarak Fold-11’de alınan sonular bulunmaktadır. outliersiz_robust eęitim seti ile Transformer Model, “Sequence” yapıda 0.663 deęerine ulařmıřtır. Ancak 4 yıllık veri setinde 2024 yılının son eyreęi zerinden yapılan testlerde 0.6062 deęeri (Gradient Boosting (Under Sampling)) grlmřtr.



4. SONUÇ VE ÖNERİLER

Tez çalışması kapsamında 2021 – 2024 yılları arasında, İzmir ve Manisa illerinde bulunan 34 farklı güç transformatörü için çalışma yapılmıştır. Yapılan çalışmada; veri ön işleme süreçleri, model çeşitliliği, veri dengesizliğine ait opsiyonel olarak kullanılan algoritmalar, TSS ve Sequence ile zamansallığın katkısı ve optimizasyon çalışmaları ile G-Mean başarı metriğine göre modeller karşılaştırılmıştır. 1220 farklı test sonucunda (44 adet Under Sampling, 44 adet SMOTE, 44 adet classweight, 44 adet ham veri, 768 adet TSS ve 276 adet sequence) en yüksek değer 0.663 çıktığı tespit edilmiştir.

Oluşturulan modeller için 59 adet özellik kullanılmış olup, meta veri ve hava durumu verileri ile çalışmalar gerçekleştirilmiştir. Yapılan çalışmada, güç transformatörlerinin günlük bazda kesinti sayılarının dengesizliği, spesifik şebeke unsuruna odaklanması ve Recall, Accuracy, Precision gibi parametrelerin dengesiz veri setlerinde yanıltıcı olmasından kaynaklı G-Mean değerinin 0.58-0.61 değerleri arasında sınırlandırıldığı gözlemlenmiştir. Literatürde yapılan birçok çalışmanın aksine, operasyonel süreçlerde kullanıma uygun model tespiti için yapılan karşılaştırmalarda G-Mean değeri kullanılması literatüre katkı sağlamaktadır.

Tez kapsamında yapılan çalışmanın; modelin hedef alanını değiştirerek değerlendirilmesi ile vardiya planlama, SAIDI/SAIFI değerlerinde iyileştirme, bakım planlama çalışmaları yapılabilmektedir. Ancak çalışma kapsamında değerlendirilen verilerin kesinti ile korelasyonu ve ilişkiselliği yeterli seviyelerde gözükmemektedir. Bu kapsamda oluşturulan modellerin “mücbir sebeplerden kaynaklanan arıza kayıtlarının raporlanması” kapsamında bahsedilen regülatif kurumlara yapılan raporlamalarda bilimsel gerekçeler sunma hususunda yeterli kanıt ve ilişkisellik üretilmediği tespit edilmiştir. İlerleyen çalışmalarda doğru tespit ve regülatif kurumlara raporlanabilir olması için akıllı sayaç verileri, trafo doluluk verileri, geçmiş dönemlere ait periyodik bakım verileri, manevra log verileri, transformatör üzerinden beslemesi gerçekleştirilen son kullanıcı tipi- sayısı bilgilerinin olduğu daha kapsamlı bir veri seti oluşturulması ile yapılan çalışmanın hava durumu ile birlikte teknik parametreleri de gözetlemesini ve denetlemesini sağlamak hedeflenmektedir.

Tez kapsamında yapılan çalışmalar ve model sonuçları kapsamında;

- Mücbir sebeplerden dolayı oluşan kesintiler için, hava durumu ilişkiselliği tespit edilmemiştir.
- 2021-2024 yılları arasındaki kullanılan verilerde k-fold çalışması kapsamında 2024 yılındaki verilerin tahminlemeyi olumsuz etkilediği tespit edilmiştir.
- G-Mean parametresinin literatürdeki diğer uygulamalara göre sektörel anlamda kullanılabilirliği daha fazla vurguladığı gösterilmiştir.
- TSS, k-fold, Optuna optimizasyon ve sınıf dengesizliği çalışmalarında ulaşılabilen değerlerin 0.58 – 0.62 arasında kalması hava durumu ilişkiselliğini daha da arttıramayacağını göstermektedir.
- Hava durumu verileriyle birlikte girdi olarak kullanılacak daha fazla meta veri kullanılmasının gerekliliği tespit edilmiştir.
- TSS kapsamında yapılan testlerde Fold-11 (04.08.2023 - 29.10.2023) test sürecinde en yüksek değerlere ulaşıldığı tespit edilmiştir (G-Mean : 0.663, ROC-AUC: 0.7138).
- K-Fold çalışması haricinde, model genellenebilirliği ve optimizasyon sonuçlarına göre Gradient Boosting- Under Sampling (G-Mean: 0.6062) en yüksek değere ulaşmıştır.

Tez kapsamında yapılan çalışmanın mevcut saha operasyonlarına katkıda bulunması için uyarı mekanizması geliştirilerek çevrim içi izleme – canlı bildirim sistemi oluşturulması hedeflenmektedir. Yapılan çalışmalara ek olarak; gelecek çalışmalarda mevcut veri setinde, güç transformatörlerinden (şebeke unsuruna) beslenen mesken, sanayi ve tarımsal sulama kullanıcıların değerlendirilmesi, manevra işlemlerinin kaydı, hat boyunca alınan izolasyon hataları, röle, SCADA ve akıllı sayaçlar tarafından üretilen verilerin kullanılması ile yalnızca güç transformatörlerinin değil, her şebeke unsuru üzerinde belirli periyotlarda oluşabilecek kesintinin tahminlemesi yapılacaktır.

5. KAYNAKÇA

- [1] M. A. AKCANCA ve S. TAŞKIN, “Akıllı Şebeke Uygulanabilirliği Açısından Türkiye Elektrik Enerji Sisteminin İncelenmesi”, *Akıllı Şebekeler ve Türkiye Elektr. Şebekesinin Geleceği Sempozyumu*, ss. 131–135, 2013, [Çevrimiçi]. Available at: www.emo.org.tr/ekler/a4d6c5a3223642b_ek.pdf.
- [2] The International Energy Agency (IEA), “No Title”, <https://www.iea.org/data-and-statistics/charts/world-electricity-final-consumption-by-sector-1974-2019>, 2021. .
- [3] B. Kocaman, “Mikro Şebekeler için Örnek Bir Enerji Yönetimi Uygulaması A Case of Energy Management Application for Microgrids”, c. 3, sayı 1, ss. 35–52, 2014.
- [4] S. Forsberg, K. Thomas, ve M. Bergkvist, “Power grid vulnerability analysis using complex network theory: A topological study of the Nordic transmission grid”, *Phys. A Stat. Mech. its Appl.*, c. 626, s. 129072, 2023, doi: 10.1016/j.physa.2023.129072.
- [5] A. Visakh ve M. P. Selvan, “Analysis and mitigation of the impact of electric vehicle charging on service disruption of distribution transformers”, *Sustain. Energy, Grids Networks*, c. 35, s. 101096, 2023, doi: 10.1016/j.segan.2023.101096.
- [6] R. J. Campbell, “Weather-related power outages and electric system resiliency”, *Transform. Electr. Power Key Issues*, ss. 103–118, 2013.
- [7] I. E. A. IEA, “World Energy Outlook 2024”, 2024.
- [8] D. A. Zaldivar, A. M. Sanchez, ve A. A. Romero, “A comprehensive methodology for the optimization of condition-based maintenance in power transformer fleets”, *Electr. Power Syst. Res.*, c. 220, sayı March, 2023, doi: 10.1016/j.epr.2023.109374.
- [9] P. J. Marcotullio *vd.*, “Local power outages, heat, and community characteristics in New York City”, *Sustain. Cities Soc.*, c. 99, sayı July, s. 104932, 2023, doi: 10.1016/j.scs.2023.104932.
- [10] ELDER, “Türkiye Akıllı Şebekeler 2023 Vizyon ve Strateji Belirleme Projesi”, 2023, [Çevrimiçi]. Available at: www.akillisebekelerturkiye.org.
- [11] İ. Yılmaz, “ELEKTRİK PİYASASINDA DİJİTALLEŞME VE AKILLI ŞEBEKE YATIRIMLARI”, c. 470, sayı Tablo 1, ss. 25–29, 2022.

- [12] X. Jia, H. Chen, Y. Yang, C. Xu, L. Duanmu, ve Z. Wang, “Research on integrative optimization operation of seawater heat pump, photovoltaic, and cross-seasonal heat storage systems”, *Renew. Energy*, c. 246, sayı March, 2025, doi: 10.1016/j.renene.2025.122864.
- [13] R. G. Verdier, “GRID4EU-Nice Grid project: How to facilitate the integration of Distributed Energy Resources into the local grid?”, *2014 Saudi Arab. Smart Grid Conf. SASG 2014*, ss. 1–3, 2014, doi: 10.1109/SASG.2014.7274297.
- [14] T.C. Resmî Gazete, *Elektrik Piyasası Kanunu*. Türkiye, 2001.
- [15] T.C. Resmî Gazete, *Elektrik Piyasası Kanunu*. Türkiye, 2013.
- [16] M. Doostan ve B. Chowdhury, “Predicting lightning-related outages in power distribution systems: A statistical approach”, *IEEE Access*, c. 8, ss. 84541–84550, 2020, doi: 10.1109/ACCESS.2020.2991923.
- [17] W. Song *vd.*, “On the prediction of power outage length based on linear multifractional Lévy stable motion”, *Pattern Recognit. Lett.*, c. 181, sayı March, ss. 120–125, 2024, doi: 10.1016/j.patrec.2024.03.013.
- [18] B. Ghasemkhani *vd.*, “Machine Learning Model Development to Predict Power Outage Duration (POD): A Case Study for Electric Utilities”, *Sensors*, c. 24, sayı 13, 2024, doi: 10.3390/s24134313.
- [19] “Enerji Verimliliği 2030 Stratejisi ve II. Ulusal Enerji Verimliliği Eylem Planı”, Ankara, 2024. [Çevrimiçi]. Available at: https://enerji.gov.tr/Media/Dizin/BHIM/tr/Duyurular/TurkiyeninEnerjiVerimliliği2030StratejisiVeIIUlusalEnerjiVerimliliğiEylemPlanı_202401161407.pdf.
- [20] A. Yuyama, Y. Kajitani, ve G. Shoji, “Simulation of operational reliability of thermal power plants during a power crisis: Are we underestimating power shortage risk?”, *Appl. Energy*, c. 231, sayı February, ss. 901–913, 2018, doi: 10.1016/j.apenergy.2018.09.089.
- [21] H. Ren, Z. J. Hou, X. Ke, Q. Huang, ve Y. Makatov, “Analysis of Weather and Climate Extremes Impact on Power System Outage”, *IEEE Power Energy Soc. Gen. Meet.*, c. 2021-July, ss. 1–5, 2021, doi: 10.1109/PESGM46819.2021.9637938.
- [22] L. Qian, T. Qingzhao, Z. Qinghui, ve W. Tianqi, “Research on power transformer fault prediction model based on LSTM neural network”, *Proc. - 2020 Int. Conf. Robot. Intell. Syst. ICRIS 2020*, ss. 702–705, 2020, doi: 10.1109/ICRIS52159.2020.00175.

- [23] V. Rexhepi, “An Analysis of Power Transformer Outages and Reliability Monitoring”, *Energy Procedia*, c. 141, ss. 418–422, 2017, doi: 10.1016/j.egypro.2017.11.053.
- [24] Z. Xu, J. Liu, W. Fan, Y. Wang, L. Luan, ve K. Zhou, “An Power Outage Prediction Method Based on XGBoost with Improved Objective Function”, *Proc. - 2022 37th Youth Acad. Annu. Conf. Chinese Assoc. Autom. YAC 2022*, ss. 1560–1565, 2022, doi: 10.1109/YAC57282.2022.10023921.
- [25] Ö. Tuttokmaği ve A. Kaygusuz, “Araştırma Makalesi / Research Article Büyük Ölçekli Elektrik Kesintilerinin İncelenmesi Investigation of Large Scale Power Blackouts”, c. 8, sayı 2, ss. 664–676, 2019.
- [26] M. Papic *vd.*, “Effects of dependent and common mode outages on the reliability of bulk electric system - Part I: Basic concepts”, *IEEE Power Energy Soc. Gen. Meet.*, c. 2014-Octob, sayı October, 2014, doi: 10.1109/PESGM.2014.6938967.
- [27] A. Ucar, M. Karakose, ve N. Kırımça, “applied sciences Artificial Intelligence for Predictive Maintenance Applications ”:., 2024.
- [28] L. Jian ve T. Tianyuan, “LS-SVM based substation circuit breakers maintenance scheduling optimization”, *Int. J. Electr. Power Energy Syst.*, c. 64, ss. 1251–1258, 2015, doi: 10.1016/j.ijepes.2014.09.013.
- [29] H. Norouzi, “Flexible assessment of line polymer insulators based on condition indexes”, *Electr. Power Syst. Res.*, c. 238, sayı October 2024, s. 111145, 2025, doi: 10.1016/j.epsr.2024.111145.
- [30] C. Jin-Qiang, “Fault Prediction of a Transformer Bushing Based on Entropy Weight TOPSIS and Gray Theory”, *Comput. Sci. Eng.*, c. 21, sayı 6, ss. 55–62, 2019, doi: 10.1109/MCSE.2018.2882357.
- [31] S. A. Gamel, S. S. M. Ghoneim, ve Y. A. Sultan, “Improving the accuracy of diagnostic predictions for power transformers by employing a hybrid approach combining SMOTE and DNN”, *Comput. Electr. Eng.*, c. 117, sayı April, s. 109232, 2024, doi: 10.1016/j.compeleceng.2024.109232.
- [32] J. F. Toubeau *vd.*, “Machine learning-assisted outage planning for maintenance activities in power systems with renewables”, *Energy*, c. 238, 2022, doi: 10.1016/j.energy.2021.121993.
- [33] F. Xu, L. Zhou, Y. Zhou, X. Ai, ve Q. Liu, “Power outage data management platform based on big data analysis technology”, *Procedia Comput. Sci.*, c. 261,

- ss. 976–982, 2025, doi: 10.1016/j.procs.2025.04.489.
- [34] C. Fan, Y. Lei, Y. Sun, ve L. Mo, “Novel transformer-based self-supervised learning methods for improved HVAC fault diagnosis performance with limited labeled data”, *Energy*, c. 278, sayı PB, s. 127972, 2023, doi: 10.1016/j.energy.2023.127972.
- [35] Z. Zhao, F. Chen, P. Lan, Y. Peng, X. Yin, ve X. Dong, “How to mine the abnormal information of power transformers: An efficient tool for quantifying the fault characteristics via multi-vibration signals”, *Adv. Eng. Informatics*, c. 62, sayı PA, s. 102561, 2024, doi: 10.1016/j.aei.2024.102561.
- [36] K. Fatima, H. Shareef, F. B. Costa, A. A. Bajwa, ve L. A. Wong, “Machine learning for power outage prediction during hurricanes: An extensive review”, *Eng. Appl. Artif. Intell.*, c. 133, sayı PB, s. 108056, 2024, doi: 10.1016/j.engappai.2024.108056.
- [37] C. Zhai, T. Y. jeh Chen, A. G. White, ve S. D. Guikema, “Power outage prediction for natural hazards using synthetic power distribution systems”, *Reliab. Eng. Syst. Saf.*, c. 208, sayı November 2020, s. 107348, 2021, doi: 10.1016/j.res.2020.107348.
- [38] T. Zhao *vd.*, “On Learning-Based Model for Dynamic Granular Prediction of Power Outages Under Extreme Events”, *2023 IEEE PES Innov. Smart Grid Technol. Lat. Am. ISGT-LA 2023*, ss. 445–449, 2023, doi: 10.1109/ISGT-LA56058.2023.10328277.
- [39] S. A. Gamel, S. S. M. Ghoneim, ve Y. A. Sultan, “Improving the accuracy of diagnostic predictions for power transformers by employing a hybrid approach combining SMOTE and DNN”, *Comput. Electr. Eng.*, c. 117, sayı March, s. 109232, 2024, doi: 10.1016/j.compeleceng.2024.109232.
- [40] S. Cao, F. Wei, X. Lin, X. Yuan, Q. Huang, ve H. Xiang, “Risk prediction based preventive typhoon defending for semi-independent power system”, *Appl. Energy*, c. 377, sayı PC, s. 124389, 2025, doi: 10.1016/j.apenergy.2024.124389.
- [41] X. Tang, T. Machimura, J. Li, W. Liu, ve H. Hong, “A novel optimized repeatedly random undersampling for selecting negative samples: A case study in an SVM-based forest fire susceptibility assessment”, *J. Environ. Manage.*, c. 271, s. 111014, 2020, doi: 10.1016/j.jenvman.2020.111014.
- [42] P. Kulczycki ve K. Franus, “Methodically unified procedures for a conditional approach to outlier detection, clustering, and classification”, *Inf. Sci. (Ny)*, c.

- 560, ss. 504–527, 2021, doi: 10.1016/j.ins.2020.08.122.
- [43] H. Yao, Y. Xu, Q. Guo, S. Chen, B. Lu, ve Y. Huang, “Study on transformer fault diagnosis based on improved deep residual shrinkage network and optimized residual variational autoencoder”, *Energy Reports*, c. 13, sayı November 2024, ss. 1608–1619, 2025, doi: 10.1016/j.egy.2025.01.037.
- [44] T. Zhao vd., “Variational data augmentation for a learning-based granular predictive model of power outages”, *Electr. Power Syst. Res.*, c. 232, sayı March, s. 110299, 2024, doi: 10.1016/j.epsr.2024.110299.
- [45] E. Cruz ve T. Munasinghe, “Exploring Power Outage Prediction Using Weather and Socioeconomic Data in the Southeastern Part of the United States”, *Proc. - 2023 IEEE Int. Conf. Big Data, BigData 2023*, ss. 6138–6142, 2023, doi: 10.1109/BigData59044.2023.10386519.
- [46] B. Raja, G. R. Venkatakrishnan, ve R. Rengaraj, “Power transformer fault diagnosis and condition monitoring using hybrid TDO-SNN technique”, *Int. J. Hydrogen Energy*, c. 68, sayı March, ss. 1370–1381, 2024, doi: 10.1016/j.ijhydene.2024.03.115.
- [47] L. Shu, T. Ji, L. Zhang, ve W. Li, “Power Equipment Fault Prediction and Diagnosis Algorithm Driven by Power Big Data”, *2024 IEEE 7th Int. Conf. Autom. Electron. Electr. Eng. AUTEEE 2024*, ss. 157–161, 2024, doi: 10.1109/AUTEEE62881.2024.10869656.
- [48] S. Yuan, S. M. Quiring, L. Zhu, Y. Huang, ve J. Wang, “Development of a Typhoon Power Outage Model in Guangdong, China”, *Int. J. Electr. Power Energy Syst.*, c. 117, sayı September 2019, s. 105711, 2020, doi: 10.1016/j.ijepes.2019.105711.
- [49] J. I. Aizpurua vd., “Probabilistic machine learning aided transformer lifetime prediction framework for wind energy systems”, *Int. J. Electr. Power Energy Syst.*, c. 153, sayı July, s. 109352, 2023, doi: 10.1016/j.ijepes.2023.109352.
- [50] J. B. Thomas ve S. K.V., “Neural architecture search algorithm to optimize deep Transformer model for fault detection in electrical power distribution systems”, *Eng. Appl. Artif. Intell.*, c. 120, sayı July 2022, s. 105890, 2023, doi: 10.1016/j.engappai.2023.105890.
- [51] C. Brester, H. Niska, R. Cizek, ve M. Kolehmainen, “Weather-based Fault Prediction in Electricity Networks with Artificial Neural Networks”, *2020 IEEE Congr. Evol. Comput. CEC 2020 - Conf. Proc.*, 2020, doi:

10.1109/CEC48606.2020.9185555.

- [52] R. Singh, “Analyzing Spatiotemporal Trends and Anomalies in Power Outage Data Across the United States”, *Proc. - 2024 IEEE Int. Conf. Big Data, BigData 2024*, ss. 7417–7421, 2024, doi: 10.1109/BigData62323.2024.10825495.
- [53] M. Zhuang ve A. Sadikovic, “Semi-Supervised Anomaly Detection Framework Using VAE-LSTM to Enhance Outage Prediction”, *2024 IEEE Power Energy Soc. Gen. Meet.*, ss. 1–5, 2024, doi: 10.1109/PESGM51994.2024.10688421.
- [54] W. O. Taylor, D. Cerrai, D. Wanik, M. Koukoura, ve E. N. Anagnostou, “Community power outage prediction modeling for the Eastern United States”, *Energy Reports*, c. 10, sayı August, ss. 4148–4169, 2023, doi: 10.1016/j.egy.2023.10.073.
- [55] Y. Kor, M. Z. Reformat, ve P. Musilek, “Predicting weather-related power outages in distribution grid”, *IEEE Power Energy Soc. Gen. Meet.*, c. 2020-Augus, 2020, doi: 10.1109/PESGM41954.2020.9281829.
- [56] E. A. Morris, K. R. W. Bell, ve I. M. Elders, “Spatial and temporal clustering of fault events on the Gb transmission network”, *2016 Int. Conf. Probabilistic Methods Appl. to Power Syst. PMAPS 2016 - Proc.*, ss. 1–9, 2016, doi: 10.1109/PMAPS.2016.7764087.
- [57] S. Wang, H. Sun, K. J. Kim, J. Guo, ve D. Nikovski, “Solving Bernoulli Bandit Problems for Weather-relative Overhead Distribution Line Failures Forecasting”, *IEEE Power Energy Soc. Gen. Meet.*, c. 2021-July, ss. 0–4, 2021, doi: 10.1109/PESGM46819.2021.9638153.
- [58] M. Abaas, R. A. Lee, ve P. Singh, “Long Short-Term Memory Customer-Centric Power Outage Prediction Models for Weather-Related Power Outages”, *2022 IEEE Green Energy Smart Syst. IGESSC 2022*, sayı M1, ss. 1–6, 2022, doi: 10.1109/IGESSC55810.2022.9955338.
- [59] K. V. Mankov, D. A. Akimov, I. P. Voltov, ve A. D. Ruchkina, “Forecasting of Emergency Power Line Outages Based on Weather Data and Past Emergency Outages”, *Proc. 2022 Conf. Russ. Young Res. Electr. Electron. Eng. ElConRus 2022*, ss. 765–767, 2022, doi: 10.1109/ElConRus54750.2022.9755779.
- [60] Q. Zhou, J. Zhang, Z. Yang, ve Y. Yang, “Distribution network outage pre-warning analytics under extreme weather conditions”, *China Int. Conf. Electr. Distrib. CICED*, c. 2014-Decem, sayı Ciced, ss. 1190–1194, 2014, doi: 10.1109/CICED.2014.6991895.

- [61] H. Hou, C. Liu, R. Wei, H. He, L. Wang, ve W. Li, “Outage duration prediction under typhoon disaster with stacking ensemble learning”, *Reliab. Eng. Syst. Saf.*, c. 237, sayı May, s. 109398, 2023, doi: 10.1016/j.ress.2023.109398.
- [62] H. Hou, X. Chen, M. Li, L. Zhu, Y. Huang, ve J. Yu, “Prediction of user outage under typhoon disaster based on multi-algorithm Stacking integration”, *Int. J. Electr. Power Energy Syst.*, c. 131, sayı February, s. 107123, 2021, doi: 10.1016/j.ijepes.2021.107123.
- [63] T. Zhao vd., “Variational data augmentation for a learning-based granular predictive model of power outages”, *Electr. Power Syst. Res.*, c. 232, sayı April, s. 110299, 2024, doi: 10.1016/j.epsr.2024.110299.
- [64] W. Liu, Y. Yang, Q. Xu, ve Y. Xia, “Multi-target prediction model of urban distribution system rainfall-caused outage based on spatiotemporal fusion”, *Int. J. Electr. Power Energy Syst.*, c. 146, sayı November 2022, s. 108640, 2023, doi: 10.1016/j.ijepes.2022.108640.
- [65] A. K. Onalapo, R. P. Carpanen, D. G. Dorrell, ve E. E. Ojo, “A Comparative Assessment of Conventional and Artificial Neural Networks Methods for Electricity Outage Forecasting”, *Energies*, c. 15, sayı 2, ss. 1–21, 2022, doi: 10.3390/en15020511.
- [66] W. O. Taylor, D. Cerrai, D. Wanik, M. Koukoulou, ve E. N. Anagnostou, “Community power outage prediction modeling for the Eastern United States”, *Energy Reports*, c. 10, sayı October, ss. 4148–4169, 2023, doi: 10.1016/j.egy.2023.10.073.
- [67] Z. E. KORKMAZ, “Modeling Longitudinal Interruption Data From Turkish ELECTRICITY DISTRIBUTION COMPANIES”, sayı May, 2019.
- [68] Orçun KİTAPCI, “Derin öğrenme modelleri ile elektrik kesinti tahmini”, 2022.
- [69] Ö. E. YÜREK, “GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES Application of Machine Learning Methods in Manufacturing Sector”, 2024.