

T.C.
İSTANBUL SABAHATTİN ZAİM ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR BİLİMLERİ VE MÜHENDİSLİĞİ BİLİM DALI

ELEKTRİK DAĞITIM ŞEBEKELERİNDE
TEKNİK OLMAYAN KAYIP KAÇAKLARIN
MAKİNE ÖĞRENMESİ İLE TESPİTİ

DOKTORA TEZİ

Mustafa ÇELİKPENÇE

İstanbul
Mart-2023

T.C.
İSTANBUL SABAHATTİN ZAİM ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR BİLİMLERİ VE MÜHENDİSLİĞİ BİLİM DALI

ELEKTRİK DAĞITIM ŞEBEKELERİNDE
TEKNİK OLMAYAN KAYIP KAÇAKLARIN
MAKİNE ÖĞRENMESİ İLE TESPİTİ

DOKTORA TEZİ

Mustafa ÇELİKPENÇE

Tez Danışmanı
Doç. Dr. Ferzat ANKA

İstanbul
Mart-2023

TEZ ONAYI

Lisansüstü Eğitim Enstitüsü Müdürlüğüne,

Bu çalışma, jürimiz tarafından Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Bilimleri ve Mühendisliği Bilim Dalında DOKTORA TEZİ olarak kabul edilmiştir.

Danışman Doç. Dr. Ferzat ANKA

Üye Dr. Öğr. Üyesi Abdullah SÖNMEZ

Üye Dr. Öğr. Üyesi Zahra ELMİ

Üye Dr. Öğr. Üyesi Ahmet KAPLAN

Üye Dr. Öğr. Üyesi Zeynep ALTAN

Onay

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım.

Doç. Dr. Erhan İÇENER

Enstitü Müdürü

BİLİMSEL ETİK BİLDİRİMİ

Doktora tezi olarak hazırladığım “**Elektrik Dağıtım Şebekelerinde Teknik Olmayan Kayıp Kaçakların Makine Öğrenmesi İle Tespiti**” adlı çalışmanın öneri aşamasından sonuçlandığı aşamaya kadar geçen süreçte bilimsel etiğe ve akademik kurallara özenle uyduğumu, tez içindeki tüm bilgileri bilimsel ahlak ve gelenek çerçevesinde elde ettiğimi, tez yazım kurallarına uygun olarak hazırladığımı, bu çalışmamda doğrudan veya dolaylı olarak yaptığım her alıntıya kaynak gösterdiğimi ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu beyan ederim.

Mustafa ÇELİKPENÇE

ÖN SÖZ

Araştırmamın her aşamasında bana gerekli desteği sağlayan Eksim Holding Yönetimi'ne, Üniversite ve Bölümüme, projeyi geliştirmede desteklerinden dolayı Dicle Elektrik Dağıtım çalışanlarına teşekkür ederim. Çalışmanın her aşamasında bana yardımcı olan ve yön veren, değerli danışman hocam Doç. Dr. Ferzat ANKA'ya, özellikle çalışmaya katalizör olarak katkılarından dolayı Yönetim Kurulu Danışmanı, değerli yaşam koçum Reşat Atalar beyefendiye ve yoğun doktora eğitimim boyunca desteklerini hiç esirgemeyen özellikle sevgili eşim Rukiye Çelikpençe'ye ve çocuklarıma teşekkürlerimi sunarım.

Mustafa ÇELİKPENÇE
İstanbul - 2023

ÖZET

ELEKTRİK DAĞITIM ŞEBEKELERİNDE TEKNİK OLMAYAN KAYIP KAÇAKLARIN MAKİNE ÖĞRENMESİ İLE TESPİTİ

Mustafa ÇELİKPENÇE

Doktora, Bilgisayar Bilimleri ve Mühendisliği

Tez Danışmanı: Doç. Dr. Ferzat ANKA

Mart, 2023 - 112 + XVI Sayfa

Kaçak elektrik kullanımı, Türkiye'deki dağıtım şirketlerinin önemli sorunlarından biridir. Özellikle Türkiye'nin güneydoğu illeri Diyarbakır, Şanlıurfa, Mardin, Siirt, Batman ve Şırnak illerini kapsayan Dicle Elektrik Dağıtım Bölgesi, 2018 yılı için %54,94 kayıp kaçak oranıyla Türkiye'de kaçak elektrik kullanımının en yüksek olduğu dağıtım bölgesidir. Bu çalışmada, makine öğrenmesi tabanlı 6 sınıflandırma algoritması kullanılarak, Teknik Olmayan Kayıp Kaçaklar tespit edilmeye çalışılmıştır. Bu amaçla Müşteri Bilgi Sistemi (MBS), Coğrafi Bilgi Sistemi (CBS), Saha Yönetim Sistemi (SYS), ve iki farklı Otomatik Sayaç Okuma Sistemi (OSOS) olmak üzere beş büyük ve canlı sistem verisi kullanılmıştır.

Çalışmanın ilk fazında, SQL teknikleri ile veri temizliği ve min-max ölçekleme ile veri normalizasyonu yöntemleri, veri ön işlemede uygulanmıştır. İkinci fazda öznelik çıkarımı gerçekleştirilerek, çıkarılan öznelikler ilgili sınıflandırma metodlarına girdi olarak kullanılmıştır. Son fazda ise, doğrulama süreci gerçekleştirilmiştir. Doğrulama için doğruluk, duyarlılık, kesinlik ve F1 skoru metrikleri kullanılmıştır. Sınıflandırma algoritmaları parametreleri optimizasyonu, öznelik sayısının değişimi ve meta sezgisel algoritmalarından İkili Parçacık Sürü Optimizasyonu (BPSO) kullanılarak, geliştirilen model metriklerinde iyileştirme çalışmaları gerçekleştirilmiştir. Gridsearch yöntemi ile RF parametreleri optimizasyonu sağlanarak, %68,6 ile en yüksek doğruluk değeri elde edilmiştir. LR ile doğruluk, duyarlılık ve kesinlik değerleri olarak sırasıyla %68,6, %75,25 ve %74,91 sonuçları elde edilmiştir. SVM ile aynı değerler, %67,93, %82,83, %66,14 olarak elde edilmiştir. XGB ise %98,41 ile en yüksek duyarlılık değerine sahiptir.

Elektrik kaçak tespitinde makine öğrenmesi yöntemleri kullanılarak yapılan bu çalışma, bu alanda Türkiye'de yapılan ilk çalışmadır. Bu çalışma, Türkiye'de Dicle

EDAŞ'ın Teknik Olmayan Kayıp Kaçak oranının düşürülmesine yardımcı olabilir. Aynı zamanda bu çalışmanın, makine öğrenmesi teknikleri ile teknik olmayan kaçakların tespiti üzerinde çalışan uzmanlar için, bilgi kaynağı olması beklenmektedir.

Anahtar Kelimeler: Elektrikte Teknik Olmayan Kaçak Tespiti, Akıllı Şebekeler, Makine Öğrenmesi, Otomatik Sayaç Okuma, Metasezgisel Algoritmalar



ABSTRACT
NON-TECHNICAL LOSS DETECTION ON ELECTRICITY
DISTRUBUTION GRIDS WITH MACHINE LEARNING

Mustafa ÇELİKPENÇE

Ph. D. Computer Science and Engineering

Supervisor: Assoc. Prof. Dr. Ferzat ANKA

March, 2023 - 112 + XVI Pages

Electricity Theft is a big problem for Power Distribution Companies (PDCs) in Turkey. Specially, Dicle Electricity Distribution Area (Dicle EDAS), include southeast cities of Turkey like Diyarbakır, Şanlıurfa, Mardin, Siirt, Batman, Şırnak cities of Turkey, is a distribution area with the highest usage of electricity theft with %54,94. In this study, Non-Technical Loss (NTL) is tried to be detected by using machine learning-based six classification methods. For this purpose, five large and live system data, Customer Information System (CIS), Geographic Information System (GIS), Field Management System (FMS) and two different Automatic Meter Reading (AMR) systems, were used.

In the first phase of the study, data cleaning with SQL techniques and normalization with Min-Max scaler is performed in data pre-processing. In the second phase, feature extraction is utilized and given as input to the related classification methods. In the last phase, the validation process is carried out. For validation, accuracy, recall and f1-score metrics are used. Metrics optimization of developed model are performed with parameter optimization of classification algorithms, by changing the number of features determined by using Binary Particle Swarm Optimization (BPSO) from meta-huristic algorithms.

Providing RF parameter optimization with Grid Search method, the best accuracy was obtained with %68,6. The evaluation metrics of the LR method used in the model proposed in this study were obtained as 68.6%, 75.25%, and 74,91%. Also, in SVM these values were obtained as 67.93%, 82.83%, and 66,14%. In addition, the best accuracy rate was obtained as 98.41% in XGB. This study using machine learning techniques for electiricity theft detection is the first study in this field in Turkey. This

study may help reduce NTLs of Dicle EDAS in Turkey and at the same time. It is also expected that this study will provide source of information for experts working on NTL detection with machine learning techniques as well.

Keywords: Non-Technical Loss Detection, Smart Grids, Machine Learning, Electricity Theft, Automatic Meter Reading, Metaheuristic Algorithms



İÇİNDEKİLER

TEZ ONAYI	i
BİLİMSEL ETİK BİLDİRİMİ.....	ii
ÖN SÖZ.....	iii
ÖZET	iv
ABSTRACT	vi
İÇİNDEKİLER	viii
TABLolar LİSTESİ.....	xi
ŞEKİLLER LİSTESİ.....	xiii
KISALTMALAR	xv

GİRİŞ.....	1
------------	---

BİRİNCİ BÖLÜM

ELEKTRİKTE KAYIP KAÇAK VE ÜLKEMİZDEKİ DURUMU	4
1.1 Elektrik Dağıtım Şebekesi.....	4
1.2. Elektrikte Kayıp Kaçak	6
1.2.1. Teknik Kayıplar	7
1.2.2. Teknik Olmayan Kayıplar.....	8
1.3 Ülkemizde Kayıp ve Kaçak Oranları	11
1.4 Değerlendirme.....	14

İKİNCİ BÖLÜM

ELEKTRİKTE TEKNİK OLMAYAN KAÇAK TESPİTİ İÇİN MAKİNE ÖĞRENMESİ UYGULAMALARI.....	16
2.1 Teknik Olmayan Kaçakların Tespitinde Kullanılan Yöntemler	16
2.2 Teknik Olmayan Kayıpların Tespiti ile İlgili Literatürde Yapılan Çalışmalar	17
2.1 Öznitelik Çıkarımları	30
2.2 Özniteliklerin Ölçeklenmesi – Veri Normalizasyonu.....	34
2.4.1 Z-Skoru Normalizasyonu	34
2.4.2 Minimum Maksimum Normalizasyonu	35
2.4.3 Ondalık Normalizasyonu.....	36

2.4.4 İstatiksel Kolon Normalizasyonu	37
2.4.5 Sigmoid Normalizasyonu	38
2.4.6 Medyan Normalizasyonu.....	38

ÜÇÜNCÜ BÖLÜM

KTS: VERİLERİN HAZIRLANMASI..... 39

3.1 Uygulama Sistemleri.....	39
3.2 Sistem Veri Tabanları ve Veri Yapıları	42
3.2.1 MBS – Müşteri Bilgi Sistemi Verileri	43
3.2.2 OSOS – Otomatik Sayaç Okuma Sistemi Verileri.....	43
3.3 Veri Toplama	47
3.4 ETL ve Veri Aktarımı	47
3.4.1 ETL Nedir?.....	47
3.4.2 Veri Ambarının Oluşturulması	48
3.5 Değerlendirme.....	49

DÖRDÜNCÜ BÖLÜM

KTS : MODELİN OLUŞTURULMASI 51

4.1. Mevcut Yaklaşım	51
4.1.1 Gerilim Uyarıları.....	51
4.1.2 Akım Uyarıları	52
4.1.3 Cosinus Uyarıları	53
4.1.4 Diğer Uyarılar	54
4.1.4.5 Demand Uyarısı	55
4.2 Önerilen Yaklaşım: Makine Öğrenmesi.....	56
4.2.1 Veri Ön işleme	57
4.2.2 Öznitelik Çıkarımı (Feature Extraction).....	61
4.2.3 Sınıflandırma ve Uygulama Geliştirme	70
4.3 Değerlendirme.....	73

BEŞİNCİ BÖLÜM

KTS: MODELİN DEĞERLENDİRİLMESİ..... 75

5.1 Karmaşıklık Matrisleri	76
5.2 K Katlamalı Çapraz Doğrulama.....	82

5.3 Precision-Recall ve ROC Eğrileri	84
5.4 Model İyileştirme Alternatif Yöntemler	86
5.4.1 Yinelemeli Öznitelik Eleme ile Öznitelik Seçimi.....	86
5.4.2 Algoritma Parametre Optimizasyonları.....	87
5.4.3 Metasezgisel Yöntemler ile Öznitelik Seçimi	89

ALTINCI BÖLÜM

SONUÇ VE ÖNERİLER.....	96
-------------------------------	-----------

KAYNAKÇA	98
-----------------------	-----------

EKLER.....	105
-------------------	------------

ÖZGEÇMİŞ.....	111
----------------------	------------

TABLolar LİSTESİ

Tablo 1.1: Dağıtım Şirketleri Bazında Hat ve Trafo Durumları-2019.....	5
Tablo 1.2: Dağıtım Şirketleri Kayıp Oranları (%)	13
Tablo 1.3: Dağıtım Sistemini Kullanan Tüketici Sayıları (Adet-%).....	14
Tablo 2.1: Örnek Profil Yükleme Verileri	18
Tablo 2.2: Master Parametre Seçim Örnekleri.....	20
Tablo 2.3: Master Parametre Seçimi	21
Tablo 2.4: Tarımsal Abonelerin Sınıflandırılması	22
Tablo 2.5: Ticari Abonelerin Sınıflandırılması	22
Tablo 2.6: Kırsal Abonelerin Sınıflandırılması.....	22
Tablo 2.7: WEKA çalışması kullanılan öznitelikler	23
Tablo 2.8: WEKA Tüm Öznitelikler Kullanıldığında Sonuçlar	23
Tablo 2.9: WEKA ile Öznitelik Seçimi Sonuçları	24
Tablo 2.10: WEKA Öznitelik Seçimi Sonrası Doğruluk ve Hata Oranları	25
Tablo 2.11: Bulanık, Mantıksal ve SVM Karmaşıklık Matrisleri.....	25
Tablo 2.12: Literatürdeki Model, Veri Seti Ve Performans Değerleri	26
Tablo 2.13: Bulanık Mantık Yaklaşımında Seçilen Parametreler.....	32
Tablo 3.1: KTS Veri Tabanları	42
Tablo 3.2: Örnek OBIS Kodları ve Açıklamaları	44
Tablo 3.3: LUNA Veri Yapıları ve Açıklamaları	46
Tablo 3.4: Veritabanı Bağlantısı Yöntem ve Araçları	48
Tablo 3.5: Kaçak Veri Ambarı Tabloları	48
Tablo 4.1: Kaçak Uyarıları Aralık Değerleri	56
Tablo 4.2: İl Değişkeni Kategorik-Nümerik Dönüşümü	60
Tablo 4.3: KTS Abone Grupları	61
Tablo 4.4: 2018 Yılı İllere Göre OSOS Kurulu Abone Sayıları	62
Tablo 4.5: Uygulama Sistemleri Öznitelik Çıkarımları	63
Tablo 4.6: KTS Değişkenleri ve Veri Tipleri	64
Tablo 5.1: KTS Modeli Veri Setleri ve Kaçak Dağılımları	75
Tablo 5.2: VeriSeti0 Doğruluk-Duyarlılık Değerleri.....	77
Tablo 5.3: VeriSeti1 Doğruluk-Duyarlılık Değerleri.....	81

Tablo 5.4: VeriSeti2 Doğruluk-Duyarlılık Değerleri.....	81
Tablo 5.5: VeriSeti3 Doğruluk-Duyarlılık Değerleri.....	81
Tablo 5.6: KTS Karmaşıklık Matrisi	82
Tablo 5.7: k=5 Katlamalı Değerlendirme Kat Detayları.....	83
Tablo 5.8: k=5 Katlamalı Karmaşıklık Matrisi.....	84
Tablo 5.9: k=10 Katlamalı Karmaşıklık Matrisi	84
Tablo 5.10: k=15 Katlamalı Karmaşıklık Matrisi	84
Tablo 5.11: Model Standart Sapma/AUC Değerleri	85
Tablo 5.12: LR-RFE Öznitelik Seçimleri Değerlendirme	86
Tablo 5. 13: GridSearch ile Seçilen Rassal Orman Hiper Parametreleri	88
Tablo 5.14: Rassal Orman - GridSearch Değerlendirme Sonuçları.....	88
Tablo 5.15: SVM Parametre Seçimine Göre Karmaşıklık Matrisi	89
Tablo 5.16: RF-BPSO ile Öznitelik Seçimi	94
Tablo 5.17: LR-BPSO ile Öznitelik Seçimi.....	94
Tablo 5.18: BPSO Öznitelik Seçimi Değerlendirme	95
Tablo 5.19: BPSO Performans Karşılaştırma	95

ŞEKİLLER LİSTESİ

Şekil 1.1: Elektrik Üretim, İletim ve Dağıtım.....	4
Şekil 1.2: Elektrik Dağıtım Şebekesi Yapısı.....	6
Şekil 1.3: 2019 Yılı Gerçekleşen ve Hedeflenen Kayıp Oranları	12
Şekil 2.1: Şüpheli bir aboneye ait profil yüklemesi	33
Şekil 2.2: Normal bir aboneye ait profil yüklemesi	33
Şekil 2.3: Z Skor Normalizasyon Örneği	35
Şekil 2.4: Minimum Maksimum Normalizasyon Örneği.....	36
Şekil 2.5: Ondalık Normalizasyon Örneği	37
Şekil 3.1: KTS Veri Ambarı Yapısı	49
Şekil 4.1: KTS Framework	57
Şekil 4.2: İstatiksel Veri Tipleri	64
Şekil 4.3: KTS Uygulama Akış Şeması	72
Şekil 5.1: Dengeli ve Dengesiz Veri Setleri.....	76
Şekil 5.2: Dengesiz Veriseti RF Karmaşıklık Matrisi.....	76
Şekil 5.3: Dengesiz Veriseti LR Karmaşıklık Matrisi	76
Şekil 5.4: Dengesiz Veriseti KNN Karmaşıklık Matrisi	77
Şekil 5.5: Dengesiz Veriseti SVC Karmaşıklık Matrisi.....	77
Şekil 5.6: Dengesiz Veriseti DT Karmaşıklık Matrisi	77
Şekil 5.7: VeriSeti1 RF Karmaşıklık Matrisi.....	78
Şekil 5.8: VeriSeti1 LR Karmaşıklık Matrisi.....	78
Şekil 5.9: VeriSeti1 KNN Karmaşıklık Matrisi	78
Şekil 5.10: VeriSeti1 SVM Karmaşıklık Matrisi	78
Şekil 5.11: VeriSeti1 DT Karmaşıklık Matrisi	79
Şekil 5.12: VeriSeti2 RF Karmaşıklık Matrisi	79
Şekil 5.13: VeriSeti2 LR Karmaşıklık Matrisi.....	79
Şekil 5.14: VeriSeti2 KNN Karmaşıklık Matrisi	79
Şekil 5.15: VeriSeti2 SVM Karmaşıklık Matrisi	79
Şekil 5.16: VeriSeti2 DT Karmaşıklık Matrisi	79
Şekil 5.17: VeriSeti3 RF Karmaşıklık Matrisi.....	80
Şekil 5.18: VeriSeti3 LR Karmaşıklık Matrisi.....	80

Şekil 5.19: VeriSeti3 KNN Karmaşıklık Matrisi	80
Şekil 5.20: VeriSeti3 SVM Karmaşıklık Matrisi	80
Şekil 5.21: VeriSeti3 DT Karmaşıklık Matrisi	80
Şekil 5.22: Precision Recall Eğrisi.....	85
Şekil 5.23: ROC Eğrisi.....	85
Şekil 5.24: Metasezgisel Algoritma Kategorileri.....	90



KISALTMALAR

- AKN: Abone Kredi Notu
- CBS: Coğrafi Bilgi Sistemleri
- Dicle EDAŞ: Dicle Elektrik Dağıtım Anonim Şirketi
- DC: Sözleşme Demand
- DCI: Difference Comparison based Inspection
- DF: Faturalanan Demand
- D_{max} : Maksimum Demand
- D_{ort} : Ortalama Demand
- EA: Etraflı Arama
- EDAŞ: Elektrik Dağıtım Anonim Şirketi
- EPDK: Enerji Piyasası Denetleme Kurulu
- EÜAŞ: Elektrik Üretim Anonim Şirketi
- EA: Exhaustive Search (Etraflı/Tam Arama)
- GA: Genetik Algoritmalar
- HRC: High Risk Customer (Yüksek Riskli Abone)
- KNN: K Nearest Neighbor (K En Yakın Komşu)
- KTS: Kaçak Takip Sistemi
- LF: Yükleme Faktörü
- LR: Lojistik Regresyon
- MBS: Müşteri Bilgi Sistemi
- NN: Neural Network (Yapay Sinir Ağları)
- NTL: Non Technical Loses
- OPF: Optimum Path Forest (En İyi Yol Ormanı)

OSOS: Otomatik Sayaç Okuma Sistemleri

OSOS1: Otomatik Sayaç Okuma Sistemi 1

OSOS2: Otomatik Sayaç Okuma Sistemi 2

PF: Güç Faktörü

P_{inst} : Kurulu Güç

PT: Trafo Gücü

RE: Reaktif Enerji

RMADS: Remote Meter Abnormality Detection Tool

SKTS: Saha Kaçak Takip Sistemleri

SOM: Self-Organization Map (Özdüzenleyici Haritalar)

SVC: Support Vector Classifier (Destek Vektör Sınıflandırma)

SVM: Support Vector Machine (Destek Vektör Makinası)

TM: Trafo Merkezi

TNB: Tenaga Nasional Berhard

TOE: Theft of Electricity (Elektrik Kaçak Kullanımı)

TEİAŞ: Türkiye Elektrik İletim Anonim Şirketi

TEDAŞ: Türkiye Elektrik Dağıtım Anonim Şirketi

TM: Trafo Merkezi

YSA: Yapay Sinir Ağları

GİRİŞ

Enerji sektörü önemi gittikçe artan bir sektör olup, tüm dünyada ülkelerin geleceklerinin belirlenmesi noktasında da kritik öneme sahiptir. Enerji kaynaklarına ulaşım, enerji kaynaklarının çeşitliliğinin sağlanması ve enerjide dışa bağımlılık oranının azaltılması, ülkelerin dünya üzerindeki statü ve konumlarının belirlenmesinde büyük öneme sahiptir. Enerji verimliliği de, en az enerji kaynaklarını arttırmak ve dışa bağımlılığı azaltmak kadar önemlidir. Sahip olduğumuz enerji kaynaklarını verimli kullanmak, enerjiye olan talep ihtiyacı açığını da azaltmış olacaktır.

Elektrik enerjisi, en önemli enerji türlerinden biridir. Elektrik üretiminden iletimine ve dağıtımına kadar geçen sürede, çeşitli kayıplar yaşanmaktadır. Bu noktalarda enerji kalite ve verimliliğini sağlamak, enerji kaybını azaltacaktır. Yaşanan bu kayıpları, teknik ve teknik olmayan kayıplar olarak ikiye ayırabiliriz. Teknik kayıpları azaltmak için, çeşitli teknik önlemler alınabilmektedir. Fakat teknik olmayan kayıp ve kaçakları önlemek, gerek bu noktadaki kaçakların tespiti, gerekse önlenmesi noktasında ciddi bir mücadele gerektirmektedir.

Elektrik dağıtım aşamasında, teknik olmayan kaçakların tespiti noktasında, özellikle kaçak kullanımın yoğun olduğu Brezilya, Malezya gibi ülkelerde bu konuda önemli çalışmalar yapılmaktadır. Özellikle son yıllarda yapay zeka teknolojileri ve bu teknolojinin alt kategorilerinden biri olan makine öğrenmesinin gelişimi ile birlikte, bu teknolojiler kaçak tespiti için kullanılmaya başlamıştır. Bu konuda yapılan farklı çalışmalar tez kapsamında incelenerek, ülkemizde kaçak tespiti için kullanılacak, makine öğrenmesi yöntemleri ile kaçak tespitine yönelik bir çalışma ortaya konmuştur.

Türkiye’de EPDK tarafından her yıl yayımlanan “Elektrik Piyasası Gelişim Raporu” incelendiğinde, kayıp kaçak kullanımının yoğun olarak üç bölgede toplandığı görülmektedir (EPDK, 2019). Diyarbakır, Şanlıurfa, Batman, Mardin, Siirt ve Şırnak illerini kapsayan Dicle Elektrik Dağıtım Bölgesi; Bitlis, Hakkari, Muş illerini kapsayan Van Gölü Elektrik Dağıtım Bölgesi ve Erzurum, Ağrı, Ardahan, Bayburt, Erzincan, Iğdır, Kars illerini kapsayan Aras Elektrik Dağıtım Bölgesi; Türkiye’de kayıp kaçağın en yüksek olduğu üç elektrik dağıtım bölgesidir. Elektrik kayıp kaçak oranı Dicle EDAŞ için %51,32, Van Gölü EDAŞ için 47,56 ve Aras EDAŞ için %21,64 olarak hesaplanmıştır (EPDK, 2019).

Tez kapsamında gerçekleştirilen bu çalışma, Türkiye’de kayıp kaçağın en yüksek olduğu Dicle Elektrik Dağıtım Bölgesi’nde gerçekleştirilmiştir. Ayrıca bu çalışma bu alanda, yapay zeka makine öğrenmesi yöntemlerinin kullanıldığı Türkiye’deki ilk çalışmadır. Bu alanda ülkemizde yapılan ilk çalışma olarak, sonraki çalışmalara öncülük etmesi ve fayda sağlaması ile birlikte, çalışmanın kaçak yoğun bir bölgede gerçekleştirilmesi, tez çalışmasının en büyük motivasyon kaynağıdır.

Tezin ilk bölümünde, elektrik dağıtım şebekesi yapısı, elektrikte kayıp kaçak kavramı, teknik ve teknik olmayan kaçaklar ve kaçak tipleri ortaya konarak ülkemizdeki elektrik dağıtım bölgeleri bazında kaçak oranlarının değişimlerine yer verilmiştir.

Çalışmanın ikinci bölümünde elektrikte teknik olmayan kaçakların tespitinde literatürde yapılan çalışmalar ve kullanılan yöntemlere yer verilmiştir. Kaçak tespitinde makine öğrenmesi kapsamında hangi özniteliklerin kullanıldığı, kullanılan özniteliklerin detaylandırılarak değerlendirilmesi ve özniteliklerin ölçeklenmesi kapsamında kullanılan algoritmalar ortaya konmuştur.

Üçüncü bölümde, geliştirilen modelde kullanılacak verilerin hazırlanması amacıyla yapılan çalışmalar yer almaktadır. Dicle Elektrik Dağıtım bölgesinde kullanılan uygulama sistemleri, bu sistemlerde tutulan veriler ve teknik altyapıları çıkarılmıştır. Tüm bu sistemlerden alınan veriler, ETL yöntemleri ile işlenerek, makine öğrenmesinde kullanılan veri ambarı yapısı oluşturulmuştur. Birden fazla uygulama sisteminin varlığı, her bir uygulama sisteminin veritabanının farklı olması ve uygulama sistemleri veritabanlarında tutulan verilerin kayıt desenlerinin öğrenilerek, bu verilerin işlenmesi ve veri migrasyonu kapsamında veri ambarının oluşturulması çalışmaları, projenin en önemli ve zaman maliyeti en yüksek bölümünü oluşturmaktadır.

Dördüncü bölümde kaçak tespiti için kullanılan mevcut yaklaşım ve önerilen yaklaşıma yer verilmiştir. Mevcut uzman yaklaşım ve bu yaklaşımda kullanılan parametreler detaylı olarak aktarılmıştır. Önerilen makine öğrenmesi yaklaşımı ve bu kapsamda oluşturulan KTS Framework yapısı açıklanmıştır. KTS oluşturulması adımlarından veri önışlemenin alt adımları olan, veri temizliği ve veri normalizasyonu çalışmalarına yer verilmiştir. Ay bazlı ve master olmak üzere, iki farklı kategoride toplam 144 öznitelik çıkarımı gerçekleştirilmiştir. Bölüm sonunda, KTS Uygulama

Akış Şeması, uygulama pseudo kodu ile birlikte, KTS modelinde kullanılan algoritmalara yer verilmiştir.

Beşinci bölümde tez kapsamında geliştirilen KTS modeli değerlendirilmiştir. Değerlendirme noktasında, karmaşıklık matrisi, k-kat çapraz doğrulama teknikleri, precision-recall ve ROC eğrileri kullanılmıştır. KTS algoritmaları performans sonuçları değerlendirilerek, model iyileştirme kapsamında; yinelemeli öznelik seçimi, algoritma parametre optimizasyonları ile metasezgisel yöntemlerden BPSO ile öznelik seçimi gibi alternatif yöntemler kullanılmıştır. Model iyileştirme sonucunda elde edilen performans değerleri karşılaştırılmıştır.

Son bölümde tez çalışmasının kısa bir özeti verilerek, elde edilen sonuçlar değerlendirilmiştir. Çalışmanın literatüre ve bu alanda ülkemizde yapılacak çalışmalara katkısı ortaya konmuştur.

BİRİNCİ BÖLÜM

ELEKTRİKTE KAYIP KAÇAK VE ÜLKEMİZDEKİ DURUMU

1.1 Elektrik Dağıtım Şebekesi

Elektrik enerjisi depolanamayan bir enerji olup, üretildiği noktalardan tüketim noktalarına, üretim sonrasında iletilmesi gerekmektedir.

Elektrik şebekesini üçe ayırabiliriz:

- Üretim yapan enerji santralleri
- Enerji iletim hatları
- Enerji dağıtım hatları



Şekil 1.1: Elektrik Üretim, İletim ve Dağıtım

Kaynak: <http://www.teias.gov.tr>

Üretim santrallerinde üretilen elektrik enerjisi, talep merkezlerine enerji iletim hatları ile aktarılmaktadır. Türkiye’de enerji üretimi bir kamu kuruluşu olan EÜAŞ – Elektrik Üretim Anonim Şirketi tarafından yönetilirken, elektrik iletimi TEİAŞ – Türkiye Elektrik İletim A.Ş., dağıtımı ise TEDAŞ – Türkiye Elektrik

Dağıtım A.Ş ile birlikte özelleştirme sonrası toplam 21 bölgedeki elektrik dağıtım kuruluşları tarafından gerçekleştirilmektedir. Türkiye’de özelleştirme sonrası dağıtım firmaları ve bölgeleri kapsamındaki iller, Ek-1’de verilmiştir. Şekil 1.1, basitçe Türkiye’deki elektrik üretim, iletim ve dağıtım yapısını göstermektedir. EPDK, 2019 yılında yayımladığı raporda, Türkiye’nin 21 bölgesinde görevli elektrik dağıtım kurumları ve dağıtım bölgesindeki hat uzunluğu, trafo kapasitesi ve trafo sayısı bilgilerini paylaşmıştır. Tablo 1.1’de görülebileceği üzere, Dicle Elektrik dağıtım bölgesi, trafo sayısı bakımından 67.761 trafo ile Türkiye’de en fazla trafoya sahip olan bölgedir.

Tablo 1.1: Dağıtım Şirketleri Bazında Hat ve Trafo Durumları-2019

Dağıtım Şirketi	Hat uzunluğu (km)	Trafo Kapasitesi (MVA)	Trafo Sayısı (Adet)
ADM	69.538	8.263	21.888
AKDENİZ	66.755	7.670	19.471
AKEDAŞ	26.524	3.804	12.440
ARAS	56.971	3.043	13.734
BAŞKENT	113.814	15.327	33.530
BOĞAZIÇI	47.138	17.850	14.682
ÇAMLIBEL	46.638	2.621	12.457
ÇORUH	59.404	3.239	12.376
DİCLE	75.984	12.494	65.368
FIRAT	47.812	3.286	12.635
GDZ	57.804	15.151	33.214
İ. ANADOLU	24.715	8.640	7.767
KAYSERİ VE CİVARI	25.220	2.579	7.840
MERAM	89.877	11.707	56.990
OSMANGAZİ	49.472	7.407	25.698
SAKARYA	37.386	7.908	17.169
TOROSLAR	93.053	15.821	47.267
TRAKYA	28.538	5.140	11.431
ULUDAĞ	50.798	9.648	25.875
VANGÖLÜ	38.839	2.692	11.954
YEŞİLIRMAK	83.892	4.770	22.147
Genel Toplam	1.190.169	169.060	485.933

Kaynak: (EPDK, 2019)

Trafolara gelen elektrik, fiderler aracılığıyla TM – Trafo Merkezlerine, trafo merkezlerinden de kutular aracılığıyla abonelere ulaştırılmaktadır. Şekil 1.2, elektrik dağıtım şebekesi yapısını göstermektedir.



Şekil 1.2: Elektrik Dağıtım Şebekesi Yapısı

Şebekeleri kullandıkları gerilime göre 4 gruba ayırabiliriz:

- Alçak gerilim şebekeleri (1-1000V arası)
- Orta gerilim şebekeleri (1kV-35kV arası)
- Yüksek gerilim şebekeleri (35kV-154kV arası)
- Çok yüksek gerilim şebekeleri (154kV'dan fazla)

1.2. Elektrikte Kayıp Kaçak

Üretilen elektrik enerjisinin tamamı abonelere ulaştırılamamaktadır, dolayısıyla dağıtım sisteminde hatırı sayılır oranda enerji kaybı oluşmaktadır (Navani, Sharma, ve

Sonal, 2012). Dağıtım hatlarındaki elektrik enerjisi kayıplarını iki ana gruba ayırabiliriz (Davidson, vd., 2002):

- Teknik Kayıplar
- Teknik Olmayan Kayıplar

Teknik olmayan enerji kaybı maliyeti, toplam kayıp enerjiden teknik kayıpların çıkarılması ile elde edilir (Davidson, 2003).

$$C_{\text{Teknik Olmayan Kayıp}} = C_{\text{Toplam Kayıp}} - C_{\text{Teknik Kayıp}}$$

Teknik kayıplar, tespit edilebilir kayıplar iken, teknik olmayan kayıplar ise farklı kaynaklara dayanan ve tespiti zor olan kayıplardır. Teknik olmayan kayıpların, toplam kayıp içerisindeki payı genellikle daha fazla olmaktadır.

1.2.1. Teknik Kayıplar

Elektrik üretim, iletim ve dağıtım aşamalarında, enerji kayıplarının olması olağandır. Bu kayıplar, ortaya konabilen kayıplar olup, teknik kayıplar olarak isimlendirilir (Angelos, vd., 2011). Teknik kayıplar, elektrik şebekesindeki iletim hattı kayıpları, bakır direnci ve transformatörlerdeki demir kayıplarını içermektedir (Davidson, vd., 2002).

Teknik kayıpları aşağıdaki tiplere ayırabiliriz (Navani, Sharma, ve Sonal, 2012).

- İletkenlerin sınırlı direncinden kaynaklanan, tüm indüktörlere ait bakır kayıpları
- Dielektrik malzeme ile iletkenler arasındaki ısınma etkisinden kaynaklı dielektrik kayıpları
- İndüktif ve radyasyon kayıpları

Teknik kayıpların nedenlerini ise aşağıdaki gibi sıralayabiliriz (A.g.e.):

- Harmonik bozulma
- Abonede uygun olmayan topraklama
- Uzun tek faz hatlar
- Dengesiz Elektrik Yükleme
- Aşırı yükleme ve düşük voltaj nedeniyle oluşan kayıplar
- Düşük standartta ekipmanların kullanılması

Dağıtım sistemlerinin kurgulanması, planlanması ve boyutlandırılması, teknik kayıpların düşürülmesinde büyük öneme sahiptir. Ayrıca dağıtım şebekelerinde zamanında gerçekleştirilecek planlı ve plansız bakımlar, aynı şekilde teknik kayıpların

azaltılmasını sağlayacaktır. Tedarik edilen enerji kalitesi de teknik kayıplarda etkili olup, aynı şekilde kaliteli bir dağıtım şebekesi altyapısı ile de verilecek enerjinin kalitesi arttırılmış olacaktır. (Angelos, vd., 2011)

1.2.2. Teknik Olmayan Kayıplar

Teknik olmayan kayıplar, birçok ülkede toplam elektrik kayıpları içinde, teknik kayıplara göre yüksek bir orana sahiptir. Bu yüzden kayıp kaçakla mücadele açısından, teknik kayıplardan önce teknik olmayan kayıpları düşürmek, daha verimli olacaktır. (Amin, Schwartz ve Tembine, 2012)

Teknik olmayan kayıpları iki ana grupta toplayabiliriz.

- Şebeke ekipmanları ile kaçak bağlantı aracılığıyla gerçekleştirilen kaçak ve sahtekârlık
- Ölçüm ekipmanlarındaki problemlerden kaynaklı ölçüm hataları

Kaçak bağlantılar ile kullanılan elektrik, hesaplanan fakat faturalandırılmayan elektrik tüketimi iken, sayaç gibi ölçüm ekipmanlarındaki sorunlardan kaynaklı elektrik kullanımı ise yanlış veya eksik faturalandırılan kayıptır. Faturalandırılan fakat ödenmeyen elektrik de teknik olmayan kayıp olarak değerlendirilir (CIRED, 2017).

Enerji hattından kaçak yollardan direkt olarak enerji çekilmesi ve sistemde kayıtlı abonelerin sayaçlara müdahale ederek yaptığı sahtekârlık sonucunda, tahakkuk eden fatura sayısı azalmaktadır. Tahakkuk eden fatura sayısının azalması, kaçak oranının artışına neden olmaktadır.

Teknik olmayan kayıplara ait bazı durumları aşağıdaki gibi sıralayabiliriz. (Glauner, vd., 2017)

- Düşük tüketim amaçlı sayaca müdahale
- Sayacı devre dışı bırakarak, elektrik şebekesinden direkt bağlantı sağlama
- Saha çalışanlarının bilinçli olarak yanlış okuma yapması
- Bozuk veya hatalı sayaçlar
- Ölçülemeyen elektrik
- Sayaç okumada yapılan teknik veya kullanıcı kaynaklı veri işleme/faturalama hataları

Teknik olmayan kayıp kaçak tespitinde, “Uzman Sistemler” ve “Makine Öğrenmesi” yöntemleri kullanılmaktadır. Uzman sistemler, saha tecrübeleri ile oluşan kurallar ile oluşturulan algoritmaları kullanan statik sistemlerdir. Makine öğrenmesi yöntemi ise, sabit bir algoritma olmadan oluşturulan model ile girdi olarak kullanılan verilerden sistemin öğrenmesini amaçlamaktadır. (A.g.e.)

1.2.2.1 Teknik Olmayan Kaçak Tipleri

Elektrikte teknik olmayan kayıpları, kaçak olarak değerlendirebiliriz. Kaçak elektrik kullanımını 4 farklı tipe ayırabiliriz. (Smith, 2004)

- Sayaca Müdahale (Sahtekârlık)
- Kaçak Bağlantı (Elektrik Çalma)
- Usulsüz Faturalama
- Ödenmeyen Faturalar

1.2.2.2 Sayaca Müdahale

Abonelerin elektrik tüketimlerinin ölçümünde, elektrik sayaçları kullanılmaktadır. Sayaçlarda okunan değerler, sayaç çarpanı ve varsa diğer etkenlerle birlikte değerlendirilerek elektrik tüketimleri hesaplanmakta ve abonelere bu tüketimleri fatura edilmektedir.

Sayaçların yapıları farklı olmakla birlikte, tüketimler genellikle elektrik kullanımına göre artan bir değer olan endeks değeri ile takip edilmektedir. Her sayaç ölçümünde alınan son endeks değeri ile bir önceki okunan ve sistemde kaydedilen ilk endeks değeri arasında hesaplanarak, abonenin tarifesine göre belirlenen birim fiyat ile çarpılarak abonelere ait fatura bedelleri hesaplanmaktadır.

Ölçüm hataları, sayaçtan gerçekten geçen enerji ile sayaçtan okunan enerji arasındaki farklardan oluşmaktadır. Hatalı ölçüm ekipmanları, bozuk sayaçlar veya hatalı sayaç ayarları, okuma anında oluşan anlık problemler, sayaçlardan yapılan ölçümlerin yanlış olmasına neden olur.

Sayaca müdahale, en çok karşılaşılan kaçak kullanım yöntemlerindedir. Çarpanları değiştirme, endekslerin geriye alınması gibi yöntemlere sayaca müdahale edilerek beklenenden az fatura çıkması sağlanmaktadır.

Dicle EDAŞ bölgesinde sahada yapılan çalışmalarda görüldüğü üzere, problemlili olduğu düşünölen sayaçlar saha personelleri tarafından sökülmemekte, Sayaç Ayar Servisi'nde incelenmemekte, müdahale tespit edilen sayaçlar için gerekli işlemler yapılmaktadır.

Yeni nesil OSOS kapsamında takılan akıllı sayaçlar ile birlikte, sayaçlara yapılan müdahaleler merkezden anlık olarak izlenebilmektedir. Akıllı sayaç takılan abonelerde, sistemlerde biriken büyük veriler üzerinde öğrenme ile, anlık tespitlerde bulunmak mümkündür.

1.2.2.3 Kaçak Bağlantı

Kaçak bağlantılar ile sayaçlar devre dışı bırakılmakta veya sayaç olmadan elektrik kullanılmaktadır. Sayacı olmayan ve herhangi bir abonelik yaptırmamış, sistemde kayıtlı bilgisi olmayan abonelerin kaçak kullanım tespiti, ancak sahada yapılacak fiziksel çalışmalar sonunda sağlanabilmektedir. Sistemlerde herhangi bir aboneliği veya tüketim verisi olmayan bu tip aboneler, kapsam dışında tutulmuştur.

Aboneliği ve sayacı olan, kaçak bağlantı sağlayarak elektrik kullanan aboneler için sayaç okuma değerleri çok düşük veya 0 endeks olarak sistemde yer alacaktır. Bu kayıtlar, sistem tarafından değerlendirilerek, gerçekten elektrik kullanıp kullanmadığında göre (tatilde olunan dönemlerde az veya hiç elektrik kullanmamak gibi) kaçak tespiti parametrelerinde kullanılabilir.

1.2.2.4. Usulsüz Faturalama

Usulsüz faturalama, sayaca fiziksel müdahale olmaksızın, kötü niyetli sistem kullanıcıları tarafından veya yanlışlıkla yapılan işlemler sonucunda sistemde gerçekleştirilen eksik faturalama işlemleridir. Sistemdeki endeks, çarpan, tarife gibi faturayı etkileyen parametrelere dışarıdan müdahale edilerek oluşturulan düşük tüketimli faturalar, usulsüz faturalama kapsamına girmektedir. Oluşturulan faturaların iptal edilerek silinmesi de bu kapsamda değerlendirilmektedir. Bu yöntemlerle yapılan usulsüzlükler sonucunda, aboneye fatura çıkmamakta veya düşük tüketimli fatura çıkmaktadır.

1.2.2.5 Ödenmeyen Faturalar

Ödenmeyen faturalar, kaçak kullanımın devam etmesi anlamına gelmektedir. Türkiye’deki mevzuata göre, abone faturalarını belirli bir dönem boyunca ödemediğinde, abone bu konu ile ilgili uyarılmaktadır. Kesme İhbarnamesi olarak isimlendirilen bu uyarı sonrasında hala faturalar ödenmediyse, abonelerin elektriği kesilmektedir. Bu işlem sırasında sahada saha çalışanları tarafından ilgili abonenin sayacı mühürlenmekte ve enerjisi kesilmektedir.

Abone kesik olan enerjisini sayacın mührünü sökerek açıp, elektrik kullanmaya devam ederse kaçak kullanım statüsüne düşmektedir. Bu tip abonelere, Elektrik Dağıtım firması tarafından usulsüz kullanım kaynaklı cezalı kaçak faturaları düzenlenmektedir. Bu ceza ilk tespit normal tüketimin 1,5 katı, sonraki tespitlerde ise 2 katı olarak aboneye yansıtılmaktadır.

Bu tip faturaların varlığı ve sayısı, abonelerin zamanında kaçak kullanımı olup olmadığını göstermektedir, bu bilgi abonelerin davranışlarının ortaya konması anlamında önem taşımaktadır. Özellikle Dicle EDAŞ Dağıtım bölgesi olan Güneydoğu Anadolu Bölgesi’ndeki abonelerde bu tip cezalı kaçak faturalarının çokluğu, kaçak kullanımının bir alışkanlık haline geldiğini göstermektedir.

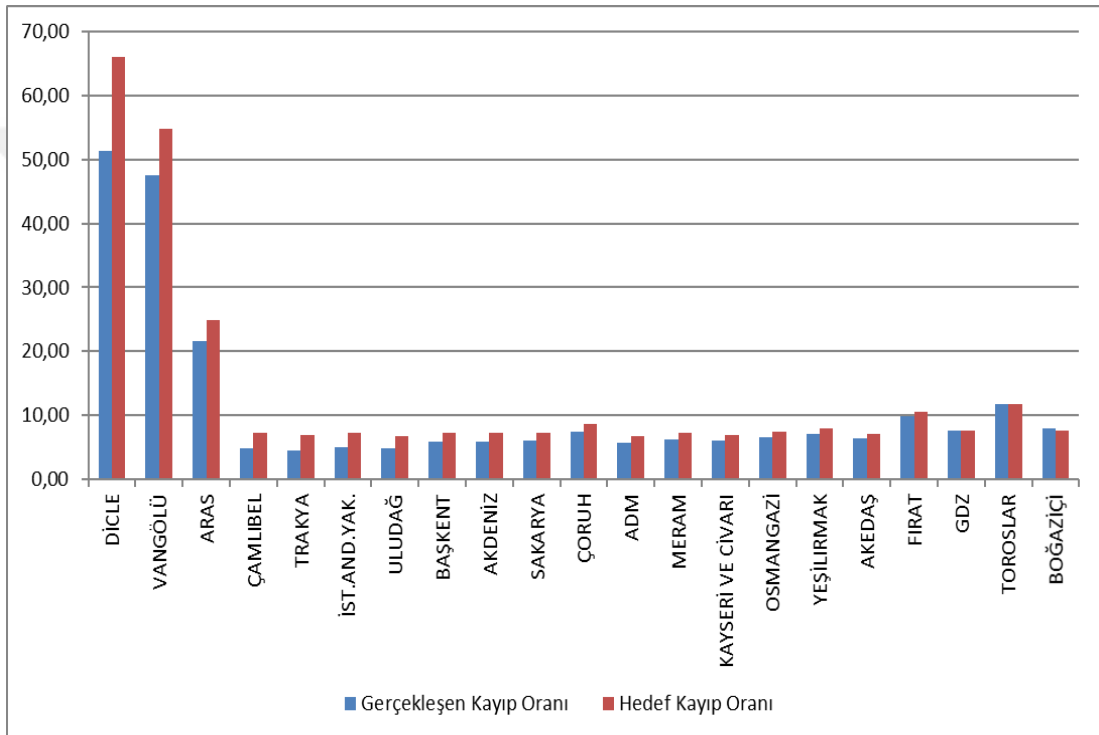
Borcundan dolayı enerjisi kesik olan bir abone, borçlarını ödediğinde sistemde oluşan “Açma Emirleri” sonucunda saha personelleri tarafından elektriği açılmaktadır. Abonenin durumu ise “Kesik” statüsünden, “Normal” statüsüne geçmektedir.

1.3 Ülkemizde Kayıp ve Kaçak Oranları

Birçok ülkede olduğu gibi, elektrikte kayıp ve kaçak, ülkemizin önemli sorunlarından biridir. Enerji piyasasını Türkiye’de düzenleme ve denetleme yetkisine sahip olan kurum olan EPDK – Elektrik Piyasası Düzenleme Kurumu, her yıl elektrik piyasası ile ilgili gelişim raporu yayımlamaktadır. EPDK, 2019 raporuna göre, 2019 yılına ait gerçekleşen ve hedeflenen elektrik kayıp kaçak oranları Şekil 1.3’te verilmiştir. İlgili rapor incelendiğinde, en yüksek kayıp oranları %54,94 ile Dicle, %49,16 ile Van gölü ve %23,55 ile Aras bölgelerinde gerçekleşmiştir. Tablo 1.2 ise, EPDK, 2019’a göre hedef, gerçekleşen kayıp oranları ile 2018 ve 2019 gerçekleşen kayıp oranlarını vermektedir. Projemizin de uygulama alanı olan Dicle EDAŞ bölgesinde, 2018’de %54,94 e düşürülen kayıp kaçak oranının, 2019 yılında %51,32 olarak gerçekleştiği

görülmüştür. Bu durum da özellikle Dicle ve Van gölü bölgelerinde yaklaşık olarak her iki kişiden birinin kaçak elektrik kullanımı yaptığını göstermektedir (Teknik kayıplar hariç).

Türkiye’de elektrik kullanan abonelerin bölge bazında sayılarının 2017-2019 arası değişimleri ise Tablo 1.3’te gösterilmiştir (EPDK, 2019). Kaçak kullanımı tespit edilen ve aboneliği olmayan abonelerin aktif abone yapılması, kaçak oranını düşürürken aynı zamanda aktif abone sayısını da arttırmaktadır.



Şekil 1.3: 2019 Yılı Gerçekleşen ve Hedeflenen Kayıp Oranları

Kaynak: (EPDK, 2019)

Tablo 1.2: Dağıtım Şirketleri Kayıp Oranları (%)

Dağıtım Şirketi	Gerçekleşen Kayıp Oranı			Hedef Kayıp Oranı	Hedef-Gerçekleşen Kayıp Oranı Farkı
	2018	2019	Değişim	2019	2019
DİCLE	54.94	51.32	3.62	65.99	14.67
VANGÖLÜ	49.16	47.56	1.60	54.74	7.18
ARAS	23.55	21.64	1.91	24.97	3.33
ÇAMLİBEL	5.08	4.75	0.33	7.31	2.56
TRAKYA	4.37	4.49	-0.12	6.87	2.38
İST.AND.YAK.	6.04	5.05	0.99	7.31	2.26
ULUDAĞ	4.20	4.82	-0.62	6.73	1.91
BAŞKENT	6.11	5.89	0.22	7.33	1.44
AKDENİZ	5.78	5.91	-0.13	7.31	1.40
SAKARYA	6.52	5.96	0.56	7.28	1.32
ÇORUH	7.85	7.38	0.47	8.68	1.30
ADM	5.53	5.63	-0.10	6.72	1.09
MERAM	6.69	6.22	0.47	7.30	1.08
KAYSERİ VE CİVARI	6.57	6.02	0.55	6.94	0.92
OSMANGAZİ	6.41	6.61	-0.20	7.43	0.82
YEŞİLIRMAK	7.63	7.13	0.50	7.94	0.81
AKEDAŞ	5.51	6.46	-0.95	7.08	0.62
FIRAT	10.32	9.93	0.39	10.53	0.60
GDZ	6.63	7.55	-0.92	7.66	0.11
TOROSLAR	11.85	11.76	0.09	11.74	-0.02
BOĞAZIÇI	6.59	7.96	-1.37	7.63	-0.33

Kaynak: (EPDK, 2019)

Tablo 1.3: Dağıtım Sistemini Kullanan Tüketici Sayıları (Adet-%)

Dağıtım Şirketi	Tüketici Sayısı			2018-2019 Değişim (%)
	2017	2018	2019	
BOĞAZIÇI	4.910.115	5.029.273	5.134.980	2.10
BAŞKENT	4.218.812	4.199.135	4.278.785	1.90
TOROSLAR	3.805.358	3.877.734	3.997.925	3.10
GDZ	3.165.847	3.312.515	3.421.183	3.28
ULUDAĞ	3.092.199	3.189.831	3.285.323	2.99
İ. ANADOLU	2.834.800	2.854.123	2.928.586	2.61
AKDENİZ	2.076.167	2.163.700	2.220.744	2.64
MERAM	2.027.866	2.094.254	2.169.733	3.60
YEŞİLIRMAK	1.998.388	2.073.523	2.147.103	3.55
ADM	1.834.517	1.902.358	1.966.146	3.35
DİCLE	1.757.417	1.838.929	1.934.641	5.20
SAKARYA	1.742.528	1.831.677	1.905.188	4.01
OSMANGAZİ	1.720.217	1.786.452	1.839.538	2.97
ÇORUH	1.303.909	1.359.781	1.389.748	2.20
TRAKYA	1.042.995	1.089.247	1.125.936	3.37
ARAS	963.072	1.002.743	1.029.869	2.71
ÇAMLİBEL	946.381	974.544	1.005.503	3.18
FIRAT	917.488	957.081	977.424	2.13
KAYSERİ VE CİVARI	695.210	718.723	748.858	4.19
AKEDAŞ	678.775	713.348	740.416	3.79
VANGÖLÜ	650.555	684.366	710.697	3.85
Genel Toplam	42.382.616	43.653.337	44.958.326	2.99

Kaynak: (EPDK, 2019)

1.4 Değerlendirme

Farklı yöntemlerle üretilen elektrik enerjisi, üretildiği gibi abonelere ulaştırılmaktadır. Elektrik depolama imkanlarının yetersizliği ve bu alanda henüz alınacak çok yol olması, üretilen elektriğin iletim hatları ile dağıtım bölgelerine minimum kayıp ve

maksimum kalitede ulařtırılmasını zorunlu kılmaktadır. Dağıtım řebeke altyapısı ile abonelere ulařtırılan elektrik, teknik ve teknik olmayan kayıplarla karřılařmaktadır.

Bu bۆlümde elektrikte kayıp kaak kavramı, teknik ve teknik olmayan kaaklar ve tipleri ۆzerinde durularak, ۆlkemizdeki elektrik kayıp kaađının durumu hakkında bilgi verilmiřtir. ۆzellikle tez alıřmasının geliřtirildiđi Dicle bۆlgesinin, Tۆrkiye'deki en bۆyۆk řebeke altyapısına ve kayıp kaak oranında en yۆksek orana sahip olması, tez alıřması iin dođru bۆlgenin seildiđini ortaya koymaktadır. Sonraki bۆlümde bu alanda literatۆrde yapılan alıřmalar, ۆzellikle de makine ۆđrenmesi yۆntemleri ile ۆretilen özۆmler ۆzerinde durulmuřtur.



İKİNCİ BÖLÜM

ELEKTRİKTE TEKNİK OLMAYAN KAÇAK TESPİTİ İÇİN MAKİNE ÖĞRENMESİ UYGULAMALARI

Türkiye’de olduğu gibi diğer birçok ülkede elektrikte kayıp kaçak, ülkelerin önemli sorunlarından biridir. Kayıp kaçak oranının düşürülmesi noktasında yapılan akademik çalışmalar, bu bölümde incelenmiştir. Elektrikte teknik kaçakların tespiti, teknik olmayan kaçakların tespitine göre daha kolaydır ve genellikle teknik unsurlara dayanır. Teknik kaçakların tespitinde, kuralların daha net olması nedeniyle, uzman sistem yaklaşımı ile hızlı sonuçlara ulaşılabilmektedir. Teknik olmayan kaçakların tespitinde ise, uzman sistemler belirli bir noktaya kadar çözüm üretebilmektedir. Kaçağı oluşturan unsurların fazlalığı, birbirleri ile ilişkisi, abone davranış modellerindeki farklılıklar ve bu farklılıklara ait detaylı veriler, makine öğrenmesi yöntemlerinin uygulanmasını mümkün kılmaktadır. Bu alanda abonelere ait yıllara dair büyük verilerin oluşması, makine öğrenmesi algoritmalarının kullanımına uygun bir altyapı oluşturmaktadır. Bu bölümde, tez kapsamında yer alan teknik olmayan kayıpların tespiti için uygulanan makine öğrenmesi algoritmaları ve bu algoritmalar ile yapılan çalışmaların sonuçlarına yer verilmiştir.

Özellikle Hindistan ve Brezilya gibi elektrikte kayıp kaçak ile mücadele eden ülkelere ait akademik çalışmalar detaylı olarak incelenmiş olup, Türkiye’de Dicle Elektrik dağıtım bölgesi ile karşılaştırılmıştır.

2.1 Teknik Olmayan Kaçakların Tespitinde Kullanılan Yöntemler

Teknik Olmayan Kayıp Kaçak tespitinde literatürde iki yöntem kullanılmıştır (Glauner, vd., 2017):

- Uzman Sistemler
- Makine Öğrenmesi

Uzman sistemler, genellikle saha tecrübelerinden oluşan bilgiler ışığında, karar destek sistemi olarak kullanılmak üzere el ile geliştirilen sistemlerdir. Uzman sistemler, statik sistemlerdir.

Makine öğrenmesi sistemleri ise, makinaların veriler üzerinden öğrenmesine ve herhangi bir programlama yapmadan sistemin kendi kendini geliştirmesine imkân tanıyan yapılardır.

Literatürde bu kapsamda kullanılan yapay zekâ algoritmalarını aşağıdaki gibi özetleyebiliriz.

- SVM – Support Vector Machine (Destek Vektör Makinası)
- LR - Lojistik Regresyon
- KNN – K-En Yakın Komşu
- K-Ortalama Kümeleme
- DT-Karar Ağaçları
- YNN - Yapay Sinir Ağları
- SOM – Self Organization Map (Özdüzenleyici Haritalar)
- GA-Genetik Algoritmalar
- OPF- Optimum Path Forest (En İyi Yol Ormanı)

2.2 Teknik Olmayan Kayıpların Tespiti ile İlgili Literatürde Yapılan Çalışmalar

İletim ve dağıtım kayıpları yüzdelerine göre sistemler, aşağıdaki gibi gruplandırılabilir (Smith, 2004).

- Çok Verimli Sistemler (Kayıp Kaçak Oranı \leq 6)
- Az Verimli Sistemler ($9 <$ Kayıp Kaçak Oranı \leq 12)
- Verimsiz Sistemler (Kayıp Kaçak Oranı $>$ 15)

Profil yüklemesi ile abonelere ait veri setleri üzerinde işlemler gerçekleştirilir. Farklı abone tiplerinin farklı davranışları olduğu ortaya konmuştur. Tarımsal Sulama, Mesken, Ticarethane gibi abone gruplarının farklı davranışlar sergilediği ortaya konmuştur. Tablo 2.1’de, teknik olmayan kayıpların tespiti amacıyla gerçekleştirilen Profil Yükleme aşamasında seçilen öznitelikler ve değerleri listelenmiştir (Nizar, Zhao, ve Dong, 2006).

2010 yılında Malezya’da TNB aboneleri üzerinde yapılan çalışmada, Kayıp Kaçak Tespit Altyapısı (NTL Detection Framework) üzerinde, Uzaktan Okuma Sayaçları Anormallik Tespit Aracı (RMADS – Remote Meter Abnormality Detection Tool) geliştirilmiştir. Bu çalışmada diğerlerinden farklı olarak, fabrikalar gibi yüksek tüketimli ve maddi anlamda getirisi yüksek aboneler üzerinde çalışılmıştır. Abonelerin davranışları analiz edilerek, sonuçlar üzerinden çalışmalar ilerletilmiştir. (Nagi, vd., 2010) Çalışmadaki ana adımlar aşağıdaki gibidir:

- Aylık olarak kullanıcı yükleme profili oluşturulması
- Özniteliklerin çıkarımı
- Yüklenen profillerin özelliklerine göre sınıflandırılması
- Her sınıf için özel tasarlanmış stratejilerin uygulanması

Tablo 2.1: Örnek Profil Yükleme Verileri

Seçilen Öznitelik	Öznitelik Değeri
Müşteri Sınıfı	Mesken, Ticarethane, Endüstriyel
Lokasyon	Kırsal, Şehir
Voltaj Seviyesi	Alçak Gerilim, Yüksek Gerilim
İklim Tipi	Yağmurlu, Sıcak
Hava Şartları	Nem, Sıcaklık, Hava Basıncı, Rüzgâr Hızı, Görüş Mesafesi
Gün Tipi	Hafta içi, Hafta Sonu, Tatil Günü

Kaynak: (Nizar, Zhao, ve Dong, 2006)

Çalışma sonucunda Teknik Olmayan Kayıp Tespitine giren aboneler davranışlarına göre 5 farklı tipte sınıfa ayrılmıştır. Tip1 ve Tip2 anomali gösteren, Tip3, Tip4 ve Tip5 ise sahtekarlık yapan sınıfa dahil edilmiştir.

Sayaca müdahale şeklinde gerçekleştirilen teknik olmayan kaçakların abone bazlı tespitinde, çeşitli algoritmalar kullanılmıştır. Fark Karşılaştırma Tabanlı Denetim Algoritması (DCI-Difference Comparison based Inspection) ile elektrik şebekesinde birbirine yakın bölgelerde bulunan abonelere ait sayaçlardan, problemlili sayaçlar tespit edilmeye çalışılmıştır (Xia, vd., 2015).

Fark Karşılaştırma Tabanlı Denetim Algoritmasında her bir alt ağaç yapısındaki kaçak elektrik miktarı fark karşılaştırmaları yapılarak kaçak şüphesi olan sayaçlar belirlenmiştir. Sayaçlar ikili ağaç yapısında düğümler olarak kabul etmiştir. Çalışmada ikili kodlama (Binary Coding) yapısı ve permütasyon işlemleri ile kayıp kaçak tespit tahminleri gerçekleştirilmiştir. Bir grup sayaç için, sayaçların her alt grubunda farklı sayaçlar kullanılmaktadır. Gözlemciler ve sayaçlar, 0 ve 1 rakamlarıyla ikili olarak kodlanmıştır (A.g.e).

Fark Karşılaştırma Tabanlı Denetim Algoritması'nda permütasyon ile anormal sayaçlar tespit edilmektedir. Çalışmada gözlemci sayısının, tanımlı sayaç sayısının yarısı kadar olması gerekmektedir. Bu şart uygulama maliyeti açısından efektif olmayan bir durum oluşturmaktadır.

Kaçak kullanan abonelerin tespit edilmesi amacıyla, destek vektör makinası kullanılarak gerçekleştirilen çalışmada, Malezya Kuala Lumpur'daki 265.870 abone verisi üzerinde çalışılmıştır (Nagi, vd., 2008a). Veri Temizliği ve abone filtreleme sonucunda aşağıdaki aboneler temizlendiğinde, abone sayısı 186.968 e düşmüştür.

- Aylık veri içinde tekrar eden abonelerin silinmesi
- 25 aylık periyotta 0 kWh tüketime sahip abonelerin silinmesi
- 25 aylık verisi bulunmayan ve periyot içinde yeni abonelik gerçekleştiren abonelerin silinerek, sadece aralıksız 25 aylık tüketime sahip abonelerin kalması

Çalışmada 25 aylık periyotta abonelerin tamamı için aylık ortalama günlük tüketim ortalamaları hesaplanmıştır. Hesaplanan parametreler minimum maksimum normalizasyonu kullanılarak normalize edildikten sonra, Gauss Kernel ile SVM modeli oluşturulmuştur. Her bir abone için, 24 aylık ortalama günlük tüketim verileri dışında, AKN parametresi ile birlikte toplamda 25 parametre ile çalışma gerçekleştirilmiştir. AKN her bir müşteri için faturaları düzenli ödeyip ödemediğine göre sistem tarafından hesaplanmaktadır ve 0 ile 5 arasında bir değer almaktadır. 5 değeri en iyi skoru gösterirken, 0 değeri ise en kötü skoru ifade etmektedir. Hiper-parametre optimizasyonu için Grid-Search algoritması kullanılmıştır. Model testinde K=10 Katlamalı Çapraz Doğrulama tekniği kullanılmış olup, model çalıştırıldığında %86,43 doğruluk (accuracy) ve %77,41 duyarlılık (recall) değerleri elde edilmiştir (Nagi, vd., 2008a).

Brezilya’da 2004 yılında yapılan çalışmada, benzer şekilde master veriler üzerinde karar ağaçları algoritması ile sınıflandırma işlemi gerçekleştirilmiştir. Bu çalışmada 52 master veri üzerinden, korelasyon fonksiyonu sonuçları ve saha tecrübeleri ile 5 tane parametre seçilmiştir. Seçilen öznitelikler ve tipleri Tablo 2.2’de gösterilmiştir (Gontijo, vd., 2004).

Tablo 2.2: Master Parametre Seçim Örnekleri

Öznitelik	Değişken Tipi
Müşteri Sınıfı	Ayrık
Müşteri Aktivitesi	Ayrık
Elektrik Bağlantısı	Ayrık
Aylık Tüketimler	Sürekli
Ortalama Tüketim	Sürekli

Kaynak: (Gontijo, vd., 2004)

Bu çalışmada 600.000 abone verisinden, tablodaki 5 özellik seçilerek, 12 aylık veri üzerinde çalışılmıştır. Son 12 ayda kaçak kullanan abonelerin seçilmesi ile sayı 100.000 müşteriye indirilmiştir. Sonrasında normal, şüpheli veya sayacı bozuk abonelerin seçilmesi ile sayı 50.000 e, aylık tüketim verisi negatif olan abonelerin de elenmesiyle 40.000 abone verisi ile devam edilmiştir. Sonuç olarak 40.000 müşteri, 12 aylık tüketim verisi, 3 ayrık ve 2 sürekli olmak üzere 5 öznitelik ile çalışılmıştır. Çalışma sonucunda aboneler kaçak kullanımlarına göre sınıflandırılmıştır. Literatürdeki diğer çalışmalar ile birlikte, ortak olarak belirlenen abone master verileri, Tablo 2.3 teki gibi oluşmuştur (Gontijo, vd., 2004).

Hindistan’da 1.350 abone üzerinde gerçekleştirilen destek vektör makinası kullanılarak oluşturulan modelde, her abone için 15’er dakikalık periyotta okunan tüketim verileri kullanılmıştır. Aboneler, günlük tüketim ortalamaları baz alınarak her biri 10 abone içeren 135 farklı gruba ayrılmıştır (Depuru, Wang ve Devabhaktuni, 2011).

Sistemde kullanılan parametreler aşağıdaki gibidir:

- Coğrafi lokasyon (doğu, batı, kuzey, güney, merkez)
- Yılın Mevsimleri (Yaz, Kış, Diğer)
- Abone sınıfı (tarımsal [küçük, büyük], ticari [küçük, orta, büyük], mesken [çok küçük, küçük, orta, büyük])

Tablo 2.3: Master Parametre Seçimi

Parametre / Öznitelik	
Müşteri Sınıfı / Abone Grubu	Tarife
Müşteri Aktivitesi / Sektör Bilgisi	Kurulu Güç
Voltaj Tipi (Alçak/Yüksek/Orta)	Sözleşme Gücü
Lokasyon (Şehir/İlçe)	Aylık Tüketimler
Voltaj (110/220)	Ortalama Tüketim
Faz Sayısı	

Kaynak: (Gontijo, vd., 2004)

Depuru, Wang ve Devabhaktuni'ye (2011) göre, abonelerin gruplarına göre sınıflandırma kriterleri Tablo 2.4, Tablo 2.5 ve Tablo 2.6'da verilmiştir. Aboneler sınıflarına göre tüketim aralıklarına bölünmüş, bu aralıklara göre de küçük abone/büyük abone ayrımına gidilmiştir. Aylık kwh tüketim değerlerinin, güne ve saate indirgenmesi ile sınıflandırma tabloları oluşturulmuştur. Buna göre tarımsal abonelerde aylık ortalama 10 kwh altı küçük abone grubuna girerken, ticari abonelerde aylık ortalama 500 kwh altı küçük abone olarak kabul edilmiştir. Kırsal küçük abonelerde ise aylık ortalama kwh üst değeri 50 kwh alınmıştır. Tarımsal abonelerde tüketimlerine göre aboneler küçük ve büyük olarak iki sınıfa ayrılırken; ticari abonelerde küçük, orta ve büyük olarak üç sınıfa; kırsal abonelerde ise çok küçük, küçük, orta ve büyük olmak üzere dört gruba ayrılmıştır.

Tablo 2.4: Tarımsal Abonelerin Sınıflandırılması

Tüketim	Küçük	Büyük
Aylık kwh	0-10	10-30
Günlük kwh	0-0.333	0.333-1
Saatlik kwh	0-0.014	0.014-0.042

Kaynak: (Depuru, Wang ve Devabhaktuni, 2011)

Tablo 2.5: Ticari Abonelerin Sınıflandırılması

Tüketim	Küçük	Orta	Büyük
Aylık kwh	0-500	500-2000	2.000-20000
Günlük kwh	0-16.677	16.677-66.677	66.677-666.677
Saatlik kwh	0-0.695	0.695-2.787	2.787-27.877

Kaynak: (Depuru, Wang ve Devabhaktuni, 2011)

Tablo 2.6: Kırsal Abonelerin Sınıflandırılması

Tüketim	Çok Küçük	Küçük	Orta	Büyük
Aylık kwh	0-50	50-300	300-600	600-1000
Günlük kwh	0-1.667	1.667-10	10-20	20-33.333
Saatlik kwh	0-0.069	0.069-0.416	0.416-0.833	0.833-1.388

Kaynak: (Depuru, Wang ve Devabhaktuni, 2011)

SVM kullanılarak yapılan çalışmanın doğruluk değeri 0,984 olarak elde edilmiştir. Aynı konuda sonraki çalışmada yine 15 dakikalık periyotta günde 4 x 24 =96 adet veri alınarak, SVM modelinde girdi parametre sayısı 96 olarak belirlenmiştir. Doğruluk oranı ise 0,92 olarak elde edilmiştir (Depuru, Wang ve Devabhaktuni, 2012).

Ramos, Souza ve Gastaldelloy'a, (2012) raporunda, Brezilya'da 5000 endüstriyel abone profili verileri üzerinde WEKA kullanılarak geliştirilen çalışmada, SVM ile birlikte YSA ve KNN algoritmaları kullanılmıştır. Bu çalışmada kullanılan öznelikler Tablo 2.7'de gösterilmiştir.

Tablo 2.7: WEKA çalışması kullanılan öznitelikler

Öznitelik Adı	Açıklama
DF - Faturalanan Demand	Aktif güç demand değeri (kW)
DC - Sözleşme Demand	Abone sözleşme demand değeri (kW)
D_{max} – Maksimum Demand	15 dakikalık periyottaki veriler içinde faturalama periyodundaki maksimum demand değeri (kW)
RE – Reaktif Enerji	Reaktif Enerji değeri (kVArh – Kilovolt amper reaktif saat)
PT - Trafo Gücü	Abonenin kurulu trafo gücü değeri (kVA)
PF - Güç Faktörü	Güç dağıtım sistemindeki verimliliği gösteren oransal değer
P_{inst} - Kurulu Güç	Aboneye ait kurulu olan tüm elektriksel ekipmanların nominal güç değerleri toplamı (kW)
LF - Yükleme Faktörü	D_{ort} – Ortalama demand ile D_{max} – maksimum demand arasındaki oran

Kaynak: (Ramos, Souza ve Gastaldello, 2012)

8 öznitelik kullanılarak, WEKA platformunda gerçekleştirilen uygulamada, 5000 profile ait veri %50 eğitim, %50 test verisi olarak ayrılmıştır. Sonuç kümesinde ise 5000 abonenin 280'inin kaçak olduğu önceden bilinmektedir. Bu anlamda gözetimli öğrenme tekniği ile ilerlenmiştir. Çalışmada modelin oluşturulmasında YSA, SVM ve KNN algoritmaları, modelin optimizasyonu aşamasında ise GA- Genetik Algoritmalar ve EA – Etraflı (Tam) Arama algoritmaları kullanılmıştır. Çalışma sonucuna göre tüm öznitelikler kullanıldığında YSA, SVM ve KNN algoritmalarının doğruluk değerleri sırasıyla 0.9448, 0.9688 ve 0.9552 olarak bulunmuştur. Tablo 2.8, doğruluk değerleri ile birlikte hata oranlarını da göstermektedir. Tüm öznitelikler kullanılarak yapılan çalışmada, en yüksek doğruluk oranını veren algoritmanın SVM olduğu görülmektedir. (A.g.e)

Tablo 2.8: WEKA Tüm Öznitelikler Kullanıldığında Sonuçlar

Sınıflandırma Algoritması	Doğruluk Oranı	Hata Oranı
YSA	% 94.48	%5.52
SVM	%96.88	%3.12
K-NN	%95.52	%4.48

Kaynak: (Ramos, Souza ve Gastaldello, 2012)

GA ve ES algoritmaları kullanılarak yapılan optimizasyon sonucunda, Tablo 2.9’da görüleceği üzere, “Ranker” metoduyla 8 öznitelik içinden PT – Trafo Gücü ve P_{inst} – Kurulu Güç olmak üzere iki öznitelik seçilmiştir. Seçilen bu öznitelikler ile algoritmalar tekrar çalıştırıldığında, SVM in yine en başarılı algoritma olduğu görülmektedir. PT – Trafo Gücü, P_{inst} – Kurulu Güç öznitelikleri ile bu özniteliklere ek olarak “Ranker” sıralamasına göre PT, P_{inst} , DC, DF, D_{max} , RE, PF, LF özniteliklerinden 3. Sıradaki öznitelik DC – Sözleşme Demand özniteliği de kullanıldığında, algoritmaların verdiği sonuçlar karşılaştırmalı olarak Tablo 2.10’da verilmiştir. Bu sonuçlara göre K-NN algoritması, 3 öznitelik ile daha iyi doğruluk oranını vermiştir (A.g.e).

Tablo 2. 9: WEKA ile Öznitelik Seçimi Sonuçları

Algoritma	Seçilen Öznitelikler
GA	PT, P_{inst}
EA	PT, P_{inst}
Ranker	PT, P_{inst} , DC, DF, D_{max} , RE, PF, LF

Kaynak: (Ramos, Souza ve Gastaldello, 2012)

Tablo 2.10: WEKA Öznitelik Seçimi Sonrası Doğruluk ve Hata Oranları

	Öznitelikler : PT, P _{inst}	Öznitelikler : PT, P _{inst} , DC
Algoritma	Doğruluk [Hata]	Doğruluk [Hata]
YSA	%94.48 [%5.52]	% 94.48 [%5.52]
SVM	%96.28 [%3.72]	%97.64 [%2.36]
K-NN	%96.20 [%3.80]	%98.12 [%1.88]

Kaynak: (Ramos, Souza ve Gastaldello, 2012)

Brezilya abone verileri üzerinde yapılan çalışmada, 100.000 abone verisi ile Mantıksal (Boolean), Bulanık (Fuzzy) ve SVM modelleri çalışılmıştır (Glauner, vd, 2016). 100.000 abone verisi üzerinden seçilen farklı alt veri kümesi örnekleri, bu modellere uygulanmıştır. AUC eğrisi altındaki alan hesaplanarak yapılan çalışmada, %5’lik Teknik Olmayan Kayıp Kaçak verisi için AUC test değerleri Mantıksal 0.465, Bulanık 0.55 ve SVM 0.55 iken; % 20’lik veri için Mantıksal 0.475, Bulanık 0.545, SVM 0.55 olarak elde edilmiştir. Tablo 2.11’de Bulanık, Mantıksal ve SVM sınıflandırmada %5’lik ve %20’lik dilimlerde hesaplanan karmaşıklık matrisleri verilmiştir.

Tablo 2.11: Bulanık, Mantıksal ve SVM Karmaşıklık Matrisleri

Sınıflandırma	(%5’lik veri)				(%20’lik veri)			
	TNR	FPR	FNR	TPR	TNR	FPR	FNR	TPR
Mantıksal	0.53	0.47	0.60	0.40	0.53	0.47	0.58	0.42
Bulanık	0.87	0.13	0.77	0.23	0.87	0.13	0.78	0.22
SVM	0.36	0.64	0.26	0.74	0.35	0.65	0.25	0.75

Kaynak: (Glauner, vd, 2017)

Bu çalışmada, verilerin en azından alttaki 4 özelliği içermesi gerektiği vurgulanmıştır.

- Farklı Abone Tipleri
- Abone ve Kaçak Abone Sayısı
- Farklı Bölgeler

- Uzun Periyotlarda Sayaç Okuma Verileri

Tablo 2.12: Literatürdeki Model, Veri Seti Ve Performans Değerleri

Kaynak	Algoritma	Abone Sayısı	Doğruluk	Kesinlik	Duyarlılık	AUC	Kaçak Oranı
(Nagi J. , Yap, Tiong, Ahmed, & Mohamad, 2010)	SVM (Gauss)	<400	0.86	-	0.77	-	-
(Nagi J., ve diğerleri, 2011)	SVM + fuzzy	100.000	-	-	0.72	-	-
(Glauner P. , ve diğerleri, 2016)	Bool rules	700.000	-	-	-	0.47	%5
(Glauner P. , ve diğerleri, 2016)	Fuzzy rules	700.000	-	-	-	0.55	%5
(Glauner P. , ve diğerleri, 2016)	SVM (linear)	700.000	-	-	-	0.55	%5
(Glauner P. , ve diğerleri, 2016)	Bool rules	700.000	-	-	-	0.48	%20
(Glauner P. , ve diğerleri, 2016)	Fuzzy rules	700.000	-	-	-	0.55	%20
(Glauner P. , ve diğerleri, 2016)	SVM (linear)	700.000	-	-	-	0.55	%20
(Nagi, Mohammad, Yap, Tiong, & Ahmed, 2008)	SVM	<400	-	-	0.53	-	-
(Nagi J. , ve diğerleri, 2008)	Genetic SVM	1.171	-	-	0.62	-	-
(Muniz , Marley, Tansheit, & Figueiredo, 2009)	Neuro-fuzzy	20.000	0.68	0.51	-	-	-
(Costa, Alberto, Portela, Maduro, & Eler, 2013)	NN	22.000	0.87	0.65	0.29	-	-
(Spiric, Stankovic, Docic, & Popovic, 2014)	Rough sets	?	0.93	-	-	-	-

**Tablo 2.12: Literatürdeki Model, Veri Seti Ve Performans Değerleri
(Devamı)**

Kaynak	Algoritma	Abone Sayısı	Doğruluk	Kesinlik	Duyarlılık	AUC	Kaçak Oranı
(Cabral, Pinto, & Pinto, 2009)	SOM	2.000	0.93	0.85	0.98	-	-
(Depuru, Wang, & Devabhaktuni, 2011)	SVM (Gauss)	1.350	0.98	-	-	-	-
(Sahoo, Nikovski, Muso, & Tsuru, 2015)	Regression	30	-	-	0.22	-	%1
(Sahoo, Nikovski, Muso, & Tsuru, 2015)	Regression	30	-	-	0.78	-	%2
(Sahoo, Nikovski, Muso, & Tsuru, 2015)	Regression	30	-	-	0.98	-	%3
(Sahoo, Nikovski, Muso, & Tsuru, 2015)	Regression	30	-	-	1	-	%4-10
(Ramos, Souza, & Gastaldello, 2012)	SVM	5.000	0.96	-	-	-	-
(Ramos, Souza, & Gastaldello, 2012)	KNN	5.000	0.96	-	-	-	-
(Ramos, Souza, & Gastaldello, 2012)	NN	5.000	0.94	-	-	-	-
(Ramos C. , Souza, Papa, & Falcao, 2009)	OPF	736	0.90	-	-	-	-
(Ramos C. , Souza, Papa, & Falcao, 2009)	SVM (Gauss)	736	0.89	-	-	-	-
(Ramos C. , Souza, Papa, & Falcao, 2009)	SVM (linear)	736	0.45	-	-	-	-
(Ramos C. , Souza, Papa, & Falcao, 2009)	NN	736	0.53	-	-	-	-
(Nizar, Zhao, & Dong, 2006)	Decision Tree	?	0.99	-	-	-	-

Tablo 2.12, literatürde teknik olmayan kayıp kaçakların tespitinde kullanılan makine öğrenmesi modelleri, karşılaştırmalı olarak veri seti büyüklükleri ve performans değerlerini göstermektedir (Glauner, vd, 2017).

Makine öğrenmesi algoritmalarında, üzerinde çalışılan verinin boyutu kadar, içeriği ve çeşitliliği de oldukça önemlidir. Farklı karakterlerde özellikler taşıyan verilerin çalışmada yer alması, daha doğru sonuçların elde edilmesini sağlayacaktır. Ters durumda, yani sadece belirli tiplerdeki veriler üzerinde çalışıldığında ise, modelin yetersiz eğitilmesi ve doğru öğrenememesi söz konusudur.

Aynı şekilde yapılan çalışmalarda, tüm veri içerisindeki kaçak veri dağılımı da, makine öğrenmesi algoritmalarının performansı açısından oldukça önemlidir. Genellikle bu alanda yapılan çalışmalarda, kaçak verilerinin normal verilere oranı çok düşük kalmaktadır. Veri seti dengeli hale getirilmeden geliştirilen uygulamalarda, aşırı öğrenme söz konusu olabilmektedir. Ayrıca amacımız kaçak olan aboneleri doğru tespit etmek olduğundan, doğruluk değeri kadar, Duyarlılık/Kesinlik değerleri de göz önüne alınmalıdır.

Ramos C. , Souza, Papa, & Falcao (2009), Ramos, Souza, & Gastaldello, (2012) ve Nizar, Zhao, & Dong (2006) tarafından yapılan çalışma sonuçlarında, sadece doğruluk değerlerine yer verilmiştir. Kaçak tahmini için önemli olan kesinlik/duyarlılık değerleri bu çalışmalarda yer almamaktadır. Doğruluk değerlerinde yüksek başarıma ulaşılrken, kesinlik/duyarlılık değerleri düşük çıkabilmektedir. Bu değerlerin birlikte değerlendirilmesi sağlıklı olacaktır.

Sahoo, Nikovski, Muso, & Tsuru (2015) çalışmasında ise, dengesiz veri setindeki kaçak oranları %1-%10 arasında arttırılarak, 0,22 - 0,78 - 0,92 - 1 doğruluk değerlerine ulaşılmıştır. Çalışmada sadece duyarlılık değeri ile değerlendirme yapılmış olup, doğruluk değerleri paylaşılmamıştır. Abone sayısı ise sadece 30 olup, bu tip bir çalışma için yeterli düzeyde değildir. Benzer şekilde sadece performans sonuçları olarak, sadece duyarlılık değerinin dikkate alındığı Nagi J., ve diğerleri çalışmasında, (2011); Nagi, Mohammad, Yap, Tiong, & Ahmed çalışmasında, (2008), Nagi J. , ve diğerleri, (2008) çalışmasında sırasıyla 0,72 – 0,53 – 0,62 – 0,29 duyarlılık değerleri elde edilmiştir. Burada tek başına değerlendirildiğinde yeterli olmasa bile, sadece Nagi J., ve diğerleri, (2011) çalışmasında elde edilen 0,72 değeri bir başarımlar olarak görülmektedir.

Tüm veri setindeki kaçak oranının yetersizliği, bir problem olarak karşımıza çıkmaktadır ve çalışmanın performansını doğrudan etkilemektedir. Nagi J. , Yap, Tiong, Ahmed, & Mohamad, (2010); Nagi J., ve diğerleri, (2011); Nagi, Mohammad, Yap, Tiong, & Ahmed, (2008); Nagi J. , ve diğerleri, (2008); Muniz , Marley, Tansheit, & Figueiredo, (2009); Costa, Alberto, Portela, Maduro, & Eler, (2013); Spiric, Stankovic, Docic, & Popovic, (2014); Cabral, Pinto, & Pinto, (2009), Depuru, Wang, & Devabhaktuni, (2011); Ramos, Souza, & Gastaldello, (2012); (Ramos C., Souza, Papa, & Falcao, (2009); Nizar, Zhao, & Dong, (2006) çalışmalarının tamamındaki veri setindeki kaçak oranı dağılımı belirtilmemiştir. Kaçak oranı dengelenmediği veri setleri üzerinde yapılan çalışmalarda, aşırı öğrenmeye bağlı yüksek doğruluk değerleri elde edilebilmektedir.

Kaçak oranlarının paylaşıldığı Nizar, Zhao, & Dong (2006) çalışmasında, sadece AUC sonuçları değerlendirilmiştir. Bu çalışmada da doğruluk ve duyarlılık/kesinlik değerlerine yer verilmemiştir.

Tablodaki en yüksek doğruluk değerine sahip olan çalışma Nizar, Zhao, & Dong (2006) çalışmasında, hem veri setindeki abone sayısı ve kaçak oranı bilgilerini, hem de karmaşıklık matrisindeki diğer değerleri içermemektedir. Tek başına doğruluk değeri başarı ölçütü olarak değerlendirilmemelidir.

Costa, Alberto, Portela, Maduro, & Eler (2013) çalışmasında ise hem doğruluk hem de duyarlılık değerleri birlikte paylaşılmıştır. Doğruluk 0,87 olarak elde edilirken, Duyarlılık değeri ise 0,29'da kalmıştır. Bu çalışmada yüksek doğruluk değerleri elde edilirken, düşük duyarlılık değerlerine ulaşıldığı görülmektedir. 0,29 değeri kaçak abonelerin tespitinde düşük bir oran olarak kalmıştır. Bu da belirttiğimiz veri setindeki kaçak oranının yetersizliği probleminin bir sonucudur.

Doğruluk/duyarlılık metrikleri birlikte değerlendirildiğinde, Tablo 2.12'deki en başarılı çalışma, kaçak oranı bilgisi paylaşılmamış olsa da, 2.000 kayıt üzerinde yapılan ve 0,93 doğruluk ile 0,98 kesinlik sonuçlarına ulaşılan Cabral, Pinto, & Pinto (2009) çalışması olarak görülmektedir.

Bu projede yapılan çalışmada da, ilk etapta %98 civarında doğruluk değerine ulaşılmıştır. Fakat bununla birlikte düşük duyarlılık değerleri ve dengesiz veri seti sonuçları da dikkate alınarak, performans açısından çalışma bir bütün olarak değerlendirilmiştir. Bu kapsamda çeşitli yöntemler kullanılarak, farklı kaçak

oranlarına sahip veri setleri ile kaçak oranı açısından dengeli bir veri seti oluşturulmaya çalışılmıştır. Ayrıca sadece Doğruluk değeri bir performans ölçütü olarak alınmayıp, sonuçlar Duyarlılık değeri ile birlikte değerlendirilmiştir. Elde edilen sonuçlar üzerinde, optimum başarıma ulaşılacak şekilde, performans artırma çalışmaları gerçekleştirilmiştir.

2.1 Öznitelik Çıkarımları

Sayaç verilerinin okunması, otomatik sayaç okuma sistemi kullanılmayan yerlerde genellikle yıllık ve aylık olarak gerçekleştirilmektedir. Literatürde yapılan teknik olmayan kaçakların tespit edilmesi ile ilgili çalışmalar da genellikle, tüketimlerin aylık olarak okunması üzerinde ilerlemiştir. Otomatik sayaç okuma sistemlerinin kurulması ile daha sık periyotta okunan veriler üzerinde yapılacak analizler de daha efektif sonuçlar alınmasını sağlayacaktır.

Abonelerin aylık tüketimleri üzerinde yapılan teknik olmayan kaçak tespiti çalışmalarında, ortalama aylık tüketim değeri hesaplanarak, öznitelik olarak kullanılmaktadır. (Nagi J. , vd., 2010; Nagi J. , vd., 2011; Glauner P. , vd., 2016; Nagi, vd., 2008b; Fourie & Calmeyer, 2005)

Ortalama günlük tüketim değeri aşağıdaki gibi hesaplanmaktadır. (Nagi J. , vd, 2010b)

$$X_h^{(m)} = \frac{L}{D_{h+1} - D_h}, \quad h = 1, 2, \dots, 24 \quad (\text{Nagi J., vd., 2010b})$$

m : $\{0, 1, \dots, m-1\}$ aralığında olmak üzere toplam abone sayısı

h : Veri Alınan Ay Sayısı

L : İlgili aydaki toplam tüketim

$(D_{h+1} - D_h)$: İki ay arasındaki gün farkı

$X_h^{(m)}$: İlgili ayda ortalama günlük tüketim (kWh)

İlgili ayda ortalama günlük tüketim değeri, verinin olduğu her ay için hesaplanan, ilgili aydaki ortalama tüketim değerini vermektedir. $h = 1, 2, \dots, 24$ değerleri, son 24 ay verisinin işlendiğini gösterir. Yukarıdaki örnekte son 2 yıllık veri üzerinde

çalışılmıştır. Bu durumda her abone için 24 tane olmak üzere, $X_h^{(m)}$ ilgili aydaki ortalama günlük tüketim değeri hesaplanacaktır. $(D_{h+1} - D_h)$ değeri, bir sonraki ay ile geçerli ay arasında tüketimin geçerli olduğu gün farkıdır. L değeri ise, ilgili aydaki toplam tüketimi göstermektedir. Her ay yürüyen bakiye şeklinde olup, tek endeks değeri alınarak okunan sayaçlar için L değeri, (Son Endeks – İlk Endeks) olarak hesaplanır. Böylelikle her abone için 24 tane günlük ortalama kWh tüketim değerleri öznitelik olarak hesaplanır.

Abonelerin ay içinde okunma zamanları farklı günler olduğundan, ayrıca bazı ayların 30, bazılarının da 31 çektiği göz önüne alınarak, bu formül kullanılarak her ay için o aydaki günlük ortalama tüketim değeri hesaplanmıştır. 24 ortalama günlük değeri, her bir abone için hesaplanarak, abone sayısı * 24 boyutunda bir matris oluşturulur (Nagi J., vd., 2010b).

Tablo 2.13: Bulanık Mantık Yaklaşımında Seçilen Parametreler

Parametre	Açıklama
Class	SVM sonucu sınıf etiketi
L23	23.günlük ortalama kWh tüketim
L24	24.günlük ortalama kWh tüketim
MinkWh	Minimum kWh değeri
MaxkWh	Maksimum kWh değeri
DiffkWh	Maksimum Tüketim ile Minimum Tüketim Farkı
DetectCount	Abonenin kaçak kullanım sayısı
Probability	SVM sonucu tahmin edilen sınıf olasılığı
TOE	Elektrik Kaçak Kullanım İşareti
HRC	Yüksek Riskli Müşteri İşareti

Kaynak: (Nagi J., vd., 2011)

AKN – Abone Kredi Notu (CWR – Credit Worthiness Rating), her abone için sistem tarafından üretilen ve abonenin kredi notunu belirten bir değerdir. Abonenin kredi notu, abonenin fatura ödemeleri, aylık ödeme durumu ve geciken ödemeleri gibi veriler ışığında hesaplanan ve her bir abonenin değerini gösteren bir özniteliktir. AKN

değeri 0 ile 5 arasında değer alabilmektedir. Bu değer de 24 ortalama günlük kWh tüketim değerleri ile birlikte 25. öznitelik olarak kullanılmıştır (Nagi J., vd., 2008b).

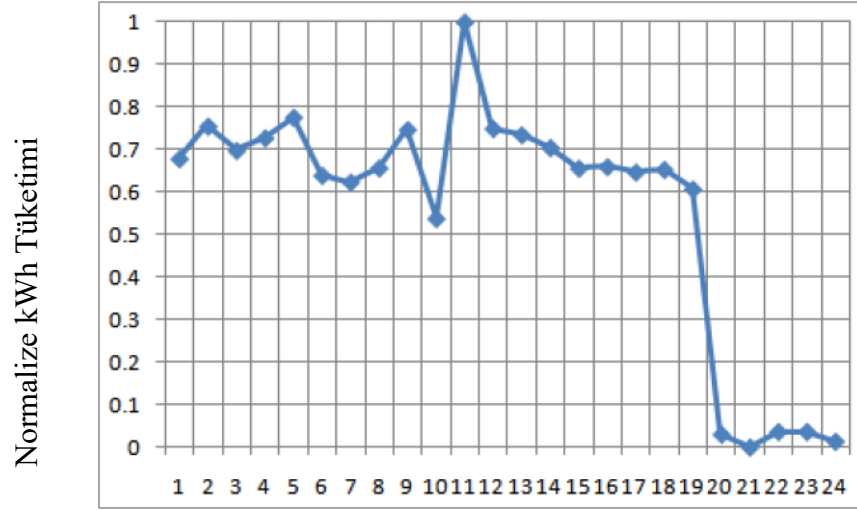
Bulanık Mantık Yaklaşımı kullanarak, teknik olmayan kayıpların Karar Destek Makinası tabanlı Geliştirilmesi ile ilgili çalışmada, 10 tane parametre seçilerek, son 6 aylık sayaç ölçüm verileri üzerinde çalışma yapılmıştır. Bu çalışmada, tüm veriler içinden potansiyel kaçak kullanımı yapan abonelerin seçiminde, SQL (Structured Query Language) dili kullanılmıştır (Nagi J. , vd., 2011). SQL ile seçilen parametreler Tablo 2.13'de gösterilmiştir.

Nagi J. , vd. (2011) çalışmada, daha önce belirlenen, 24 tüketim değeri ve 1 Abone Kredi Notu olmak üzere toplam 25 öznitelikten farklı olarak; Ortalama Tüketim, Maksimum Tüketim, Minimum Tüketim, Standart Sapma, Toplam Kaçak Sayısı öznitelikleri kullanılmıştır.

Yine bu alanda yapılan başka bir çalışmada, sınıflandırma algoritmalarından SVM algoritması kullanılarak geliştirilen modelde, kaçak tahmin sonuçlarına göre aboneler 4 tip sınıfa ayrılmıştır (Nagi J., vd., 2008b).

- Onaylı Kaçak Şüpheliler
- Onaysız Kaçak Şüpheliler
- Onaylı Temiz Şüpheliler
- Onaysız Temiz Şüpheliler

Aboneler, onaylı veya onaysız kaçak şüpheliler olarak sınıflandırılmıştır. Onaylı tanımına giren abonelerin profil grafiklerinde ani değişimler tespit edilmiştir. Onaysız grubundaki abonelerde ise keskin bir ani grafik değişimine rastlanmamıştır. Önceden kaçak kullanımı olan aboneler kaçak şüpheli, kaçak kullanımı olmayan aboneler ise temiz şüpheli olarak sınıflandırılmıştır. Şekil 2.1'de daha önceden de kaçak kullanmış ve tekrar kaçak şüpheli sınıfına giren, Onaylı Kaçak Şüpheli sınıfa giren bir abonenin profil grafiği verilmiştir. Şekil 2.2'de ise daha önceden kaçak kullanmamış, fakat tüketimlerinde ani değişimler olan ve Onaylı Temiz Şüpheliler sınıfına giren bir abonenin profil grafiği verilmiştir.



24 Aya ait ortalama günlük tüketim değerleri

Şekil 2.1: Şüpheli bir aboneye ait profil yüklemesi

Kaynak: (Nagi J., vd., 2008b)



Şekil 2.2: Normal bir aboneye ait profil yüklemesi

Kaynak: (Nagi J., vd., 2008b)

2.2 Özniteliklerin Ölçeklenmesi – Veri Normalizasyonu

Normalizasyon yöntemleri ile verilerin normalize edilerek, yapay zeka uygulamalarının performanslarının artırılması, veri ön işleme süreçlerinde önemli bir adım haline gelmiştir. Normalizasyon işlemleri sonucunda verilerin boyutları aynı kalırken, verinin maksimum ve minimum noktaları değişerek, veri başka bir uzaya taşınmış olur (Akdemir, 2009).

Literatürde birçok veri normalizasyon yöntemi bulunmaktadır. Bu yöntemler arasında en çok kullanılan yöntemlerden bazıları aşağıdaki gibidir:

- Z-Skoru Normalizasyonu
- Minimum Maksimum Normalizasyonu
- Medyan Normalizasyonu
- Sigmoid Normalizasyonu
- İstatiksel Kolon Normalizasyonu
- Ondalık Normalizasyon

Mühendislik uygulamalarında daha çok Minimum Maksimum Normalizasyonu ve Ondalık Normalizasyonu kullanılırken, ile tıp alanında ise Z-Skoru Normalizasyonu ağırlıklı olarak kullanılmaktadır.

Bu çalışmada normalizasyon için, minimum maksimum normalizasyonu kullanılmıştır.

2.4.1 Z-Skoru Normalizasyonu

Her bir öznitelik için standart sapma ve aritmetik ortalama hesaplanır. Şekil 2.3'te gösterilen z dönüşümü formülü kullanılarak, değişkenin ortalaması ve standart sapmasına bağlı olarak veriler normalleştirilir.

$$z = \frac{x_i - \mu}{\sigma} \quad (2.1)$$

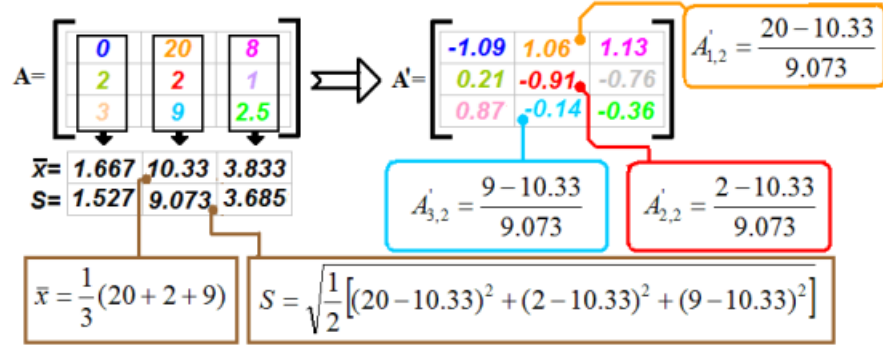
z : Normalize Edilmiş Veri

x_i : Normalize Edilmemiş Girdi Değeri

μ : Aritmetik ortalama

σ : Standart Sapma

Yukarıdaki formül, örnek bir A matrisinin Z-Skoru Normalizasyonunun nasıl hesaplandığını göstermektedir (Akdemir, 2009).



Şekil 2.3: Z Skor Normalizasyon Örneği

Kaynak: (Akdemir, 2009)

2.4.2 Minimum Maksimum Normalizasyonu

Veriler doğrusal olarak normalize edilir. Herhangi iki veri arasındaki fark veya değerlerin büyüklüklerinin orantısallığı değişmez. (Akdemir, 2009) Verinin alabileceği en düşük değer minimum değeri, en yüksek değer ise maksimum değeri ifade eder. Veriler genellikle 0 ile 1 arasında olmak üzere, sabit bir aralıkta ölçeklendirilir. Böylelikle aykırı uç değerlerin etkisi azaltılabilir. Minimum-maksimum normalizasyon formülü aşağıdaki gibi ifade edilir :

$$Z = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2.2)$$

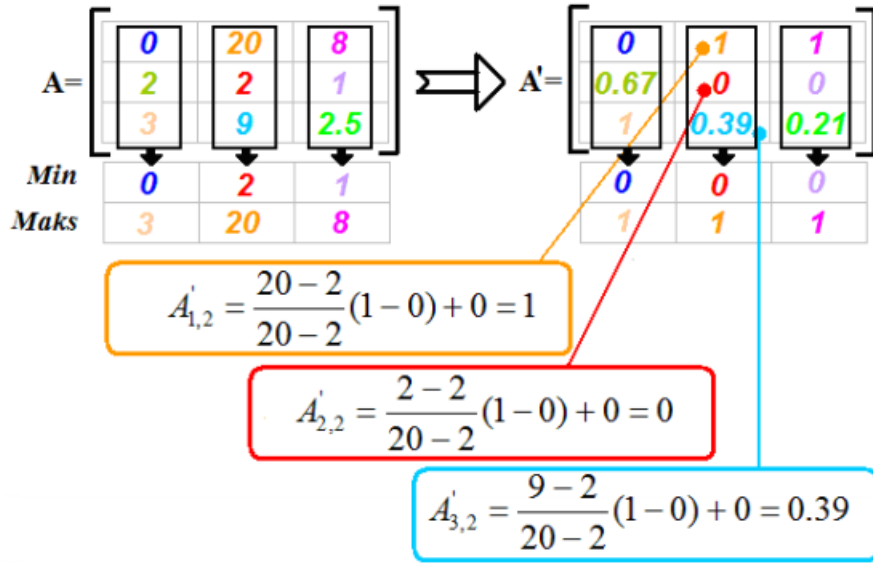
Z: Normalize Edilmiş Veri

X: Normalize Edilmemiş Girdi Değeri

min(x): Veri setindeki minimum X değeri

max(x): Veri setindeki maksimum X değeri

Bu yöntemde veri setindeki değerlerde negatif olan değerleri ortadan kaldırır ve bu değerler pozitif hale gelir. Şekil 2.4'te bir A matrisine minimum maksimum normalizasyonu metodunun uygulanışı örnek olarak verilmiştir (Akdemir, 2009).



Şekil 2.4: Minimum Maksimum Normalizasyon Örneği

Kaynak: (Akdemir, 2009)

2.4.3 Ondalık Normalizasyonu

Ondalık normalizasyonda, değişkene ait değerlerdeki ondalık kısımlar hareket ettirilerek normalizasyon işlemi gerçekleştirilir. Değişkenin maksimum mutlak değerine bağlı olarak, hareket ettirilecek ondalık nokta sayısı belirlenir. Örneğin maksimum değer 900 olduğu bir veri setinde, $n=3$ olduğundan 900 sayısı, virgöl 3 hane kaydırılarak 0,9 olarak normalleşir. (Ergün, 2019)

Veri setindeki değerleri 1'den küçük hale getirmek için, veriler 10 veya 10'un katlarına bölünür. Aşağıda Ondalık Normalizasyon formülü verilmiştir. J değeri, V değerini 1'den küçük yapan en küçük değer olarak tanımlanabilir. (Akdemir, 2009)

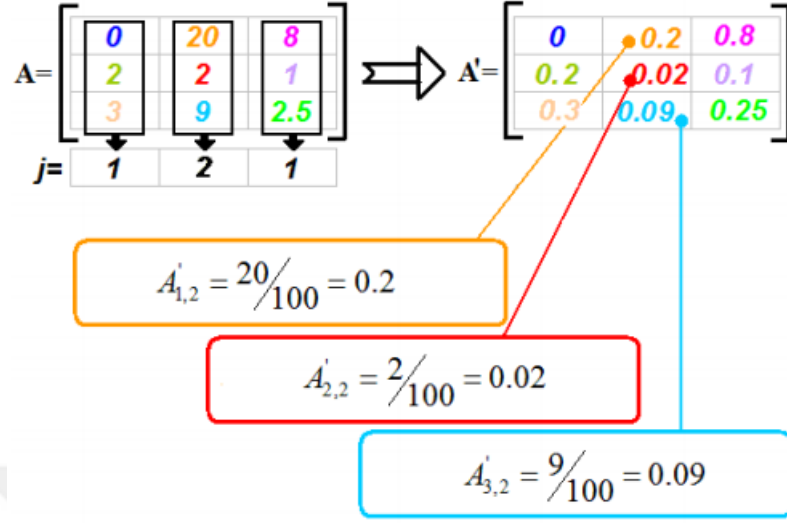
$$v' = \frac{v}{10^j} \quad j: \text{Max}(|v'|) < 1 \text{ olacak şekildeki en küçük tam sayı} \quad (2.3)$$

v' : Normalize Edilmiş Veri

v : Normalize Edilmemiş Girdi Değeri

Şekil 2.5'de Ondalık Normalizasyon uygulanmış bir A matrisi örneği yer almaktadır. Her bir sütunda, maksimum değeri 1'den küçük yapan j değeri bulunarak, tüm sütundaki verilere ondalık normalizasyon formülüyle uygulanır. Örnekte birinci ve

üçüncü sütunlar için tüm veriler 10, ikinci sütun için ise 100 değerine bölünmüştür.
($10^2=100$)



Şekil 2.5: Ondalık Normalizasyon Örneği

Kaynak: (Akdemir, 2009)

2.4.4 İstatiksel Kolon Normalizasyonu

İstatiksel kolon normalizasyon yönteminde, her bir kolon normalizasyon değeri ile normalize edilir. Her veri örneği, normalize edilmiş kolon değerine bölünür ve küçük bir bias (yanlılık) değeri ile çarpılır. Aşağıda İstatiksel Kolon Normalizasyonu formülü verilmiştir.

$$x' = \frac{x_i - n(c_a)}{n(c_a)} \times 0.1 \quad (2.4)$$

Kaynak: (Jayalakshmi & Santhakumaran, 2011)

x' : Normalize Edilmiş Veri

x_i : Normalize Edilmemiş Girdi Değeri

$n(c_a)$: Normalize kolon değeri

2.4.5 Sigmoid Normalizasyonu

Sigmoid normalizasyonu ile veriler 0 ile 1 ya da -1 ile 1 arasında sınıflandırılır. Sigmoid normalizasyonu için belirli fonksiyonlar bulunmaktadır. Özellikle Tanjant Sigmoid fonksiyonu, süreçleri hızlandırma anlamında iyi bir seçimdir (Jayalakshmi & Santhakumaran, 2011).

$$x' = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.5)$$

Kaynak: (Jayalakshmi & Santhakumaran, 2011)

x' : Normalize Edilmiş Veri

x : Normalize Edilmemiş Girdi Değeri

e : Doğal logaritma değeri

2.4.6 Medyan Normalizasyonu

Her bir girdi için medyan değeri alınarak, her girdi için medyan normalizasyonu gerçekleştirilir. Medyan aşırı sapmalardan etkilenmez (**Jayalakshmi & Santhakumaran, 2011**). Medyan normalizasyon formülü aşağıda gösterilmiştir.

$$x' = \frac{x_i}{\mathit{median}(a_i)}$$

Kaynak : (Jayalakshmi & Santhakumaran, 2011)

x' : Normalize Edilmiş Veri

x_i : Normalize Edilmemiş Girdi Değeri

$\mathit{median}(a_i)$: Girdi setinin medyan değeri

ÜÇÜNCÜ BÖLÜM

KTS: VERİLERİN HAZIRLANMASI

Bir önceki bölümde belirtildiği gibi, makine öğrenmesi projelerinde verinin boyutu kadar çeşitliliği de büyük önem taşımaktadır. Bu kapsamda veriler hazırlanırken, veri setinde abonelere ait maksimum detayda ve çeşitlilikte verilerin bulunmasına özen gösterilmiştir.

Dicle Elektrik Dağıtım A.Ş. bünyesinde kullanılan uygulamalara ait veri tabanı sistemlerinde çalışmalar gerçekleştirilmiştir. Dicle Elektrik'te abone yönetiminde ana sistem olarak MBS – Müşteri Bilgi Sistemi kullanılmasına rağmen; otomatik sayaç okuma, coğrafi bilgiler, kaçak takibi ve saha operasyonları, sayaç değişimleri ve sayaç arıza yönetimi operasyonları için ayrı ayrı birden fazla uygulama kullanılmaktadır. Proje kapsamında kullanılan tüm bu uygulamalar, bu uygulamalara ait iş süreçleri ve veri tabanlarına her bir uygulamaya ait tutulan kayıtların kayıt desenleri üzerinde detaylı çalışmalar gerçekleştirilmiştir. Kayıt yapılarının belirlenmesinin ardından, proje kapsamındaki sistemlere ait veri tabanlarındaki ham veriler veri toplama yöntemleri ile işlenerek, oluşturulan yeni veri ambarına aktarılmıştır. Veri ambarına aktarılan veriler, veri temizliği işlemleri sonrası, proje kapsamında kullanılabilir hale getirilmiştir.

Uygulama süreçlerinin öğrenilmesi, bu uygulamalara ait verilere ulaşılarak verilerin hazırlanması, veri ambarına aktarılan verilerin temizlenerek işlenmesi, projenin zaman kapsamında en fazla zaman alan bölümü olmuştur. Bu bölümde bu konuda yapılan çalışmaların detaylarına yer verilmiştir.

3.1 Uygulama Sistemleri

Dicle EDAŞ bünyesinde kullanılan sistemlerden veriler alınmıştır:

- Müşteri Bilgi Sistemi
- OSOS - Otomatik Sayaç Okuma Sistemleri
- Saha İş Gücü Yönetim Sistemi
- Coğrafi Bilgi Sistemi

- Saha Kaçak Takip Sistemi

Dicle Elektrik Dağıtım'da kullanılan bu sistemlerden ana sistem olan, Müşteri Bilgi Sistemi'nde aşağıdaki bilgiler yer almaktadır:

- Müşteri (Abone) Bilgileri
- Sayaç Bilgileri
- Sayaç Değişim Bilgileri
- Kaçak İhbarnameleri
- Kaçak Tutanakları
- Kaçak Faturaları
- Sayaç Okuma (Endeks) Verileri

Bu kapsamda, elektrik abonelerinin nüfus bilgileri, adres bilgileri, sözleşme bilgileri, bağlantı anlaşması bilgileri Müşteri Bilgi Sistemi'nde tutulmaktadır. Ayrıca abonelerde takılı durumda olan tüm sayaç bilgileri ile birlikte sayaç değişimlerine ait geçmiş bilgiler de Müşteri Bilgi Sistemi'nde bulunmaktadır. Tüm abonelerin sayaçları sahada genellikle ayda bir kez okunmakta, okunan sayaçlara ait endeks değerleri ile sahada faturalama işlemleri gerçekleştirilmektedir.

Enerjisi kesilen aboneler elektrik kullandığında, Elektrik Dağıtım kurumu (EDAS) üzerinden kaçak faturaları düzenlenmektedir. Kaçak faturaları, enerjisi kesilen abonenin, kesilen elektriği açılmadan elektrik kullanması sonucu veya sahada kaçak tarama faaliyetleri sonrasında tespit edilmesi sonucu düzenlenebilmektedir. Sahada kaçak tespit edildiğinde, kaçak tutanakları tutulmakta ve sistemde kaçak ihbarnameleri düzenlenmektedir. Düzenlenen kaçak ihbarnameleri sonucunda, kaçak tüketim gerçekleştirdiği belirlenen aboneler, SMS veya kağıt bildiri ile uyarılmaktadır. Kaçak kullanım sonucu çıkan faturalarını ödemeyen aboneler, belirli bir süre sonunda hukuğa sevk edilmektedir.

Enerjisi kesik iken sayaca müdahale edip, mührü açarak elektrik kullanmaya devam eden abonelere, bir sonraki faturalarında kaçak faturası olarak 1,5 kat cezalı olarak fatura kesilmektedir. Aynı şekilde devam ederse, sonraki faturalar 2 kat cezalı olarak çıkmaktadır. Sayacında problem olan veya kaçak kullanım şüphesi bulunan abonelerin sayaçları sökülerek, Sayaç Ayar Merkezi'nde incelenmektedir. İnceleme sonucunda

sayaca müdahale tespit edilen abonelere, kaçak faturaları düzenlenmektedir. Sayaç Ayar Masası'nda yapılan işlemler, Sayaç Ayar Uygulaması'nda takip edilmektedir. Sayaç Ayar Uygulaması, Müşteri Bilgi Sistemleri ile entegre bir şekilde çalışmaktadır.

Müşteri Bilgi Sistemlerinde, abonelere ait faturaya baz teşkil eden sayaç okuma (endeks) değerleri yer almaktadır. OSOS (Otomatik Sayaç Okuma Sistemleri) ile abonelerin sayaçları anlık olarak takip edilebilmektedir. Dicle Elektrik Dağıtım bünyesindeki abonelerin yaklaşık 300 binine, uzaktan okunabilen sayaçlar takılmıştır (2018 verilerine göre). OSOS ile abonelere sahada okumaya gerek kalmadan, merkezden uzaktan okuma ile fatura kesilebilmektedir. OSOS ile sayaçlar modeme bağlanmakta, modem aracılığıyla GPRS teknolojisi ile veriler merkeze iletilmektedir.

OSOS ile Müşteri Bilgi Sistemleri arasındaki entegrasyon ile, OSOS ile daha sık okunan sayaç verileri, ay sonunda son sayaç değerleri (endeks) ile Müşteri Bilgi Sistemlerine aktarılarak, faturalama işlemleri Müşteri Bilgi Sistemleri'nde gerçekleştirilmektedir.

OSOS ile uzaktan okunan sayaçlara ait daha çok veri, daha fazla sıklıkla periyodik olarak alınabilmektedir. Örneğin "kapak açma uyarısı", "demand" değerleri değişimleri gibi bilgiler anlık olarak takip edilmekte, sayaçlara yapılan şüpheli müdahaleler alarm olarak izlenebilmektedir.

Müşteri Bilgi Sistemleri üzerinden enerji kesme/açma, sayaç değişiklik emirleri otomatik/manuel olarak oluşturulmaktadır. İlgili emirler Saha Yönetim Sistemi'ne aktarılarak, saha operasyonları "Saha Yönetim Sistemi" üzerinden yönetilmektedir. Müşteri Bilgi Sistemi ve Saha Yönetim Sistemi entegrasyonları ile saha faaliyetleri uçtan uca yönetilebilmektedir.

Coğrafi Bilgi Sistemlerinde tüm abonelere ait lokasyon ve koordinat bilgileri gibi coğrafi bilgileri yer almaktadır. Abonelerin enlem ve boylam bilgileri ile, harita üzerinde lokasyonları takip edilmektedir. Coğrafi Bilgi Sistemleri'nde yer alan katmanlar ile abonelere ait binalar, şebeke tanımları ile abone-şebeke bağlantı bilgileri tutulmaktadır. Elektrik dağıtım ağı şebeke tanım bilgileri, Coğrafi Bilgi Sistemleri'nde yer almaktadır.

Saha Kaçak Takip Sistemi üzerinde, sahada gerçekleştirilen kaçak taramaları döküman ve dosyaları kaydedilmektedir. Bu sistem ile Müşteri Bilgi Sistemi arasındaki

entegrasyon ile abone verilerine ulaşılabilir. Ayrıca sahada gerçekleştirilen kaçak tarama sonuçları Saha Kaçak Takip Sistemi'ne işlenmektedir.

Enerji açma, enerji kesme, kesme ihbarı bırakma, kaçak tespit gibi tüm saha çalışmaları Saha İş Gücü Yönetim Sistemi'nde takip edilmektedir. Tez çalışması sonucu Makine Öğrenmesi ile ortaya konacak tahmin sonuçlarının doğruluğu, SYS-Saha (İş Gücü) Yönetim Sistemi'ndeki gerçek sonuçlar ile karşılaştırılarak değerlendirilmektedir.

3.2 Sistem Veri Tabanları ve Veri Yapıları

Dicle EDAŞ'ta kullanılan uygulamaların her biri, uygulamalara özel olmakla birlikte MSSQL ve Oracle veri tabanlarından birini kullanmaktadır. Her bir uygulamanın kullandığı veri tabanında, kendine özgü bir kayıt yapısı bulunmaktadır. Dicle EDAŞ'ta kullanılan aboneler ile ilişkili sistemler ve veri tabanı bilgileri Tablo 3.1'de verilmiştir. Bu sistemler arasında abone ana verilerini tutan MBS ile tüm uzaktan okunan sayaçların 15 dakikalık, saatlik, günlük ve aylık verilerini tutan OSOS 1 ve OSOS 2 sistemleri ana sistemler olarak kabul edilebilir.

Tablo 3.1: KTS Veri Tabanları

Uygulama Sistemi	Açıklama	Veritabanı
MBS	Müşteri Bilgi Sistemi	Oracle
OSOS 1	Otomatik Sayaç Okuma Sistemi 1	MS SQL
OSOS 2	Otomatik Sayaç Okuma Sistemi 2	Oracle
CBS	Coğrafi Bilgi Sistemi	Oracle
SAP	SYS - Saha Yönetim Sistemi	Oracle
SKTS	Saha Kaçak Takip Sistemi	MS SQL
SAU	Sayaç Ayar Uygulaması	MS SQL

3.2.1 MBS – Müşteri Bilgi Sistemi Verileri

Teknik olmayan kayıpların analizinde Müşteri Bilgi Sistemleri kullanılmaktadır. Müşteri Bilgi Sistemi'nden alınan veriler, işleme konulmadan önce veri ön işleme işlemleri gerçekleştirilmelidir (Nizar, vd., 2006).

Müşteri Bilgi Sistemi, Dicle Elektrik'te elektrik tüketim, faturalama, tahsilat, kesme-çama gibi işlemlerin gerçekleştirildiği ana sistemdir. Bu sistemde müşterilerin ana verileri, abonelik ve sözleşme bilgileri, aylık tüketim verileri, faturaları ve tahsilatları yer almaktadır. Projede abonelere ait ana veriler, Oracle veri tabanını kullanan MBS sisteminden alınacaktır.

3.2.2 OSOS – Otomatik Sayaç Okuma Sistemi Verileri

Dicle Elektrik Dağıtım'da, Türkiye'deki en büyük OSOS yapısı bulunmaktadır. Bölgedeki 6 ilde uzaktan okunan toplam sayaç sayısı, Türkiye'deki diğer bölgelerdeki tüm illerin toplamının 2 katından fazladır. Kayıp kaçak ile mücadele kapsamında, abonelerin OSOS sayaçları takılması işlemleri artarak devam etmektedir.

OSOS verilerinde, Abone sayaçları + Yedek Sayaçlar + Toplayıcı sayaçlar yer almaktadır. Proje kapsamında sadece abone sayaçları kullanılacaktır.

Verinin çok hızlı büyümesiyle birlikte, yönetimi de günden güne zorlaşmaya başlamıştır. Bu kapsamda birden fazla sistem ve yapı yer almaktadır. Mevcut OSOS durumu aşağıdaki gibidir:

- 2013 Aralık tarihinden sonra HAYEN sistemleri ile OSOS başlamıştır.
- 2017 Mayıs itibarıyla HAYEN sisteminden ARIL sistemine geçiş yapılmıştır.
- HAYEN-ARIL geçişi ile birlikte aboneler yeni sisteme aktarılırken, eski tüketim verileri eski sistemde kalmıştır.
- 2016 Ocak ayı itibarıyla, LUNA-METRUM sistemi de aktif olarak başlamıştır.
- LUNA-METRUM sistemi MS SQL veri tabanında, yatay sayaç projesi kapsamındaki ağırlıklı mesken abonelerine ait tüketim verilerini içermektedir.
- Güncel ARIL sistemi ise Oracle veri tabanında, ağırlıklı yüksek tüketimli, trafolu abonelerin verisini içermektedir.

Tablo 3.2: Örnek OBIS Kodları ve Açıklamaları

OBIS Kodu	Açıklama
0.0.0	Seri Numarası
0.0.0	Landis Seri Numarası
0.1.0	Reset Tarihi
0.1.0	Landis Reset Sayısı
0.1.0*01	Landis Reset Tarihi 01 Ay Önceki
0.1.0*02	Landis Reset Tarihi 02 Ay Önceki
0.1.0*12	Landis Reset Tarihi 12 Ay Önceki
0.1.2	Landis Reset Saati
0.1.2*01	1 Ay Önceki Reset Tarihi
0.1.2*8	8 Ay Önceki Reset Tarihi
0.1.3	Landis Reset Tarihi
0.8.0	Demand Periyot
0.9.1	Sayaç Saati
0.9.1	Landis Sayaç Saati
0.9.2	Sayaç Tarih
0.9.2	Landis Sayaç Tarihi
0.9.5	Gün
1.01.2001	Seri no okuma Güncelleme
1.06.2000	Demand
1.06.2000	Landis Demand
1.6.0*01	1 Ay Önceki Demand
1.6.0*01	Landis Demand 01 Ay Önceki
1.6.0*02	2 Ay Önceki Demand
1.6.0*02	Landis Demand 02 Ay Önceki
1.6.0*19	19 Ay Önceki Demand
1.6.0*2	2 Ay Önceki Demand
1.6.0*20	20 Ay Önceki Demand
1.6.0*24	24 Ay Önceki Demand

Yukarıdaki durum ışığında, birden fazla OSOS sistemi ile çalışma yapılmıştır. Kullanılan sistemler ve veri tabanları her ne kadar farklı olsa da, farklı yapıda benzer verileri tutmaktadır.

LUNA-METRUM sistemi analizi kapsamında, OSOS Müdürlüğü ile birlikte veri tabanı düzeyinde çalışmalar gerçekleştirilerek, tüm yapı teknik olarak analiz edilmiştir. OSOS sayaç verileri, “Readout” ve “Yük Profili” verisi olmak üzere iki şekilde alınmaktadır. Tüm alınan verilere ait bir OBIS Kodu tanımlıdır. Tablo 3.2’de, yaklaşık 1000 adet olan OBIS Kodlarından örnekler paylaşılmıştır.

3.2.2.1 Sayaç Readout Verileri

Readout Verileri, sayaçtan T anında alınan bilgilerdir. Sayaç kapağının açılması, enerjinin kesilmesi gibi olaylara ait bilgiler bu şekilde tutulmaktadır. Readout bilgileri günde bir kez T anında alınarak, veri tabanına kaydedilmektedir. LUNA-METRUM da readout verisi, OkumaTipi=11 olarak kaydedilmektedir.

Dicle OSOS sisteminde, gece saat 02:00 da tüm sayaçlar için merkezi sistemden Readout isteği gönderilmektedir. Okunamayan sayaçlar için saat başı bu istek tekrarlanmaktadır. Readout verileri alındığında veri tabanına OBIS Kodları ile kaydedilmektedir. Bir sayaçta yaklaşık 100 adet OBIS Kodu yer almaktadır.

3.2.2.2 Sayaç Yük Profili Verisi

Yük profili verisi, sayaçların otomatik olarak kaydettiği verilerdir. Bu veriler sayacın kendi belleğinde tutulmakta olup, bir sayaç kendi belleğinde son 6 ayın verisini içermektedir. Sayacın markasına ve üreten firmanın periyoduna göre, yük profili verileri 15 dakika, 30 dakika ve 1 saatlik periyotlarla OBIS Kodlarına göre sayaç belleğine kaydedilmektedir. LUNA-METRUM da yük profili verisi, OkumaTipi=22 olarak kaydedilmektedir.

Dicle OSOS sisteminde, merkezi sistemden gece saat 03:00 da tüm sayaçlar için bir önceki günün 24 saatine ait yük profili verileri isteği gönderilmektedir. Yük Profili verileri alındığında veri tabanına OBIS Kodları ile kaydedilmektedir. Sayaçların zaman saatleri farklılık gösterebilmektedir.

Yük profili verilerinin tamamı ortalama verileri içermektedir. Yatay sayaç projesinde yer alan tüm sayaçlar, LUNA-METRUM sisteminde yer almaktadır.

3.2.2.3 LUNA Veri Tabanı Yapısı

LUNA SQL Server veri tabanında proje kapsamında kullanılacak veri tabanı tabloları ve açıklamaları, Tablo 3.3'te gösterilmiştir. OBISKODU tablosunda, tanımlı tüm OBIS Kodları ve açıklamaları yer almaktadır.

ABONE tablosunda, abonelere ait master OSOS bilgileri yer almaktadır. Projede master verilerin kaynağı MBS olarak belirlendiğinden, öncelikli olarak MBS verisi master olarak kullanılacaktır.

SAYAC tablosunda, sayaç markası, modeli, sayaç durumu, sayaç numarası, sayaç çarpanı, takılma tarihi gibi sayaç tanım bilgileri yer almaktadır. ABONEGECMISI tablosunda bir abonedeki eski ve yeni sayaç bağlantı bilgilerine ulaşılabilir.

OBISOKUMA tablosu, aynı isimli (OBISOKUMA1, OBISOKUMA2...gibi) 7 tabloya ait ham verileri içeren tablodur. Okunan akım, voltaj, tüketim, cosinüs...gibi değerlerin tamamı bu tabloda yer almaktadır.

OKUMASAATLIK tablosu, OBISOKUMA tablosundan oluşturulan ve saatlik olarak okuma verilerini içeren tablodur.

OKUMAGUNLUK tablosu, OKUMASAATLIK tablosundan oluşturulan ve günlük olarak okuma verilerini içeren tablodur. Her abonenin günlük toplam tüketim değerleri bu tabloda yer almaktadır.

Tablo 3.3: LUNA Veri Yapıları ve Açıklamaları

Tablo Adı	Tablo Açıklaması
OBISKODU	OBIS Kodları ve Açıklamaları
ABONE	Abone Master Bilgileri
SAYAC	Sayaç Bilgileri
OBISOKUMA	Tüm okumalar özet tablo
OKUMASAATLIK	Saat bazında okunan sayaç bilgileri değerleri
OKUMAGUNLUK	Gün bazında okunan sayaç bilgileri değerleri
ABONEGECMISI	Aboneye ait sayaç değişimleri bilgileri

3.3 Veri Toplama

Proje kapsamında, farklı sistemlerdeki farklı veri tabanlarından seçilen veriler toplanarak konsolide edilmiştir. Tablo 3.1’de projede kullanılan sistemler ve veri tabanı bilgileri yer almaktadır.

Müşteri Bilgileri Sisteminden, abonelere ait demografik veriler alınırken; OSOS 1 ve OSOS 2 Otomatik Sayaç Okuma Sistemleri’nden elektriksel ölçüm sayısal verileri, CBS – Coğrafi Bilgi Sistemlerinden abonelerin lokasyon bilgileri, SAP sisteminden ise saha operasyonlarına ait kaçak bilgileri alınmıştır.

Bu kapsamda; MBS, OSOS 1, OSOS 2, CBS sistemlerinden elde edilen bilgiler girdi (öznitelik/parametre) olarak, SAP – Saha Yönetim Sisteminden alınan bilgiler ise çıktı olarak kullanılmıştır.

3.4 ETL ve Veri Aktarımı

3.4.1 ETL Nedir?

ETL, İngilizce Extract (Çıkart), Transform (Dönüştür) ve Load (Yükle) kelimelerinin baş harflerinden oluşmaktadır. Kullanılacak verinin dış kaynaklardan çıkarılması, işlenip temizlenerek istenilen yapıya dönüştürülmesi ve sonrasında istenilen ortama (veri ambarı veya veri tabanı...) yüklenmesi işlemidir.

Tablo 3.1’ de gösterilen Oracle ve MS SQL veri tabanlarından çıkartılan veriler, ilgili dönüşüm işlemleri sonrasında işlenerek, yeni oluşturulan MS SQL veri tabanına yüklenmiştir. Böylelikle işlenen tüm veriler, makine öğrenmesi algoritmalarında kullanılmak üzere MS SQL veri tabanında oluşturulan veri ambarına aktarılmıştır.

Aktarımlar “SQL Server Export and Import Data” uygulaması kullanılarak gerçekleştirilmiştir. Oracle veri tabanlarından aktarımlar için “.Net Data Provider For Oracle”, MS SQL veri tabanlarından aktarımlar için “SQL Server Native Client 11.0” bağlantı yöntemleri kullanılmıştır. İlgili veritabanlarına erişim için kullanılan araçlar ve bağlantı yöntemleri, Tablo 3.4’te gösterilmiştir.

Tablo 3.4: Veritabanı Bağlantısı Yöntem ve Araçları

Veritabanı	Bağlantı	Bağlantı Aracı
Oracle	.Net Data Provider For Oracle	Toad For Oracle
MS SQL	SQL Server Native Client 11.0	SQL Server Management Studio

3.4.2 Veri Ambarının Oluşturulması

MS SQL üzerinde DBKACAK isminde bir veri ambarı oluşturularak, tüm sistemlerden gelen veriler bu veri ambarı altında toplanmıştır. Tablo 3.5'te DBKACAK veri ambarında oluşturulan ve diğer veri tabanlarından aktarılan verilerin yer aldığı, veri tabanı tablo isimleri yer almaktadır.

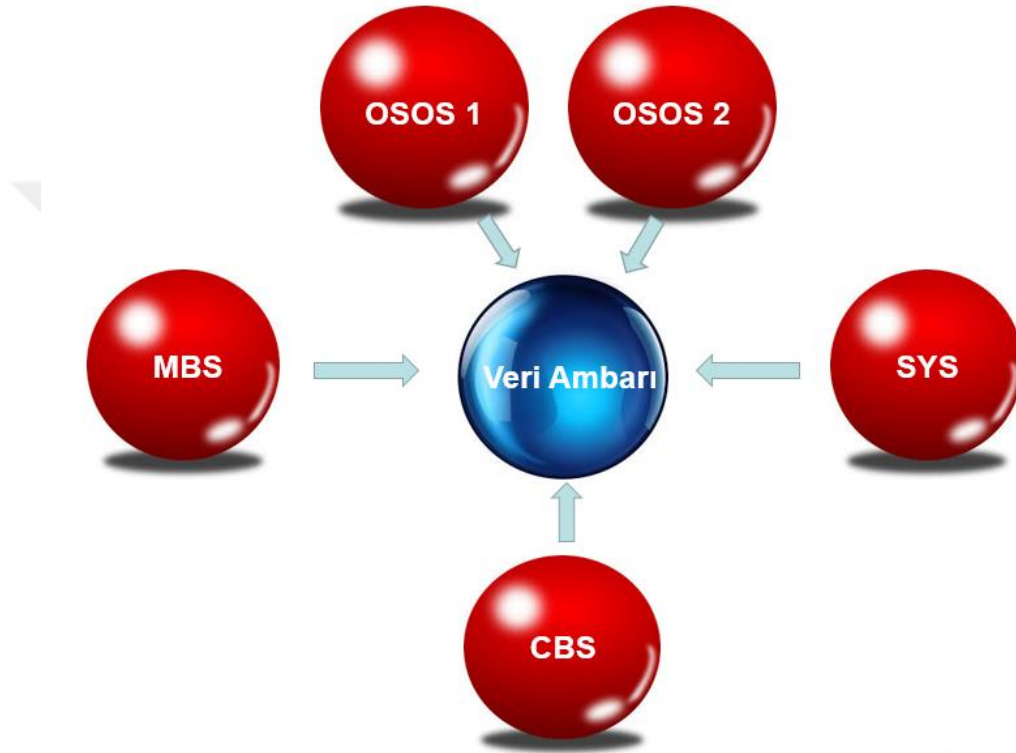
Tablo 3.5: Kaçak Veri Ambarı Tabloları

Tablo Adı	Tabloda Tutulan Veriler
MBSInfoFull	MBS ve CBS Demografik Bilgileri
OSOSFull	Akım, Gerilim, Cosinus Sapma
OSOSFull2	Ortalama Günlük Tüketim, Max Demand
SysInfoFull	Saha Kaçak Bilgileri

MBSInfoFull tablosu, MBS-Müşteri Bilgi Sistemi ve CBS-Coğrafi Bilgi Sistemleri'nden alınan demografik verilerin bulunduğu tablodur. İki farklı Oracle veri tabanında yer alan tablolardan aktarım yapılarak oluşturulmuştur.

OSOSFull tablosu, biri MS SQL, diğeri Oracle veri tabanından alınan OSOS-Otomatik Sayaç Okuma Sistemleri verilerinden oluşturulmuştur. Akım, gerilim ve cosinus açısı sapma değerleri gibi elektriksel değerleri içermektedir. OSOSFull2 tablosu da aynı şekilde, MS SQL ve Oracle veri tabanlarından alınan OSOS-Otomatik Sayaç Okuma Sistemleri abone tüketim verilerinden oluşturulmuştur. Ortalama günlük tüketim, max demand değerlerini içermektedir. İki farklı OSOS'a ait verilerden, tüketim ile ilgili olan veriler bir veri ambarında; akım gerilim verileri de diğer bir veri ambarında tutulduğundan, OSOSFull ve OSOSFull2 olarak iki ayrı tabloya aktarım gerçekleştirilmiştir. OSOSFull abone tüketim verilerini, OSOSFull2 ise akım gerilim değerlerini içermektedir.

Saha kaçak tespitlerinin yer aldığı bilgiler, SAP-SYS sisteminden alınarak, SysInfoFull tablosuna aktarılmıştır. SysInfoFull tablosu abonelere ait saha operasyonlarının sonuçlarını içermektedir. Şekil 3.1’de oluşturulan veri ambarına, hangi sistemlerden verilerin aktarıldığı gösterilmiştir. Veri aktarımı öncesinde aktarılamayan veriler için, “Veri Temizliği” bölümünde belirtilen veri temizliği operasyonları gerçekleştirilmiştir.



Şekil 3.1: KTS Veri Ambarı Yapısı

3.5 Değerlendirme

Dicle EDAŞ uygulama sistemlerinde yer alan ve farklı veri tabanlarında tutulan veriler, yeni oluşturulan KACAKDB veri ambarına aktarılmıştır. Aktarılan veriler; biri MSSQL, diğeri Oracle olmak üzere, iki büyük veri tabanında yapısal (structured) olarak tutulmaktadır. ETL yöntem ve araçları ile iki farklı veri tabanında yer alan bu veriler, MSSQL veri tabanında oluşturulan veri ambarına taşınmıştır.

Oluşturulan yeni veri ambarına taşıma işlemleri, oldukça detaylı bir çalışma olup, her bir veri tabanı ve veri seti için ayrı ayrı çalışmalar yapılmıştır. Ayrıca farklı veri tabanlarına erişim ve veri aktarımları için, farklı araçlar (tool) kullanılmıştır. Her bir

veri tabanı için kullanılan SQL dili de, benzer yapıda olmasına rağmen, detayda sözdizimi (syntax) farklılıkları göstermektedir. Aktarım için kullanılan SQL scriptleri de, veri tabanlarına bağlı olarak farklı sözdizimlerinde yazılmıştır.

Ayrıca veri aktarımları sonrası, aktarılan verilerin doğruluğunun teyidi için, SQL ile kontrol betikleri oluşturulmuştur. Bu kontrollerde farklar çıktığında farklılıkları giderme noktasında, aktarılmayan kayıtların tekrar aktarılması için, veri temizliği vb. çalışmalar gerçekleştirilmiştir. Aktarılan verilerin büyüklüğüne istinaden, gerektiğinde oluşturulan ara tablolar ile, tablo ayırma ve tablo birleştirme yöntemleri uygulanmıştır. Tutarsız tablo verileri ile karşılaşıldığında, veri tutarlılığının sağlanması ile ilgili gerekli önlemler alınmıştır. Tutarsız verilerin tutarlı hale getirilebilmesi için gerekli veri güncellemeleri veya bozuk verilerin temizlenmesi/silinmesi operasyonları gerçekleştirilmiştir. Aktarıma engel problemleri tek bir satır veri bile bırakılmamıştır.

SQL betiklerinin yazılması ve çalıştırılması, veri kontrol sonuçları, tutarsız verilerin tespiti ve veri bütünlüğünün sağlanması işlemleri gibi yapılan çalışmaların tamamı detaylı olarak ayrı dosyalarda takip edilmiştir.

Tüm bu detaylı çalışmalardan sonra, DBKACAK veri ambarına aktarılan veriler, makine öğrenmesi modellerinin oluşturulabilmesi amacıyla kullanılabilir hale getirilmiştir. Özellikle çalışmanın bu aşamasındaki verilerin anlamlandırılması ve işlenerek hazır hale getirilmesi adımları, projenin zaman maliyeti olarak %70 ini oluşturmaktadır.

DÖRDÜNCÜ BÖLÜM

KTS : MODELİN OLUŞTURULMASI

KTS oluşturulması kapsamında veriler hazırlandıktan sonra, hazırlanan bu veriler ile model oluşturma ve algoritmaların bu veriler ile eğitilmesi çalışmalarına geçilmiştir. Bu bölümde KTS öncesi mevcutta kullanılan yaklaşım ve KTS ile önerilen yaklaşım ortaya konmuştur. Önerilen yaklaşıma istinaden oluşturulan model ve bu modelde kullanılan algoritmalar belirlenerek, projede uygulanmıştır.

4.1. Mevcut Yaklaşım

Klasik uzman sistem yaklaşımında, sahada kazanılan tecrübeler ışığında algoritmalar oluşturulur. Bu algoritmalar statik algoritmalarlardır. Dicle EDAŞ bünyesinde, kayıp kaçak ile mücadele kapsamında yıllar boyunca elde edilen bilgiler doğrultusunda bir uzman sistem yaklaşımı oluşturulmuştur.

Bu bölümde uzman sistem yaklaşımında kayıp kaçak tahmininde kullanılan verilere yer verilmiştir. Kaçak tahmininde kullanılan parametreler, ağırlıklarına göre sıralanarak, bir ağırlık tablosu oluşturulmuştur. Oluşturulan bu ağırlık tablosu kaçak tahmininde kullanılmıştır. Örnekler; içerdikleri parametrelerin ağırlıklarına göre toplanarak, toplamda en yüksek puanı alan örnekten, en düşük puanı alan örneğe göre sıralanmıştır. Kaçak tahmininde, en yüksek puanlı örneklerden başlanarak ilerlenmiştir.

Bu çalışmada en az bir parametre değerinin olması bile örneğin kaçak tahmin listesine girmesine izin vermektedir. Bu durumda büyük bir liste oluşmakta ve sahada kaçak tahmininde bu listenin tamamına bakılamamakta, yukarıdan aşağıya doğru başlanarak gelinebilen noktaya kadar ilerlenmektedir.

4.1.1 Gerilim Uyarıları

Gerilim uyarıları, yüksek gerilim, alçak gerilim ve sıfır gerilim uyarısı olmak üzere üç farklı uyarıdan oluşmaktadır. Belirli tipte abonelerde, voltaj değerinin belirli bir sınırın altında kalması, üstüne çıkması veya belirli bir aralıkta değerinde sıfır olması durumlarında gerilim uyarıları oluşmaktadır.

4.1.1.1 Gerilim Yüksek Uyarısı

Primer abonelerde, herhangi bir fazın voltaj değeri, 3 saat art arda 65 Volt ve üstündeyse gerilim yüksek uyarısı oluşmaktadır.

Sekonder abonelerde, herhangi bir fazın voltaj değeri, 3 saat art arda 270 Volt ve üstünde ise gerilim yüksek uyarısı oluşmaktadır.

4.1.1.2 Gerilim Düşük Uyarısı

Primer abonelerde herhangi bir fazın voltaj değeri, 3 saat art arda 40 Volt ve altında ise ve ayrıca sifıra eşit değilse, gerilim düşük uyarısı oluşmaktadır.

Sekonder abonelerde herhangi bir fazın voltaj değeri 3 saat art arda 150 Volt ve altında ise ve ayrıca sifıra eşit değilse, gerilim düşük uyarısı oluşmaktadır.

4.1.1.3 Gerilim Sıfır Uyarısı

Primer ve Sekonder abonelerde, en az bir en fazla iki fazın voltaj değeri 3 saat art arda sifıra eşit olması durumunda gerilim= 0 uyarısı oluşmaktadır.

4.1.2 Akım Uyarıları

Akım uyarıları da akımın olması gerekenden düşük olması veya akım değeri sıfır olmak üzere hiç akım ölçülememesi durumudur. Ayrıca akım değerlerindeki yüksek artışlar, tarımsal sulama aboneleri için kaçak kullanım şüphesi ile manyetik ve nötr uyarıları oluşturmaktadır.

4.1.2.1 Akım Düşük Uyarısı

Primer abonelerde herhangi bir fazdaki akım değeri, 3 saat art arda olarak 0'dan büyük ve 0,2'den küçük olduğu ($0 < \text{Akım} < 0,2$) durumlarda akım düşük uyarısı oluşmaktadır.

Sekonder abonelerde herhangi bir fazdaki akım değeri 3 saat art arda olarak 0'dan büyük ve 0,5'ten küçük olduğu ($0 < \text{Akım} < 0,5$) durumlarda akım düşük uyarısı oluşmaktadır.

4.1.2.2 Akım Sıfır Uyarısı

Herhangi bir fazdaki akım değeri 3 saat art arda sıfıra eşit olduğunda ve aynı fazda gerilim sıfıra eşit olmadığı durumlarda akım=0 uyarısı oluşmaktadır.

4.1.2.3 Tarımsal Şüpheli Akım Uyarısı

Sayacın üç fazına ait akım değerleri ayrı ayrı olmak üzere, akım değerleri aynı saatteki akım faz toplamına bölünerek, 3'er saatlik toplamları alınıp, fazlara dağılan yüzdelikleri hesaplanır. Herhangi bir fazın yüzdesi %25'in altında ise "Tarımsal Şüpheli Akım Uyarısı" oluşmalıdır.

4.1.2.4 Manyetik veya Nötr Uyarısı

Gün içerisinde birbirini takip eden saatlerdeki akım değerleri arasında, sonraki akım değeri önceki akım değerinin 100 katı ve üzerinde bir değer ise, Manyetik veya Nötr Uyarısı oluşur.

Ayrıca baz istasyonu, resmi kurumlar ve aydınlatma aboneleri dışındaki aboneler için, 1 saatten fazla manyetik müdahale uyarısı veren aboneler için Manyetik veya Nötr Uyarısı oluşur.

4.1.3 Cosinus Uyarıları

Cosinus uyarıları, OSOS'tan ölçülen cosinus değerlerinin anormallik gösterdiği durumlardır.

4.1.3.1 Cos Sıfır Uyarısı

Herhangi bir fazda akım değeri primer aboneler için 0,2 A, sekonder aboneler için 0,5 A'den büyük iken, aynı faza ait cos değeri 3 saat art arda sıfıra eşit olduğunda Cos= 0 uyarısı oluşur.

4.1.3.2 Cos Düşük Uyarısı

Herhangi bir fazda akım değeri, primer aboneler için 0,2 A, sekonder aboneler için 0,5 A'den büyük iken, aynı faza ait cos değerinin 3 saat art arda -0.5 ve +0.5 arasında olması ve sıfıra eşit olmaması durumunda, Cos düşük uyarısı oluşmaktadır.

4.1.3.3 Tarımsal Şüpheli Cos Uyarısı

Sayacın üç fazına ait cos değerleri ayrı ayrı olmak üzere, cos değerleri aynı saatteki cos faz toplamına bölünerek, 3'er saatlik toplamları alınıp, fazlara dağılan yüzdeleri hesaplanır. Herhangi bir fazın yüzdesi %25'in altında ise "Tarımsal Şüpheli Cos Uyarısı" oluşmalıdır.

4.1.3.4 Tarımsal Düşük Cos Uyarısı

Tarımsal harici abonelerde olduğu gibi, herhangi bir veya tüm fazın akımı 3 saat art arda çarpansız olarak 0,5 Amperden büyük iken, aynı faza ait cos değeri -0,5 ile 0,5 arasında ise "Tarımsal Düşük Cos Uyarısı" oluşmalıdır.

Not: Tarımsal aboneler için; gerilim, manyetik ve geri kalan tüm uyarılar, tarımsal harici aboneler ile aynı kriterlere sahiptir.

4.1.4 Diğer Uyarılar

Gerilim, akım ve cosinus değerleri ölçümleri dışında kalan uyarılardır. Demand ve tüketim değerleri ile, gövde kapağı uyarısı bilgilerinden oluşur.

4.1.4.1 Endeks Düşük Uyarısı

Okuma sırasına göre, bir saat sonraki endeks değerinin sayaç değişimi olmaksızın bir önceki değerden daha düşük olması durumunda Endeks Uyarısı oluşmaktadır. Eğer sayaç arızalı değilse, endeks değeri tüketime bağlı olarak sürekli artan bir değerdir ve azalması mümkün değildir. Endeks değerinde azalma olması, sayaca müdahale edildiğinin bir göstergesi olabilir.

4.1.4.2 Gövde Kapağı Uyarısı

OSOS'ta gövde kapağı uyarısı veren sayaçların mevcut aboneleri için, "Gövde Kapağı Uyarısı" oluşmaktadır. Gövde kapağı uyarısı, sayaç gövdesinde yer alan kapağın açıldığı anlamına gelmektedir. Uzaktan okunan sayaçların gövde kapaklarının normal şartlar altında kapalı olması, açılmaması gerekmektedir. Sayaçlar ancak arıza veya sayaç kontrol amacıyla yetkili saha personelleri tarafından, ilgili ekiplerin bilgisi dahilinde açılmaktadır. Bu uyarı sayaca müdahale olabileceği anlamına gelmektedir.

4.1.4.3 Kaçak Sonrası Tüketim Uyarısı

Her abone için, kaçak öncesi, sonrası ve dönemsel tüketimler kıyaslanacaktır.

1. Kaçağın tutulduktan sonraki dönem sayısı ile öncesindeki fatura sayıları eşit olmalıdır.
2. Kaçak sonrası tüketim ortalamasının, kaçak öncesi tüketim ortalamasına göre %30 artması beklenmektedir.

Yukarıdaki iki şarta uymayan aboneler için “Kaçak Sonrası Tüketim Uyarısı” oluşmalıdır.

4.1.4.4 Sayaç Değişim Sonrası Tüketim Uyarısı

Sayaç değişimi yapılan ve sayaç inceleme sonucu kaçak çıkan abonelerde, ticari aboneler için, abone sektör değişimi veya güç artışı olmadığı halde tüketim artmamişsa, “Sayaç Değişim Sonrası Tüketim Uyarısı” oluşmalıdır.

4.1.4.5 Demand Uyarısı

Abonenin maksimum Demand değerinin son Demand değerine oranı 5 ten büyük ise “Demand Uyarısı” oluşmaktadır.

NOT: Sahadaki bazı sayaçlar (TF19,TF20 gibi...) ile modemler arasında uyumsuzluk bulunmaktadır. Bu uyumsuzluk sonucunda akım, cos ve gerilim değerleri sisteme hatalı bir şekilde yansımaktadır. Hatalı değerler abonenin şüpheli konumuna düşmesine neden olmaktadır. Bu aboneler veri temizliği kapsamına alınarak, çalışmalarda özel olarak değerlendirilmiştir.

Sahada yapılan çalışmalar sonucunda belirlenen uyarılara ait puan tablosu aşağıdaki gibidir. Puanlar tamamen saha tecrübelerine istinaden belirlenmiştir. Puanlar, makine öğrenmesinde kullanılan özniteliklerin ağırlıkları olarak da düşünülebilir.

Mevcut uzman sistem yaklaşımında, Tablo 4.1 kullanılmaktadır. Bu tabloda, sahanın bilgi birikiminden elde edilen uyarılar, ağırlıklarına göre puanlanmıştır. Her bir uyarının bir ağırlık değeri vardır. Sistemde birden fazla uyarı geldiğinde, her bir uyarıya karşılık gelen ağırlık değerleri toplanarak, toplamda en fazla puana sahip abonelere kaçak takibinde öncelik verilmiştir. Bu tabloya göre “Gerilim Uyarısı”,

“Cos Uyarısı” ve “Gövde Kapağı Uyarısı”, mevcut uyarılar içinde en yüksek ağırlığa sahip olup, bu uyarılara sahip abonelerde kaçak kullanım ihtimalinin yüksek olduğu anlamına gelmektedir. Örneğin saha operasyonları haricinde OSOS sayaçlarının kapaklarının açılması, güçlü bir kaçak kullanım şüphesi olarak değerlendirilmektedir.

Tablo 4.1: Kaçak Uyarıları Aralık Değerleri

Uyarı Adı	Puan
Gerilim Uyarısı	5
Akım Uyarısı	4
Cos Uyarısı	5
Manyetik veya Nötr Uyarısı	3
Endeks Düşme Uyarısı	1
Gövde Kapağı Uyarısı	5
Kaçak Sonrası Tüketim Uyarısı	3
Sayaç Değişim Sonrası Tüketim Uyarısı	3
Demand Uyarısı	3

Bu yaklaşımda sadece belirlenen uyarılar dikkate alınmaktadır ve bu uyarılar ağırlıklarına göre kaçak tahminine etki etmektedir. Belirlenen uyarılar dışında kalan ve kaçak kullanan aboneler, kapsam dışında kalacağından, bu abonelerin tespitinde uzman yaklaşımı yetersiz kalmaktadır.

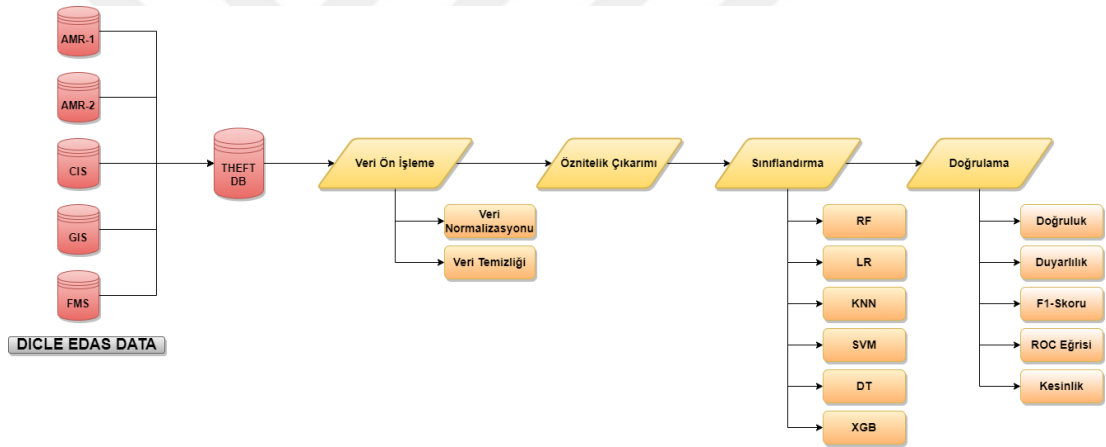
4.2 Önerilen Yaklaşım: Makine Öğrenmesi

Mevcut yaklaşım bir uzman sistem yaklaşımı olup, tamamen sahanın tecrübeleri doğrultusunda geliştirilen algoritmalara dayanmaktadır. Bu yaklaşım statik bir yaklaşım olup, tüm veriler değerlendirilmemektedir. Yine de sahanın tecrübesinin aktarılması, milyonlarca abone içinde doğru abonelerin önceliklendirilmesi anlamında büyük fayda sağlamaktadır.

Bu çalışmada önerilen yaklaşım ise makine öğrenmesi yöntemlerini kullanarak, kaçak tahmininin gerçekleştirilmesidir. Makine öğrenmesi modeli ile yapılacak kaçak abonelerin tahmin edilmesi çalışması, geçmiş verilerden öğrenerek, daha doğru kaçak tahmini yapabilme imkanı sağlamaktadır. Özellikle büyük veri içerisinde doğru verilerin seçilmesi, verilerin birbirleri ile olan ilişkilerinin değerlendirilmesi; uzman sistemlere göre avantaj sağlamaktadır. Böylelikle detayda dikkate alınmayan bazı

durumlar, makine öğrenmesi ile değerlendirilerek kaçak tahmininde kullanılabilir. Bu bölümde, makine öğrenmesi ile değerlendirilerek kaçak tahmininde kullanılabilir.

Şekil 4.1, KTS Framework yapısını göstermektedir. Veriler farklı sistemlerden Bölüm 3'te detaylandırılan yöntemlere alınarak, KTS için oluşturulan kaçak veri ambarına aktarılmıştır. Kaçak veri ambarındaki veriler üzerinde, veri ön işleme işlemleri uygulanmaktadır. Sonraki adımda veri setinden öznitelikler çıkarılmaktadır. Veri setinden çıkarılan öznitelikler, sınıflandırma algoritmalarında girdi olarak kullanılmaktadır. Bu bölümde sırasıyla; veri ön işleme adımları, öznitelik çıkarımları ve sınıflandırma algoritmalarının uygulanması adımlarına yer verilmiştir. Doğrulama adımında ise Doğruluk, Duyarlılık, F1-Skoru, Kesinlik değerleri ile ROC Eğrisi kullanılmıştır. Doğrulama adımındaki sonuçlar ve değerlendirmeler, 5.Bölüm kapsamında yer almaktadır.



Şekil 4.1: KTS Framework

4.2.1 Veri Ön İşleme

3. bölümde farklı sistemlerden toplanan veriler temizlenerek, ETL araçları ile veri ambarına aktarılmıştı. Bu bölümde ise veri ambarının ilgili tablolarından alınan veriler, veri ön işleme kapsamında veri temizliği ve veri normalizasyonu işlemlerinden geçirilerek, oluşturulacak makine öğrenmesi modelinde kullanılabilir hale getirilmiştir. Veri temizliği kapsamında kirli veri çeşitli yöntemlerle temiz hale getirilirken, veri normalizasyonunda ise veri ölçeklenerek işlenebilir hale getirilmiştir.

4.2.1.1 Veri Temizliği

Farklı sistemlerdeki veri tabanlarından elde edilen tablolardaki hatalı veya eksik

verilerde, veri temizliđi operasyonları gerekleřtirilmiřtir. Bu kapsamda mükerrer verileri tekilleřtirme, eksik verileri tamamlama, hatalı verileri düzeltme, sayısal olmayan veri tiplerini sayısal hale getirme iřlemleri uygulanmıřtır.

Veri Temizliđi iřlemleri, MSSQL DB tarafında “SQL Script” dilinde yazılan SQL kodları ile sađlanmıřtır. Veri temizliđi kapsamında yapılan iřlemler, ařađıdaki gibidir.

- Mükerrer Verilerin Temizlenmesi
- Hatalı Verilerin Silinmesi
- Hatalı Veri Tiplerinin Deđiřtirilmesi
- Hatalı Verilerin Deđiřtirilmesi

Elde edilen verilerde, tekrarlanan veya aynı veriyi ifade eden alanlarda, verilerin tekilleřtirilmesi sađlanarak, mükerrer veriler temizlenmiřtir. Örneđin İl alanında yer alan “URFA” ve “řANLIURFA” verileri, “řANLIURFA” olarak birleřtirilmiřtir. Bu kapsamda “URFA” ieren veriler, “řANLIURFA” olarak güncellenmiř olup, SQL kodu ařađıdaki gibidir.

```
Update MBSData set Province='řANLIURFA' Where PROVINCE='URFA'
```

Anahtar alan deđerleri hatalı (sayısal yerine karakter, boř deđer, null) deđerler ieren kayıt satırları temizlenir. Abone No alanı ile ilgili örnek SQL kodları ařađıda gösterilmiřtir. Benzer řekilde ilgili alanlarda hatalı veriler ieren kayıtlar silinmiřtir.

```
Delete From OSOSData Where IsNumeric(ABONENO)=0
```

```
Delete From SYSDData Where isnull(ABONENO,0)=0
```

Sayısal olması gerekirken, karakter olarak aılan alanlar sayısal alanlara dönüřtürülmüřtür. Bu ařamada hatalı veriler de temizlenmiřtir. İlgili alanlar veri ierdiđinden, öncelikle dođru tipte yeni bir alan aılarak, veriler bu alana aktarılmıř, sonrasında aktarılan verilerin bulunduđu alan tablodan silinerek, yeni alanın adı eskisi ile aynı olacak řekilde deđiřtirilmiřtir. Bu yöntemle hatalı veri tipleri düzeltilmiřtir.

```
Alter Table OSOSData Add ABONENO2 numeric(10,0)
```

```
Update OSOSData set ABONENO2=ABONENO
```

```
Alter Table OSOSData Drop Column ABONENO
```

Alter Table OSOSData Change ABONENO2 ABONENO

Anahtar alan deęerleri doęru olan satırlarda, eksik veya hatalı bilgi ieren kolonlardaki veriler deęiřtirilmiřtir. rnek kodlar ařaęıdaki gibidir.

```
Update Thefts set SayacModeli='YOK' Where isnull(SayacModeli,"")=
```

```
Update Thefts set SayacModeli=Trim(SayacModeli)
```

```
Update Thefts set OrtalamaAkim1Ay1=0 Where isnull(OrtalamaAkim1Ay1,-1)=-1
```

```
Update Thefts set OrtalamaAkim2Ay1=0 Where isnull(OrtalamaAkim2Ay1,-1)=-1
```

```
Update Thefts set OrtalamaAkim3Ay1=0 Where isnull(OrtalamaAkim3Ay1,-1)=-1
```

```
Update Thefts set OrtalamaGerilim1Ay1=0 Where isnull(OrtalamaGerilim1Ay1,-1)=-1
```

```
Update Thefts set OrtalamaGerilim2Ay1=0 Where isnull(OrtalamaGerilim2Ay1,-1)=-1
```

```
Update Thefts set OrtalamaGerilim3Ay1=0 Where isnull(OrtalamaGerilim3Ay1,-1)=-1
```

Veri aktarımı ile DBKACAK tablosunda oluřturulan; MBSInfoFull, OSOSFull, OSOSFull2, SysInfoFull tablolarındaki veriler zerinden THEFTS veri ambarı tablosu oluřturulmuřtur. Makine ęrenmesi algoritmaları, nihai tablo olan THEFTS tablosu zerinde geliřtirilmiřtir.

Phyton'da pyodbc ktphanesi kullanılarak, MSSQL direkt srcs (native driver) Theft tablosuna baęlantı saęlanmıřtır. Tm veriler, bu baęlantı zerinden, pandas ktphanesi kullanılarak "KacakTahminVerileri" veri erevesine (DataFrame) alınmıřtır.

KacakTahminVerileri'nde yer alan kategorik veriler, nmerik verilere dnřtrlmřtir. Bu kapsamda; kategorik veri ieren Abone Grubu, lm Durumu, İl, Saya Modeli ve Kaak verileri, nmerik verilere dnřtrlmřtir. Bu dnřm esnasında; sklearn ktphanesinde yer alan veri n iřleme yntemlerinden Label Encoding, One Hot Encoding yntemleri kullanılmıřtır. Paralara ayrılarak, veri dnřm gerekleřtirilen veriler, pandas ktphanesinde yer alan komutlar ile tekrar birleřtirilmiřtir.

Tablo 4.2'de MBSInfoFull tablosundan aktarılan, demografik verilerden İl alanına ait kategorik verilerin, nmerik hale dnřm sonrası rnek verilerini gstermektedir.

Tablo 4.2: İl Değişkeni Kategorik-Nümerik Dönüşümü

Index	ŞIRNAK	DİYARBAKIR	ŞANLIURFA	SİİRT	MARDİN	BATMAN
0	0	1	0	0	0	0
1	0	0	1	0	0	0
2	0	1	0	0	0	0
3	0	0	0	1	0	0
4	0	0	0	1	0	0
5	0	0	0	0	1	0
6	0	0	0	0	1	0
7	1	0	0	0	0	0
8	0	0	0	0	0	1
9	0	1	0	0	0	0
10	0	0	1	0	0	0

4.2.1.2 Veri Normalizasyonu

Özellik ölçeklemenin önemli olduğu algoritmalar aşağıdaki gibidir [26]:

- K-En Yakın Komşu Algoritması (Özellikle Öklid Mesafesi ile hesaplamada)
- Lojistik Regresyon
- Karar Destek Makinası
- Yapay Sinir Ağları
- K-Ortalama Algoritması
- LDA – Linear Discriminant Analysis (Doğrusal Ayrımcılık Analizi)
- PCA - Principal Component Analysis (Temel Bileşenler Analizi)
- Kernel Principal Component Analysis (Çekirdek Temel Bileşenler Analizi).

Naive Bayes, Karar Ağaçları ve Rassal Orman gibi algoritmalarda ise öznitelik ölçeklemenin önemi yukarıdaki algoritmalar kadar olmasa da, yine de öznitelik ölçekleme – veri normalizasyonu tavsiye edilmektedir [26].

Projede geliştirilen uygulamada, tüm algoritmalarda öznitelik standartlaştırma işlemleri uygulanmıştır. Veri normalizasyonu yöntemlerinden, Z-Skoru Normalizasyonu gerçekleştirilmiştir. Bu amaçla Phyton scikit-learn kütüphanesinde yer alan, veri önileme ve normalizasyon işlemleri altındaki StandartScaler() metodu kullanılmıştır.

4.2.2 Öznitelik Çıkarımı (Feature Extraction)

Öznitelik çıkarımında, 3.bölümde detaylarına yer verilen, Şekil 3.1’de gösterildiği şekilde, birden fazla veri tabanından verilerin aktarılması ile MS SQL veri tabanında oluşturulan, DBKACAK veri ambarındaki tablolar kullanılmıştır.

Çalışmada kullanılacak verilerde, farklı abone tipleri mutlaka dikkate alınmalıdır. Genellikle mesken ve endüstriyel aboneler üzerinde yapılan çalışmalarda, farklı davranış profilleri ortaya çıkmaktadır. Mesken sınıfındaki aboneler elektriği daha çok hafta sonları tüketirken, endüstriyel aboneler ise daha çok hafta içinde elektrik tüketmektedir (Glauner, vd, 2017). Benzer şekilde tarımsal sulamaların yapıldığı bölgelerde, özellikle tarımsal sulama grubundaki abonelerin tüketimlerinde, dönemsel olarak farklılıklar olmaktadır. Tez çalışmasında yer alan tüm abone grupları Tablo 4.3’te verilmiştir.

Tablo 4.3: KTS Abone Grupları

Abone Grupları		
Resmi daireler	Ticarethane	Hayır Kurumları
Kitler	Belediyeler	Aydınlatma
Cami – İbadethane	Üreticiler	Tedarikçiler
Hizmet Binaları	İçme Suyu	Mesken
Arıtma Tesisleri	Os/jeotermal	Tarımsal Sulama
Özel İdare Ayd.	Şantiye	Sanayi

Çalışmada kullanılacak abone sayısı da önemlidir. Veri temizliği sonrası bazı abonelerin elenerek, sayının düşeceği de göz önüne alınırsa, verilerde yer alan abone sayılarının yüksek olması modelin başarımına etki edecektir. Bu tezde, Dicle Elektrik Dağıtım Bölgesi’ndeki toplam 1.877.097 aboneden, otomatik sayaç okuma sistemine 2018 Ocak itibarıyla dahil olan toplam 275.809 adet abone verisi üzerinde çalışma gerçekleştirilmiştir. Benzer şekilde sahadaki kaçak tespitleri sonucu kaçak olarak belirlenen abone sayılarının da çok olması, modelin eğitiminde oldukça önemlidir. Bu tezde kullanılan verilerde, kaçak olduğu tespit edilen 37.281 abone bulunmaktadır. Yine de toplam aboneler içindeki kaçak abonelerinin sayısı, oransal olarak yetersiz kalmaktadır.

Abonelerin farklı bölgelerde ve farklı iklime sahip alanlarda yayılmış olması, çalışmaya olumlu katkı sağlayacaktır. Farklı bölgelerde iklime bağlı olarak, farklı elektrik tüketimleri gerçekleştirmektedir. Bu çalışmada Dicle Elektrik Dağıtım Bölgesi kapsamında yer alan Diyarbakır, Mardin, Şanlıurfa, Batman, Siirt ve Şırnak bölgelerindeki aboneler üzerinde çalışılmıştır. Her ne kadar genel olarak sıcak bir iklime sahip olan bölgede aboneler yayılmış olsa da, yine de lokasyon farklılıkları anlamında zengin bir veri bulunmaktadır. Tablo 4.4, bu tezde kullanılan illerdeki abone sayılarını göstermektedir.

Tablo 4.4: 2018 Yılı İllere Göre OSOS Kurulu Abone Sayıları

İl	OSOS Kurulu Abone Sayısı
Diyarbakır	75.264
Mardin	50.811
Şanlıurfa	62.002
Batman	26.651
Siirt	13.853
Şırnak	8.435

Uzun periyotta sayaç okuma verilerinin çalışmada yer alması önemlidir. En azından tüm mevsimlerde okuma verilerinin olması, mevsimsel tüketim değişikliklerinin ortaya konmasına imkân tanıyacaktır. Bu anlamda en azından bir yıllık veri üzerinde çalışılması önerilmektedir (Glauner, vd., 2017). Bu tezde kullanılan veri periyodu 2018 yılının tamamını içerdiğinden, çalışma yıl içerisindeki mevsimsel değişiklikleri kapsamaktadır.

Makine öğrenmesi algoritmalarında kullanılacak öznitelikler, farklı uygulama sistemlerindeki veri tabanlarından çıkarılmıştır. Tablo 4.5’de hangi uygulama sistemlerinden, hangi özniteliklerin alındığı bilgisi yer almaktadır.

Değişken veri tiplerini, Şekil 4.2’de olduğu gibi, “Kategorik” ve “Sayısal” olarak ikiye ayırabiliriz. Sayısal veriler, ölçülebilen ve sayarak elde edilebilen verilerdir. Vücut ağırlığı, elektrik tüketim değeri, basınç, sıcaklık değerleri gibi değişkenler, kategorik değişkenlere örnek olarak gösterilebilir.

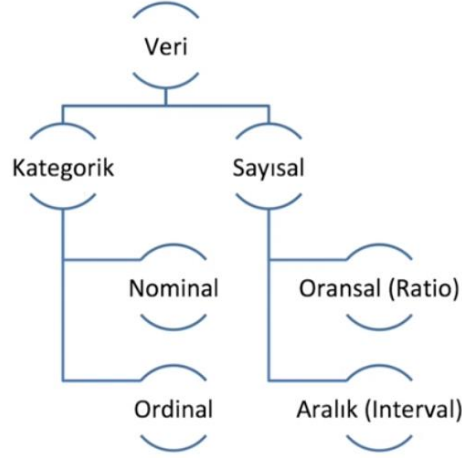
Tablo 4.5: Uygulama Sistemleri Öznitelik Çıkarımları

Uygulama Sistemi	Öznitelik – (Feature)
MBS	Abone Grupları
MBS	Ölçüm Durumları
CBS	İller
MBS	Sayaç Modelleri
MBS	Köy/Şehir
MBS	Kurulu Güç
MBS	Ortalama Tüketim
OSOS1, OSOS2	Günlük Ortalama Aylık Tüketim (Ay1-Ay12)
OSOS1, OSOS2	Aylık Max Demand (Ay1-Ay12)
OSOS1, OSOS2	Ortalama Akım1,2,3 (Ay1-Ay12)
OSOS1, OSOS2	Ortalama Gerilim1,2,3 (Ay1-Ay12)
OSOS1, OSOS2	CosFOrtalamaSapma 1,2,3 (Ay1-Ay12)

Kategorik veriler ise, değişkenlerin aldığı değerlerin belirli gruplar altında toplanabildiği, sözel ifade edilebilen veya sayısal bir değişken olarak sınıflandırılabilen verilerdir. Makine öğrenmesi algoritmalarında, kategorik verilerin sayısal verilere dönüştürülmesi gerekecektir. Bu tez çalışmasındaki tüm kategorik değişkenler, sayısal değişkenlere dönüştürülmüştür.

Kategorik verileri, “Nominal” ve “Ordinal” olarak ikiye ayırabiliriz. Ordinal veriler, sıralı verileri ifade etmektedir. Aralarında sıralama olarak bir ilişki de olsa, ölçülemeyen verilerdir. İl plaka kodları, bir sokaktaki evlerin kapı numaraları gibi değişkenler, ordinal verilere örnek olarak verilebilir. Nominal veriler, sıralı olmayan kategorik verilerdir. Cinsiyet, il, ilçe, abone grubu... gibi veriler, nominal verilerdir.

Sayısal verileri, “Oransal” ve “Aralıklı” olarak ikiye ayırabiliriz. Oransal verilere örnek olarak; her ay kaçak elektrik kullandığı tespit edilen abone sayısı oranı, abone başına aylık ortalama tüketim değeri, günlük maksimum demand değeri gibi veriler verilebilir. Adından da anlaşılacağı üzere, bu verilerde oransallık söz konusudur.



Şekil 4.2: İstatiksel Veri Tipleri

“Aralıklı” sayısal verileri ise, ölçümle belirlenen değerlerdir. Örneğin bir okuldaki öğretmen sayısı, kaçak kullanımı tespit edilen abone sayısı, abone elektrik tüketim değeri, sıcaklık değeri gibi değerler, “Aralıklı” olarak ifade edilen sayısal verilerdir.

Tez çalışmasında kullanılan veri tipleri, Tablo 4.6’da gösterilmiştir.

Tablo 4.6: KTS Değişkenleri ve Veri Tipleri

Veri Adı	Veri Tipi
Günlük Ortalama Aylık Tüketim (Ay1-Ay12)	Sayısal
Aylık Max Demand (Ay1-Ay12)	Sayısal
Ortalama Akım _{1,2,3} (Ay1-Ay12)	Sayısal
Ortalama Gerilim _{1,2,3} (Ay1-Ay12)	Sayısal
COS ϕ OrtalamaSapma _{1,2,3} (Ay1-Ay12)	Sayısal
Kaçak Bilgisi	Kategorik
Abone Grupları	Kategorik
Ölçüm Durumları	Kategorik
İller	Kategorik
Sayaç Modelleri	Kategorik
Köy/Şehir	Kategorik
Kurulu Güç	Sayısal
Ortalama Tüketim	Sayısal

4.2.2.1 Ay Bazlı Değişkenler

Ay bazlı değişkenler; yıl içerisindeki her ay için hesaplanan bir tane değişken olmak üzere, toplamda 12 adet değişkeni ifade etmektedir. Bu çalışmadaki aylık değişkenler, 2018 yılı Ocak-Aralık ayları için hesaplanan değerlerden oluşmaktadır.

“Günlük Ortalama Aylık Tüketim” değişkeni, her ay için o aydaki toplam tüketimin, tüketim gün sayısına bölünmesi ile elde edilen tüketim değeridir. Örneğin 2018 Ocak ayı için bir abonenin, Ocak ayındaki bir günde ortalama ne kadar elektrik tükettiğini ifade etmektedir.

“Aylık Max Demand” değişkeni, her ay için o aydaki en yüksek demand değerini vermektedir.

“Ortalama Akım1”, “Ortalama Akım2” ve “Ortalama Akım3” değişkenleri sırasıyla sayaçtan ölçülen Akım1, Akım2 ve Akım3 değerlerinden hesaplanmaktadır. OrtalamaAkım1, her ay için o aydaki Akım1 değerleri toplamının, akım değerinin alındığı gün sayısına bölünmesi ile elde edilen ortalama değerdir. O aydaki günlük ortalama Akım1 değerini vermektedir. “Ortalama Akım2” Akım2 değeri için, “Ortalama Akım3” de Akım3 değeri için benzer şekilde ilgili ay içerisindeki ortalama günlük ilgili akım değerini vermektedir.

“Ortalama Gerilim1”, “Ortalama Gerilim2” ve “Ortalama Gerilim3” değişkenler ise sırasıyla sayaçtan ölçülen Gerilim1, Gerilim2 ve Gerilim3 değerlerinden hesaplanmaktadır. OrtalamaGerilim1, her ay için o aydaki Gerilim1 değerleri toplamının, gerilim değerinin alındığı gün sayısına bölünmesi ile elde edilen ortalama değerdir. O aydaki günlük ortalama Gerilim1 değerini vermektedir. “Ortalama Gerilim2” Gerilim2 değeri için, “Ortalama Gerilim3” de Gerilim3 değeri için benzer şekilde ilgili ay içerisindeki ortalama günlük ilgili gerilim değerini vermektedir.

“ $\text{COS}\phi 1$ OrtalamaSapma”, “ $\text{COS}\phi 2$ OrtalamaSapma” ve “ $\text{COS}\phi 3$ OrtalamaSapma” değişkenleri de benzer şekilde sayaçtan ölçülen 3 farklı cosinus değerine ait ortalama sapma değerlerini içermektedir. Her bir değişken, 2018 yılı tüm ayları için hesaplanmaktadır ve Ocak-Aralık arası tüm aylara ait ortalama sapma verilerini içermektedir. Örneğin $\text{COS}\phi 1$ OrtalamaSapma, Ocak-Aralık arası 12 değişkenden oluşmakta olup, ilgili formüle göre hesaplanan her ay için sapma değerlerini içermektedir.

Ay bazlı deęişkenlerde her bir deęişken 12 farklı deęişken olarak ay bazlı hesaplanmaktadır.

X seçilen toplam öznitelik sayısını, M ise veri setindeki benzersiz abone sayısını ifade etmek üzere, tüm öznitelikler aşağıdaki formülle ifade edilebilir :

$$X = \{X^{(m)}, m = 1, \dots, M\} \quad (4.1)$$

M : Veri Setindeki Abone Sayısı

Çalışmada 2018 yılına ait veriler kullanıldığı için, $h = 1, \dots, 12$ deęerleri, 2018 yılına ait Ocak-Aralık arası ayları göstermektedir.

$h = 1, \dots, 12$ h : 12 aydan oluşan zaman aralığını ifade eder.

Her aya ait ortalama günlük tüketimler ise (4.2) deki formülle ifade edilmektedir :

$$X_i^{(m)} = P_h/G_h, i = 1, \dots, 12, h = 1, \dots, 12 \quad (4.2)$$

Öznitelikler içindeki ilk 12 deęer, ay bazında OrtalamaGünlükTüketimAy deęerleri ifade etmektedir.

OrtalamaGünlükTüketimAy : X_1, \dots, X_{12}

$$X_i^{(m)} = P_h/G_h, i = 1, \dots, 12, h = 1, \dots, 12 \quad (4.3)$$

P_h : h . aydaki toplam tüketim

G_h : h . aydaki tüketim deęeri alınan gün sayısı

Maksimum demand deęeri, ilgili aydaki en yüksek demand deęerini ifade etmektedir.

MaxDemandAy : X_{12}, \dots, X_{24}

$$X_i^{(m)} = D_h, i = 12, \dots, 24, h = 1, \dots, 12 \quad (4.4)$$

D_h : h . aydaki maksimum demand deęeri

OrtalamaGerilim1, OrtalamaGerilim2, OrtalamaGerilim3 deęerleri, sayaçtan okunan 3 adet gerilim deęerinin aylık ortalamasını ifade etmektedir.

OrtalamaGerilim1: X_{25}, \dots, X_{36}

$$X_i^{(m)} = \frac{V_{1h}}{NV_{1h}}, i = 25, \dots, 36, h = 1, \dots, 12 \quad (4.5)$$

V_{1h} : h . ayda okunan gerilim1 deęerleri toplamı

NV_{1h} : h . ayda okunan toplam gerilim1 sayısı

OrtalamaGerilim2: X_{37}, \dots, X_{48}

$$X_i^{(m)} = \frac{V_{2h}}{NV_{2h}}, i = 37, \dots, 48, h = 1, \dots, 12 \quad (4.6)$$

V_{2h} : h . ayda okunan gerilim2 deęerleri toplamı

NV_{2h} : h . ayda okunan toplam gerilim2 sayısı

OrtalamaGerilim3: X_{49}, \dots, X_{60}

$$X_i^{(m)} = \frac{V_{3h}}{NV_{3h}}, i = 49, \dots, 60, h = 1, \dots, 12 \quad (4.7)$$

V_{3h} : h . ayda okunan gerilim3 deęerleri toplamı

NV_{3h} : h . ayda okunan toplam gerilim3 sayısı

OrtalamaAkım1, OrtalamaAkım2, OrtalamaAkım3 deęerleri, sayaçtan okunan 3 adet akım deęerinin aylık ortalamasını ifade etmektedir.

OrtalamaAkım1: X_{61}, \dots, X_{72}

$$X_i^{(m)} = \frac{C_{1h}}{NC_{1h}}, i = 61, \dots, 72, h = 1, \dots, 12 \quad (4.8)$$

C_{1h} : h . ayda okunan akım1 deęerleri toplamı

NC_{1h} : h . ayda okunan toplam akım1 sayısı

OrtalamaAkım2: X_{73}, \dots, X_{84}

$$X_i^{(m)} = \frac{C_{2h}}{NC_{2h}}, i = 73, \dots, 84, h = 1, \dots, 12 \quad (4.9)$$

C_{2h} : h . ayda okunan akım2 değerleri toplamı

NC_{2h} : h . ayda okunan toplam akım2 sayısı

OrtalamaAkım3: X_{85}, \dots, X_{96}

$$X_i^{(m)} = \frac{C_{3h}}{NC_{3h}}, i = 85, \dots, 96, h = 1, \dots, 12 \quad (4.10)$$

C_{3h} : h . ayda okunan akım1 değerleri toplamı

NC_{3h} : h . ayda okunan toplam akım3 sayısı

$COS\varphi_1$ OrtalamaSapma, $COS\varphi_2$ OrtalamaSapma, $COS\varphi_3$ OrtalamaSapma değerleri, sayaçtan okunan $\cos \varphi$ değerlerinin belirlenen aralık dışında olanlarının, toplamının ortalamasını ifade etmektedir.

$COS\varphi_1$ OrtalamaSapma: X_{97}, \dots, X_{108}

$$C\varphi_1 = 0,8 - \cos \varphi_1 ; \cos \varphi_1 < 0,8$$

$$C\varphi_1 = 1 - \cos \varphi_1 ; \cos \varphi_1 > 1$$

$$C\varphi_1 = 0 ; 0,8 < \cos \varphi_1 < 1$$

$$X_i^{(m)} = \frac{C\varphi_{1h}}{NC\varphi_{1h}}, i = 97, \dots, 108, h = 1, \dots, 12 \quad (4.11)$$

$C\varphi_{1h}$: h . ayda $\cos \varphi_1$ sapma farkları toplamı

$NC\varphi_{1h}$: h . ayda sapma gösteren toplam $\cos \varphi_1$ sayısı

$COS\varphi_2$ OrtalamaSapma: X_{109}, \dots, X_{120}

$$C\varphi_2 = 0,8 - \cos \varphi_2 ; \cos \varphi_2 < 0,8$$

$$C\varphi_2 = 1 - \cos \varphi_2 ; \cos \varphi_2 > 1$$

$$C\varphi_2 = 0 ; 0,8 < \cos \varphi_2 < 1$$

$$X_i^{(m)} = \frac{C\varphi_{2h}}{NC\varphi_{2h}}, i = 109, \dots, 120, h = 1, \dots, 12 \quad (4.12)$$

$C\varphi_{2h}$: h . ayda $\cos \varphi_2$ sapma farkları toplamı

$NC\varphi_{2h}$: h . ayda sapma gösteren toplam $\cos \varphi_2$ sayısı

$\text{COS}\varphi_3$ Ortalama Sapma: X_{121}, \dots, X_{132}

$C\varphi_3 = 0,8 - \text{Cos}\varphi_3$; $\text{Cos}\varphi_3 < 0,8$

$C\varphi_3 = 1 - \text{Cos}\varphi_3$; $\text{Cos}\varphi_3 > 1$

$C\varphi_3 = 0$; $0,8 < \text{Cos}\varphi_3 < 1$

$$X_i^{(m)} = \frac{C\varphi_{3h}}{NC\varphi_{3h}}, i = 121, \dots, 132, h = 1, \dots, 12 \quad (4.13)$$

$C\varphi_{3h}$: h . ayda $\text{Cos}\varphi_3$ sapma farkları toplamı

$NC\varphi_{3h}$: h . ayda sapma gösteren toplam $\text{Cos}\varphi_3$ sayısı

4.2.2.2 Master Değişkenler

Master değişkenler abonelere ait ay veya yıl bazlı değişmeyen ana verileri içeren değişkenlerdir. Ay bazlı değişkenlere ait değerler OSOS'tan alınırken, master değişkenlere ait değerler, OSOS dışında kalan MBS ve CBS gibi sistemlerden elde edilmektedir. Bu değişkenler aşağıdaki gibidir:

X_{133} : Abone Grupları

X_{134} : AGOG Durumu (AG/OG)

X_{135} : Ölçüm Durumları (Primer, Seconder, Diğer)

X_{136} : Tarife

X_{137} : Sektör

X_{138} : İl

X_{139} : İlçe

X_{140} : Köy Durumu

X_{141} : Sayaç Modeli

X_{142} : Bağlantı Durumu

X_{143} : Kurulu Güç

X_{144} : Ortalama Tüketim

Abone grupları, abonelerin bağlı olduğu mesken, ticarethane gibi sınıfları ifade etmektedir. AG/OG durumu, abonenin alçak gerilim veya orta gerilim olarak

kullandığı elektriğin gerilim durumunu belirtmektedir. Ölçüm durumu, abonenin elektriğinin primer/secondar olarak ölçüm durumunu; tarife ise faturalamaya baz teşkil eden elektrik kullanım tarifelerini kapsamaktadır. Sektör bilgisi, abonenin hangi sektörde faaliyet gösterdiğini, İl ve İlçe bilgileri de lokasyon olarak abonenin bulunduğu lokasyonu ifade eder. Köy durumu, abonenin köyde yaşayıp yaşamadığını belirtir. Kırsal abonelerin davranış şekillerinde farklılıklar olabilmektedir. Sayaç modeli, abonenin kullandığı elektrik sayacının modelini gösterir. Bağlantı durumu abonenin elektrik hatlarına bağlantı şeklini ifade ederken, kurulu güç ise sözleşmede yer alan watt cinsinden kullandığı güçtür. Ortalama tüketim ise her bir abone için yürüyen ağırlıklı ortalama olarak ilerleyerek, her tüketim faturalaması sonrası güncellenen bir değer olup, abonelik başından beri abonenin günlük tüketimini göstermektedir.

4.2.3 Sınıflandırma ve Uygulama Geliştirme

Çeşitli veri kaynaklarından toplandıktan sonra veri aktarımı ile aktarılan; veri temizliği ve veri ön işleme işlemleri ile kullanılabilir hale getirilen veriler kullanılarak, uygun makine öğrenmesi sınıflandırma algoritmaları ile modeller oluşturulmuştur.

Makine öğrenmesi kapsamında tezde kullanılan sınıflandırma algoritmaları aşağıdaki gibidir.

- Lojistik Regresyon (Logistic Regression)
- Karar Ağaçları (Decision Tree)
- Rassal Orman (Random Forest)
- Destek Vektör Sınıflandırma (SVC-Support Vector Classification)
- K-En Yakın Komşu (KNN)
- XGBoost – Extreme Gradient Boosting

Proje uygulaması, Windows 10 Pro işletim sistemi üzerinde, Anaconda platformu, Spyder 5.1.5 Grafiksiz Kullanıcı Arayüzü (GUI) ile Python 3.9 dili kullanılarak geliştirilmiştir. Oracle 12c, Microsoft SQL Server 2017 ve SQLExpress veri tabanları kullanılmıştır. Geliştirilen uygulamanın akış şeması, Şekil 4.3'te gösterilmiştir.

Phyton'da kullanılan ana kütüphaneler aşağıdaki gibidir:

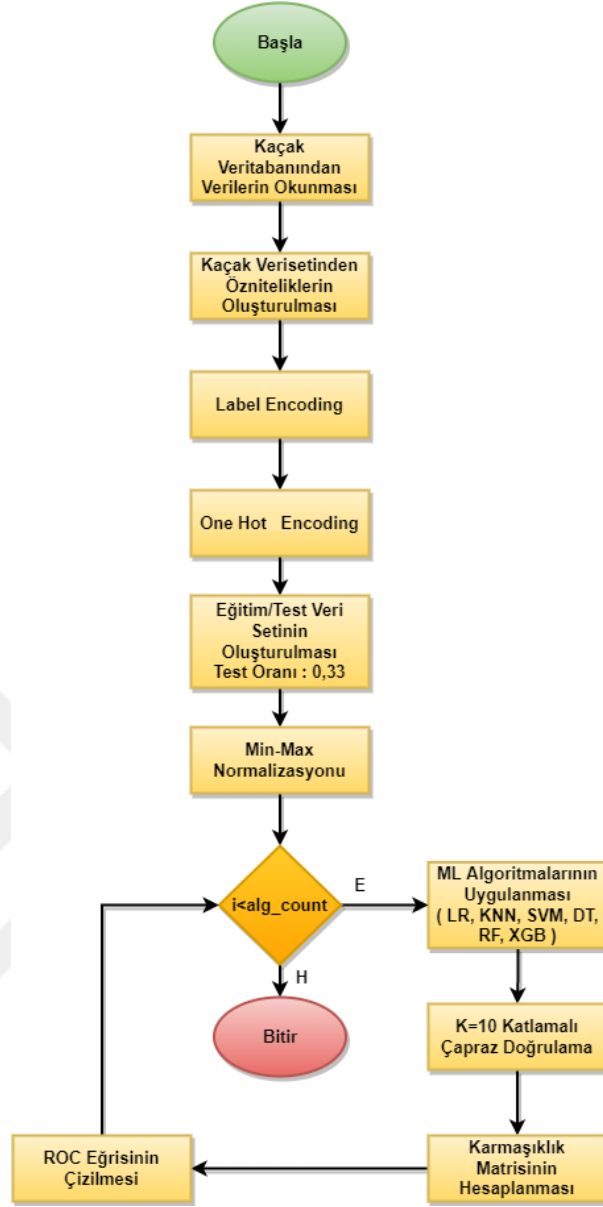
- Pyodbc
- Pandas
- Numpy
- Matplotlib
- sklearn
- statsmodels.api
- xgboost
- seaborn

Geliştirilen uygulamadaki veriler, veri ambarının oluşturulduğu Microsoft SQL Server veritabanından, Python kodu içinden SQL Server bağlantı sürücülerini (driver) ve MSSQL script dili kullanılarak okunmuştur. Kaçak tahmin verileri içindeki "Tesisat No" alanı anahtar alan olarak kullanılmış olup, 2018 yılında sözleşmesi devam eden tekil bir aboneyi ifade etmektedir.

Abone grubu, ölçüm durumu, il alanları için kategorik-nümerik dönüşümü için sırasıyla Label-Encoding ve One-Hot Encoding işlemleri uygulanmıştır. Farklı veri setlerinde gerçekleştirilen işlemler sonrası oluşturulan öznitelikler, KaçakTahmin veri setinde toplanmıştır. Oluşturulan KaçakSonuc veri seti ise, abonelere ait 0/1 değerleri ile normal/kaçak bilgisini içermektedir.

Kaçak Tahmin Verileri, 0.33 oranında eğitim ve test verisine bölünmüştür. Tüm verinin %66 lık bölümünde eğitim veri kümesinde makine öğrenmesi eğitimi gerçekleştirilirken, %33 lük bölümünde test veri kümesinde tahminleme gerçekleştirilmiştir.

KaçakTahmin veri setinde, min-max normalizasyonu gerçekleştirilmiştir. Normalizasyon sonucunda elde edilen veriler, belirlenen makine öğrenmesi algoritmalarında kaçak tahmini için kullanılmıştır. KaçakSonuc veri setindeki kaçak sonuçları kullanılarak, gözetimli öğrenme (Supervised Learning) gerçekleştirilmiştir.



Şekil 4.3: KTS Uygulama Akış Şeması

K=10 katlamalı çapraz doğrulama kullanılarak, optimum sonuçların elde edilmesi amaçlanmıştır. İş akışında doğrulama adımında yer alan karmaşıklık matrisi sonuçları ve model iyileştirme çalışmalarına, 5. Bölüm’de yer verilmiştir.

Geliştirilen projenin pseudo kodu aşağıdaki gibidir:

X_i ($i=1, 2, \dots, m$)

- o $A_i = (LR, KNN, SVM, DT, RF, GBC, ABC, HGBC)$

$D_0 =$ KACAK Veriseti

KacakDB = MSSQL DB den Veri Getir()

KacakDB den Veriyi D_0 Al.

$F_t = D_0$ Öznitelik Seti

$L_i =$ Kategorik X_i Verisi için Label Encoding Verisi

$O_i = L_i$ için One Hot Encoding

KacakVeri = Dataset(O_i, F_t , nümerik X_i)

KacakEğitimVerisi, KacakTestVerisi = KacakVerisi Ayır, test_boyutu = 0.33

Eğitim/Test Verisi min_max normalizasyonu yap.

ML Algoritma Sayısı = Kullanılacak Toplam Algoritma Sayısı.

eğer ($i < \text{MLAlgoritma_Sayısı}$)

A_i ML Algoritmasını seç

KacakEğitimVerisi için $k=10$ Çapraz Katlamalı Doğrulama gerçekleştir.

KacakTestVerisi test et.

Karmaşıklık Matrisi hesapla.

ROC/Kesinlik Duyarlılık eğrilerini çiz.

Doğrulama ve en iyi çözümün bulunması

4.3 Değerlendirme

Bu bölümde abone kaçak tahmini için, Dicle Elektrik Dağıtım bölgesinde uygulanan mevcut uzman yaklaşımı ortaya konarak, makine öğrenmesi ile kaçak tahmini için yeni bir model önerilmiştir. Mevcut yaklaşımdaki saha çalışmalarının tecrübesi ışığında ortaya konan algoritma detaylarına yer verilmiştir. Önerilen makine öğrenmesi önerilen yaklaşımı ile, var olan daha fazla verinin kaçak tahmininde dikkate alınması ve öğrenen bir yapı ortaya konması amaçlanmıştır.

Mevcut yaklaşımda, sayaçtan alınan akım, gerilim ve cosinus değerleri ile birlikte sayaç değişimi sonrası abonelerin tüketimleri, uzaktan okunan sayaçların kapaklarının bilgi dışında açılması, tüketimdeki dalgalanmalar dikkate alınmıştır. Makine öğrenmesi yaklaşımında ise toplamda 144 öznitelik çıkarılmıştır. Büyük veri üzerinde oluşturulan modelde, eğitim/test çalışmaları gözetimli öğrenme kapsamında gerçekleştirilmiştir.

Yapılan çalışmada; Lojistik Regresyon, Karar Ağaçları, Rassal Orman, Destek Vektör Sınıflandırma, K-En Yakın Komşu makine öğrenmesi algoritmalarının yanı sıra Boosting algoritmaları da kullanılmıştır. Oluşturulan modelin değerlendirilmesi ve

performans iyileştirme çalışmalarına 5. Bölüm’de yer verilmiştir.



BEŞİNCİ BÖLÜM

KTS: MODELİN DEĞERLENDİRİLMESİ

Veri setinde kaçak olan abonelerin sayısı ile kaçak olmayan abonelerin sayısı dengesiz dağılmıştır. Dengelenmemiş veri seti VeriSeti0 ile oluşturulan model üzerinden elde edilen sonuçlar, modelin aşırı öğrenme gerçekleştirdiğini göstermektedir.

Model karmaşıklık matrisleri sonuçları göz önüne alınarak, veri setinin dengeli dağılım olacak şekilde yeniden düzenlenmesi gerektiği belirlenmiştir. Bu amaçla daha dengeli dağılım sahip birden fazla veri seti oluşturularak, oluşturulan bu veri setleri üzerinden model tekrar test edilmiştir.

Tablo 5.1 yeni oluşturulan veri setlerindeki kayıt sayılarını ve kaçak/kaçak olmayan veri dağılımlarını göstermektedir.

Tablo 5.1: KTS Modeli Veri Setleri ve Kaçak Dağılımları

Veri Seti	Normal Sayısı	Kaçak Sayısı	Toplam	Kaçak Sayısı Oranı
VeriSeti0	114.620	5.620	120.240	%4.67
VeriSeti1	5.000	5.620	10.620	%52.92
VeriSeti2	7.655	5.620	13.275	%42.34
VeriSeti3	10.310	5.620	15.930	%35.28

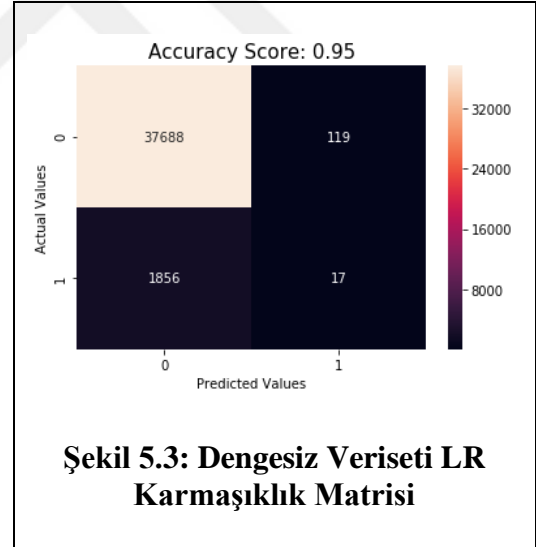
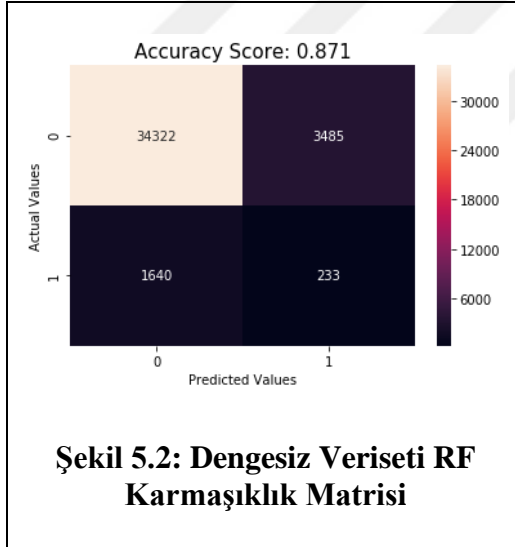
Veri Seti1’de, kaçak olan abonelerin sayısı ile kaçak olmayan abonelerin sayısı dengeli olacak şekilde hazırlanmıştır. 5620 kaçak abone, 5000 normal abone olmak üzere 10.620 kayıttan oluşan veri setinde kaçak olan abonelerin toplam abone içindeki oranı %52,92’dir. VeriSeti2 ve VeriSeti3, kaçak abone verileri aynı kalıp, normal abone sayıları artırılarak oluşturulan yeni veri setleridir. Model oluşturulurken tüm veri setleri, %66’sı eğitim %33’ü test verisi olacak şekilde ayrılmıştır. Şekil 5.1, Dengesiz ve dengeli verisetlerinde normal/kaçak dağılımlarını göstermektedir.



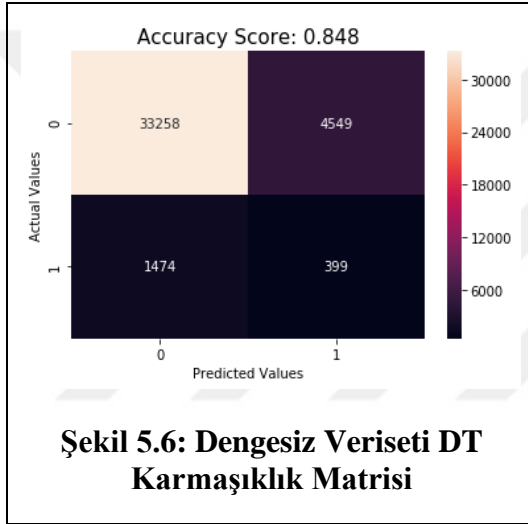
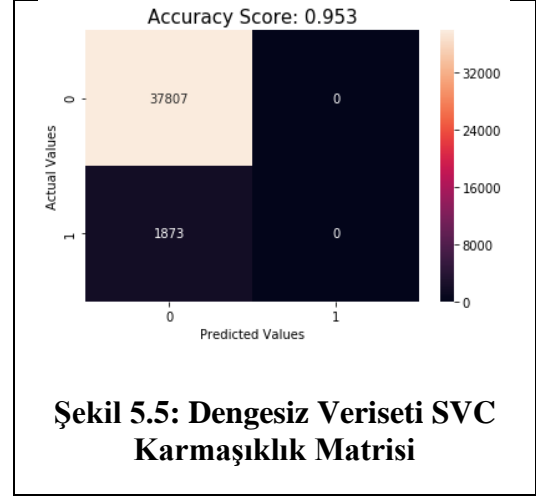
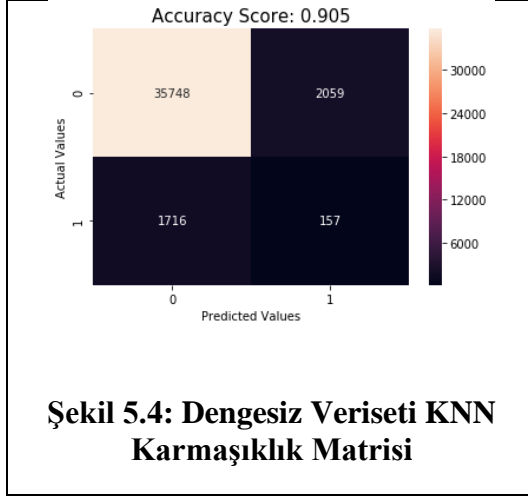
Şekil 5.1: Dengeli ve Dengesiz Veri Setleri

5.1 Karmaşıklık Matrisleri

Dengelenmemiş veri seti VeriSeti0'a ait Rastgele Orman, Lojistik Regresyon, KNN, SVC, Karar Ağaçları algoritmaları için hesaplanan karmaşıklık matrisleri sırasıyla Şekil 5.2, Şekil 5.3, Şekil 5.4, Şekil 5.5, Şekil 5.6'da gösterilmiştir.



VeriSeti0 ile oluşturulan modellerde, karmaşıklık matrisi incelendiğinde, doğruluk değerleri yüksek çıkmaktadır. Kaçak abonelerin toplam aboneler içindeki doğru tespiti, projemiz için önemli olduğundan, duyarlılık değerini de dikkate almak gereklidir. VeriSeti0 için doğruluk ve duyarlılık değerleri karşılaştırıldığında, sonuçlar Tablo 5.2'deki gibi çıkmaktadır.



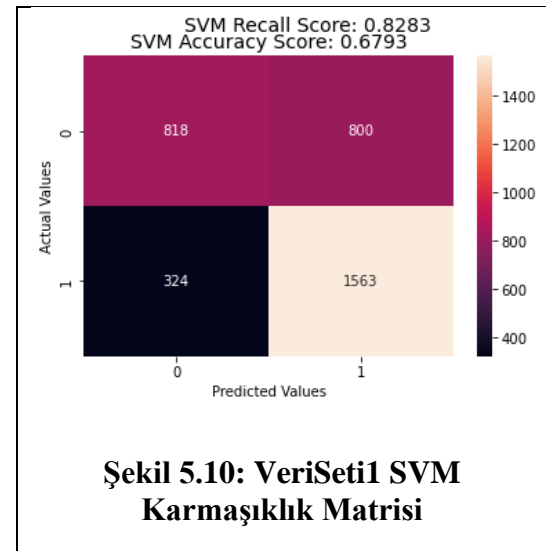
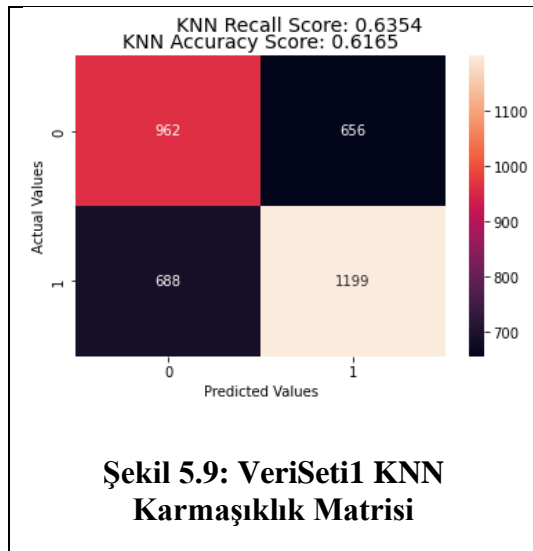
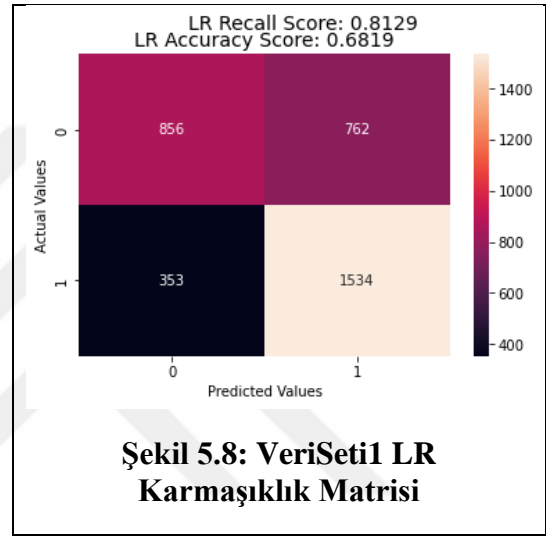
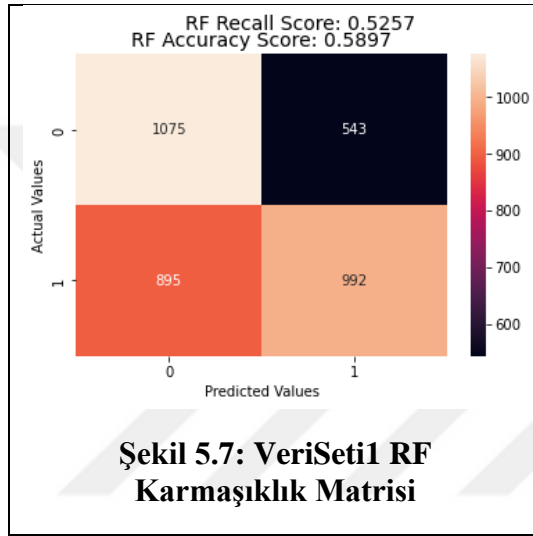
Tablo 5.2: VeriSeti0 Doğruluk-Duyarlılık Değerleri

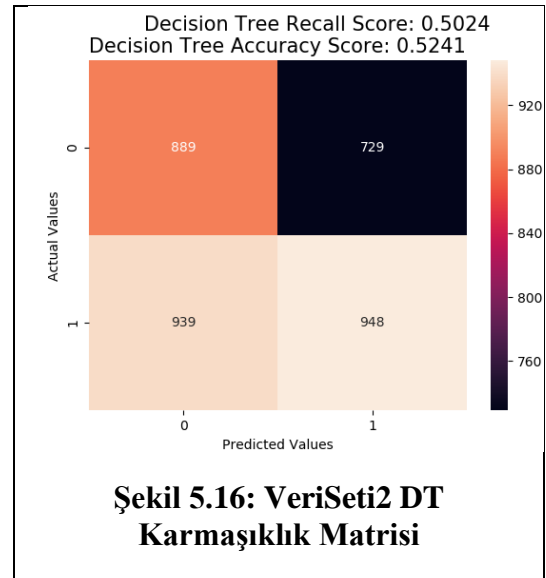
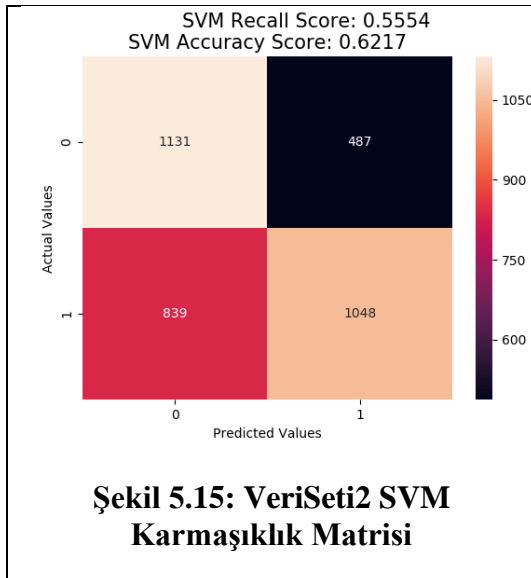
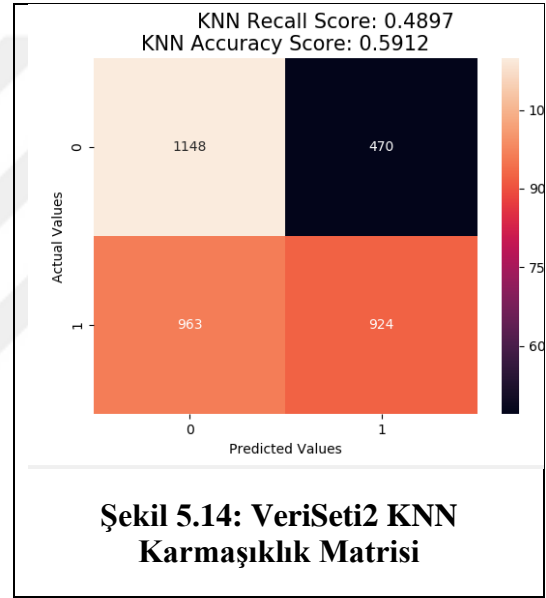
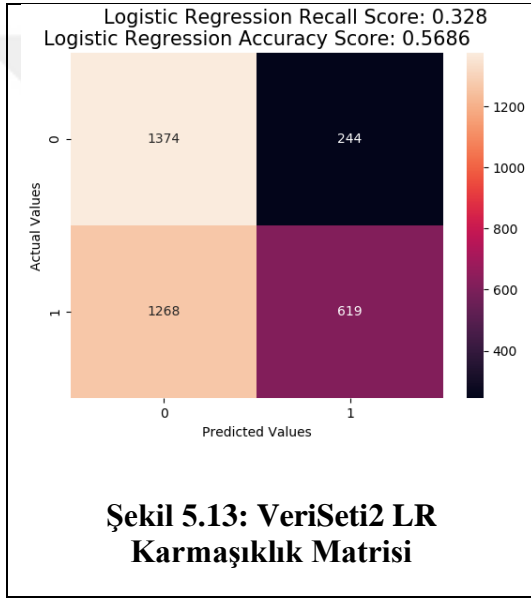
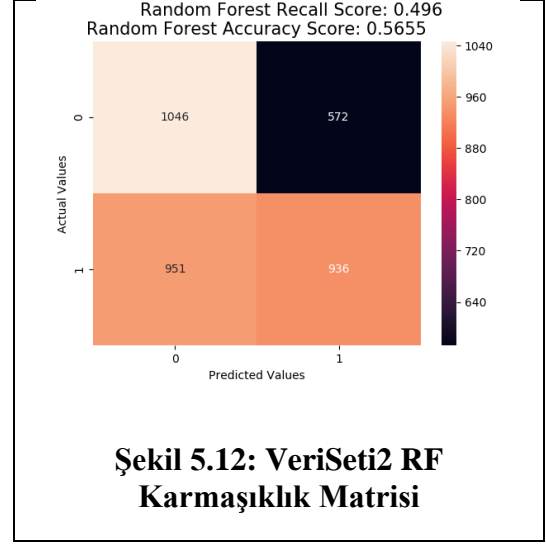
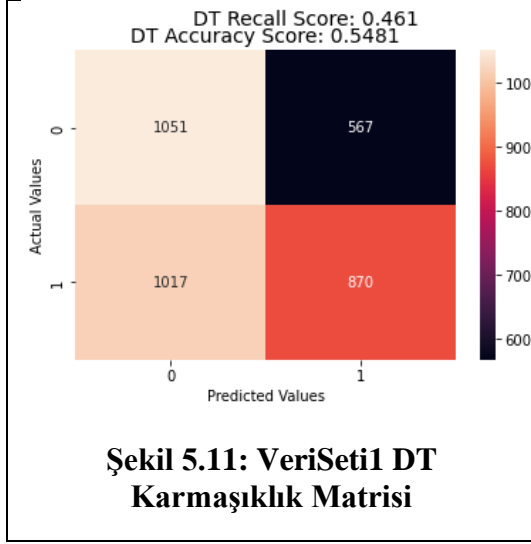
Algoritma	Doğruluk	Duyarlılık
Random Forest	0.871	0.124
Lojistik Regresyon	0.950	0.009
KNN	0.905	0.084
SVC	0.953	0
Decision Tree	0.848	0.213

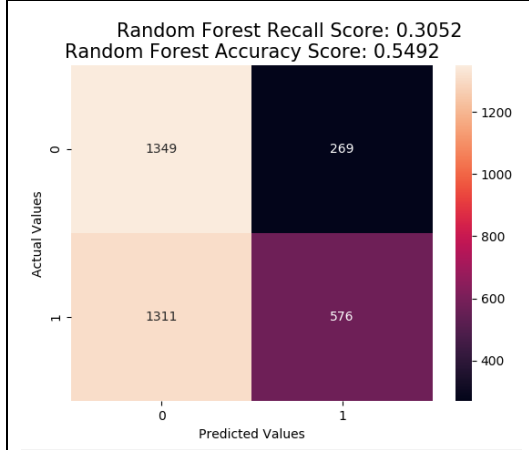
Tablo 5.2'deki sonuçlar incelendiğinde, doğruluk değerleri yüksek çıkarken, duyarlılık değerlerinin düşük çıktığı görülmektedir. Bu sonuç modelimizin aşırı öğrenme

(ezberleme) geliřtirdiđini gstermektedir. Aynı zamanda veri setinin dengeli olmadıđını ortaya koymaktadır. Dengeli bir veri seti elde etmek iin, modelde kullanılan veri setindeki rnekle rin, normal/kaak deđerlerinin dengeli dađılımlarını sađlayacak Tablo 5.1'deki gibi yeni veri setleri oluřturulmuřtur.

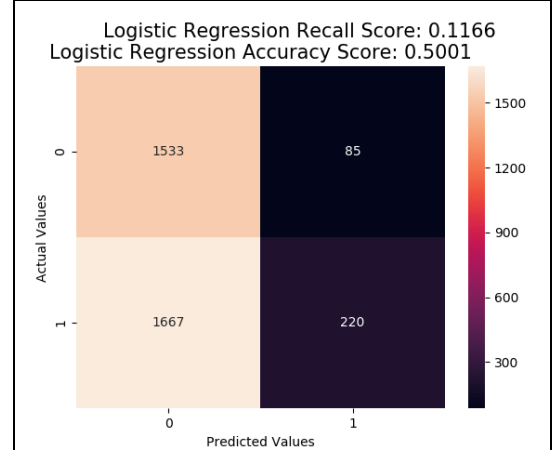
Kaak ve normal abonelerin dengeli olarak dađıtıldıđı veri setleri; VeriSeti1, VeriSeti2 ve VeriSeti3 iin elde edilen karmařıklık matrisleri, algoritma bazında Őekil 5.7, Őekil 5.8, Őekil 5.9, Őekil 5.10, Őekil 5.11'de verilmiřtir.



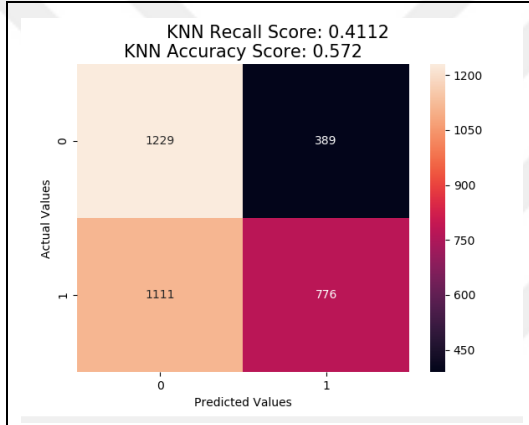




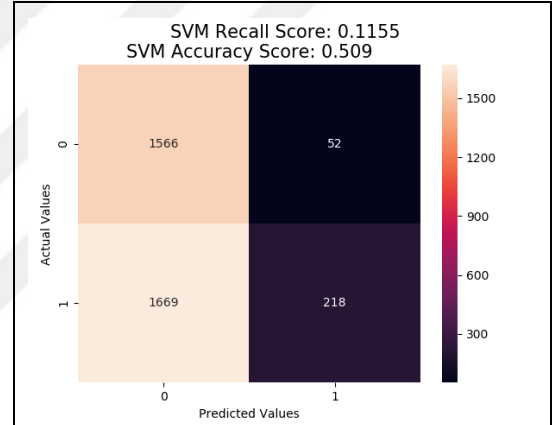
Şekil 5.17: VeriSeti3 RF Karmaşıklık Matrisi



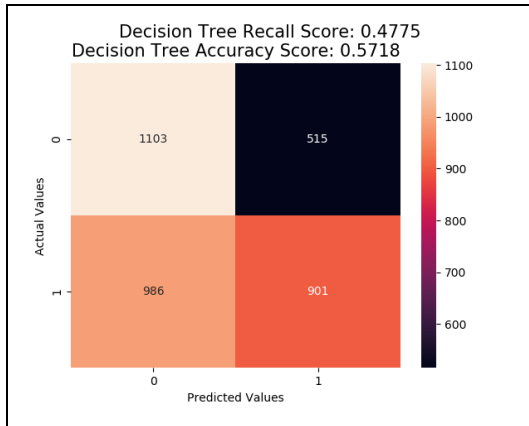
Şekil 5.18: VeriSeti3 LR Karmaşıklık Matrisi



Şekil 5.19: VeriSeti3 KNN Karmaşıklık Matrisi



Şekil 5.20: VeriSeti3 SVM Karmaşıklık Matrisi



Şekil 5.21: VeriSeti3 DT Karmaşıklık Matrisi

VeriSeti1, VeriSeti2, VeriSeti3 için doğruluk ve duyarlılık değerleri, Tablo 5.3, Tablo 5.4 ve Tablo 5.5’de verilmiştir.

Tablo 5.3: VeriSeti1 Doğruluk-Duyarlılık Değerleri

Algoritma	Doğruluk	Duyarlılık
RF	0.5897	0.5257
LR	0.6819	0.8129
KNN	0.6165	0.6354
SVM	0.6793	0.8283
DT	0.5481	0.4610

Tablo 5.4: VeriSeti2 Doğruluk-Duyarlılık Değerleri

Algoritma	Doğruluk	Duyarlılık
RF	0.5655	0.496
LR	0.5686	0.328
KNN	0.5912	0.4897
SVM	0.6217	0.5554
DT	0.5241	0.5024

Tablo 5.5: VeriSeti3 Doğruluk-Duyarlılık Değerleri

Algoritma	Doğruluk	Duyarlılık
RF	0.5492	0.3052
LR	0.5001	0.1166
KNN	0.572	0.4112
SVM	0.509	0.1155
DT	0.5718	0.4775

Veri setlerindeki kaçak/normal dağılımları değiştikçe, algoritmaların performanslarının da farklı yönde değiştiği görülmektedir. VeriSeti0 üzerinde çalışıldığında yüksek çıkan doğruluk değerleri, daha dengeli veri setlerinde azalırken; tersine dengesiz veri seti VeriSeti0’da düşük olan duyarlılık değerlerinin, dengeli veri

setleri ile birlikte yükseldiği görülmektedir. Normal sayısının arttığı VeriSeti2 ve VeriSeti3 sonuçlarına bakıldığında, VeriSeti1 de elde edilen performansın hem doğruluk hem de duyarlılık değerleri anlamında giderek düştüğü görülmektedir. Bu sonuçlara göre, veri setleri içinde optimum çözümü sağlayan veri seti VeriSeti1 olduğundan, modelde kullanılmak üzere bu veri seti seçilmiştir. Modelde kullanılmak üzere seçilen VeriSeti1 karmaşıklık matrisi, Tablo 5.6'da gösterilmiştir.

Tablo 5.6: KTS Karmaşıklık Matrisi

Sınıflandırma	Doğruluk	Duyarlılık	Kesinlik	F1 Skoru
LR	68.19	81.29	66.81	73.34
KNN	61.65	63.54	64.64	64.08
SVM	67.93	82.83	66.14	73.55
DT	54.81	46.10	60.54	52.35
RF	58.97	52.57	64.63	57.98

Bu tablo incelendiğinde, LR algoritmasının kullanıldığı model %68,19 ile en yüksek doğruluk değerini elde ederken, SVM modeli %82,83 ile en yüksek duyarlılık değerine sahiptir. DT, tüm değerlerde en kötü performansı göstermiştir. Bu sonuçlarla SVM ve LR toplamda en iyi performansı gösteren algoritmalar olmuştur.

5.2 K-Kat Çapraz Doğrulama

Sınıflandırma modellerinin değerlendirilmesi ve modelin eğitilmesinde, K Katlamalı Çapraz Doğrulama (K Fold Cross Validation) tekniği kullanılmıştır. K değeri k=5, k=10, k=15 olarak birden fazla değer ile test edilerek, sonuçlar değerlendirilmiştir. Kaçak tahmin veri kümesi; 5,10 ve 15 katlı gruba ayrılarak, oluşturulan modeller bu gruplar için test edilmiştir. k=5 katlamalı çapraz doğrulama için, her bir katta elde edilen karmaşıklık matrisi değerleri ve bu değerlerin ortalamaları, Tablo 5.7'de verilmiştir.

Tablo 5.7: k=5 Katlamalı Değerlendirme Kat Detayları

	Algoritma	Kat 1	Kat 2	Kat 3	Kat 4	Kat 5	Ortalama
LR	Doğruluk	0.66127899	0.6718201	0.67111736	0.67814476	0.66619817	0.6697
	Duyarlılık	0.80831099	0.80965147	0.82329317	0.82195448	0.81793842	0.8162
	Kesinlik	0.68092679	0.6707476	0.70427826	0.70503448	0.68692958	0.6896
	F1 Skoru	0.71445498	0.72119403	0.72438163	0.72835113	0.72009428	0.7217
KNN	Doğruluk	0.62473647	0.60435699	0.62473647	0.62895292	0.62543921	0.6216
	Duyarlılık	0.60991957	0.59517426	0.64390897	0.62248996	0.62115127	0.6185
	Kesinlik	0.60208121	0.58706084	0.60044066	0.60528656	0.60253695	0.5995
	F1 Skoru	0.63019391	0.61199173	0.64304813	0.63786008	0.63518138	0.6317
RF	Doğruluk	0.6528461	0.66127899	0.67252284	0.69501054	0.6753338	0.6714
	Duyarlılık	0.63002681	0.65549598	0.67068273	0.68406961	0.68808568	0.6657
	Kesinlik	0.69897271	0.70247811	0.73326485	0.7342128	0.71521185	0.7168
	F1 Skoru	0.65550907	0.66986301	0.68256131	0.70192308	0.68993289	0.6800
DT	Doğruluk	0.61419536	0.6373858	0.64230499	0.6303584	0.62895292	0.6306
	Duyarlılık	0.63270777	0.6233244	0.67068273	0.6398929	0.65997323	0.6453
	Kesinlik	0.59262467	0.61172399	0.61268056	0.60518528	0.60270318	0.6050
	F1 Skoru	0.63228399	0.64315353	0.663137	0.64507422	0.65125495	0.6470
SVM	Doğruluk	0.66127899	0.66338721	0.66479269	0.66549543	0.66408995	0.6638
	Duyarlılık	0.81367292	0.81635389	0.83400268	0.79518072	0.82864793	0.8176
	Kesinlik	0.69552015	0.66078105	0.69464535	0.70395723	0.69499576	0.6900
	F1 Skoru	0.71580189	0.71773718	0.72315728	0.71394231	0.72144522	0.7184
XGBoost	Doğruluk	0.69711876	0.70765987	0.71820098	0.69922699	0.69360506	0.7032
	Duyarlılık	0.72252011	0.71447721	0.75368139	0.72958501	0.72289157	0.7286
	Kesinlik	0.7711929	0.77634392	0.78403905	0.78418801	0.77772647	0.7787
	F1 Skoru	0.71438038	0.71929825	0.73739358	0.71805007	0.71240106	0.7203

k=5 kat katlamalı çapraz doğrulama için, veriseti 5 parçaya bölünerek, her bir parça için karmaşıklık matrisi değerleri hesaplanmıştır. Tablo 5.6'da algoritma bazında her bir kat için elde edilen doğruluk, duyarlılık, kesinlik ve F1 skoru değerleri verilmiş

olup, bu deęerlerin ortalaması da ortalama stununda gsterilmiřtir. Aynı řekilde k=10 ve k=15 kat katlamalı apraz doęrulama detaylı verileri, Ek-1 de verilmiřtir.

Her bir k-kat apraz doęrulama iin, ortalama deęerleri ieren karmařıklık matrisleri sonuları; k=5, k=10 ve k=15 iin sırasıyla Tablo 5.8, Tablo 5.9 ve Tablo 5.10'da gsterilmiřtir.

Tablo 5.8: k=5 Katlamalı Karmařıklık Matrisi

Sınıflandırma	Doęruluk	Duyarlılık	Kesinlik	F1 Skoru
LR	0.6697	0.8162	0.6896	0.7217
KNN	0.6216	0.6185	0.5995	0.6317
SVM	0.6638	0.8176	0.6900	0.7184
DT	0.6306	0.6453	0.6050	0.6470
RF	0.6714	0.6657	0.7168	0.6800
XGB	0.7032	0.7286	0.7787	0.7203

Tablo 5.9: k=10 Katlamalı Karmařıklık Matrisi

Sınıflandırma	Doęruluk	Duyarlılık	Kesinlik	F1 Skoru
LR	0.6711	0.8170	0.6916	0.7228
KNN	0.6252	0.6223	0.6021	0.6351
SVM	0.6670	0.8307	0.6897	0.7237
DT	0.6354	0.6400	0.6094	0.6481
RF	0.6812	0.6734	0.7348	0.6887
XGB	0.7103	0.7351	0.7868	0.7269

Tablo 5.10: k=15 Katlamalı Karmařıklık Matrisi

Sınıflandırma	Doęruluk	Duyarlılık	Kesinlik	F1 Skoru
LR	0.6713	0.8173	0.6936	0.7230
KNN	0.6264	0.6196	0.6035	0.6350
SVM	0.6668	0.8288	0.6927	0.7231
DT	0.6344	0.6445	0.6085	0.6489
RF	0.6820	0.6670	0.7366	0.6875
XGB	0.7155	0.7356	0.7905	0.7307

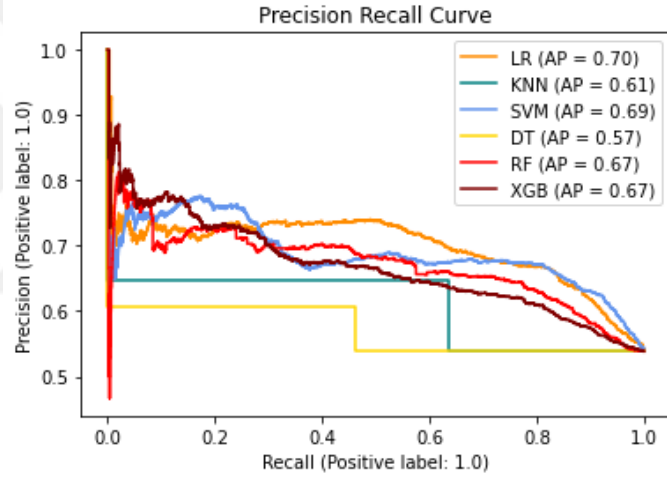
5.3 Precision-Recall ve ROC Eęrileri

K-Kat apraz Doęrulama (K Fold Cross Validation) ynteminde k=10 olarak seildięinde, modelde kullanılan algoritmalara ait standart sapma ve AUC deęerleri

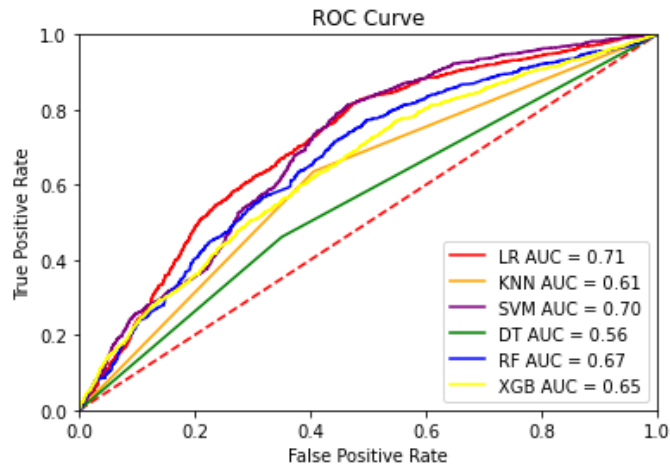
Tablo 5.11’de verilmiştir. Oluşturulan modeldeki sınıflandırma algoritmalarına ait Precision Recall ve AUC eğrileri, Şekil 5.22 ve Şekil 5.23’te verilmiştir.

Tablo 5.11: Model Standart Sapma/AUC Değerleri

Sınıflandırma	Std (%)	AUC
LR	1.64	0.71
KNN	1.24	0.61
SVM	1.66	0.70
DT	1.79	0.56
RF	1.54	0.67
XGB	0.93	0.54



Şekil 5.22: Precision Recall Eğrisi



Şekil 5.23: ROC Eğrisi

5.4 Model İyileştirme Alternatif Yöntemler

5.4.1 Yinelemeli Öznitelik Elemesi ile Öznitelik Seçimi

Yinelemeli Öznitelik Elemesi (RFE-Recursive Feature Elimination) geriye doğru özniteliklerin seçilmesi yöntemidir. Bu yöntemde sınıflandırma performansını optimize eden öznitelik alt kümesini bulmak için, tüm öznitelikleri derecelendirerek, amaç fonksiyonu doğrultusunda en düşük skora sahip öznitelikleri elemektedir. (Eskidere, 2012).

Bu çalışmada, LR-RFE Lojistik Regresyon Yinelemeli Öznitelik Elemesi ile öznitelik alt kümesi oluşturulmuştur. Toplam 144 öznitelik içinden, 10 ile 50 arası öznitelik seçilerek; seçilen özniteliklerle model tekrar çalıştırılmış ve çıkan karmaşıklık matrisi sonuçları değerlendirilmiştir. $K=10$ çapraz katlamalı doğrulama işlemi gerçekleştirilmiştir. Tablo 5.12, LR-RFE ile seçilen öznitelikleri kullanarak, Lojistik regresyon ve VeriSet1 ile oluşturulan modele göre, karmaşıklık matrisi sonuçlarını göstermektedir.

Tablo 5.12: LR-RFE Öznitelik Seçimleri Değerlendirme

Parametre Sayısı	Doğruluk	Duyarlılık	Kesinlik	F1 Skoru
144	0.6810	0.8161	0.6664	0.7337
10	0.5448	0.9341	0.6013	0.6828
20	0.6170	0.8160	0.6909	0.6909
30	0.6687	0.8307	0.6970	0.7247
40	0.6697	0.8278	0.6968	0.7246
50	0.6722	0.8165	0.6941	0.7234
60	0.6725	0.8219	0.6930	0.7248
70	0.6738	0.8216	0.6963	0.7255
80	0.6738	0.8229	0.6949	0.7259
90	0.6731	0.8208	0.6944	0.7249
100	0.6717	0.8187	0.6946	0.7235
110	0.6721	0.8189	0.6947	0.7239
120	0.6715	0.8178	0.6942	0.7233

Bu tabloya göre, LR-RFE ile 80 parametre seçildiğinde doğruluk değerinin 0,6738 iken duyarlılık değerinin 0,8229; 10 parametre seçildiğinde ise doğruluk değeri 0,5448

iken duyarlılık değerinin 0,9341 olarak elde edildiği görünmektedir. Seçilen 70 ve 80 öznitelik ile en yüksek doğruluk değeri elde edilirken; 10 parametre kullanıldığında en yüksek duyarlılık değerine ulaşılmıştır. Duyarlılık değerini kesinlik değeri ile birlikte değerlendirdiğimizde, 30 parametre ile yapılan çalışma; 0,6970 değeri ile en yüksek kesinlik ve aynı zamanda 0,8307 ile de 2. en yüksek duyarlılık değerine ulaşmıştır. 80 parametre, 0,7259 F1 skoru ile en yüksek doğruluk değeri ile birlikte, en yüksek F1 skoruna da ulaşmıştır. Bu sonuçlarla, 30 ve 80 parametre ile gerçekleştirilen çalışmalar, en başarılı performansları göstermiştir.

5.4.2 Algoritma Parametre Optimizasyonları

Çalışmada kullanılan algoritmaların parametre değerleri olarak, öndeğer olarak geçerli olan değerler seçilmiştir. Bu bölümde seçilen algoritmaların parametreleri değiştirilerek, en iyi parametrelerin seçimi ile performans arttırımı amaçlanmıştır. Belirlenen parametre havuzundan, belirlenen performans kriterine göre en iyi sonucu veren parametre seçimine sağlayan GridSearch yöntemi, k=10 katlamalı çapraz doğrulama ile birlikte kullanılmıştır. Seçilen yeni algoritma parametreleri ile model tekrar çalıştırıldığında elde edilen sonuçlar paylaşılmıştır.

5.4.2.1 GS-RF GridSearch ile Rassal Orman Hiper Parametre Optimizasyonu

GridSearch yöntemi, hiper parametre optimizasyonu için yaygın olarak kullanılan algoritmalarından biridir (James & Yoshua, 2012). Parametreleri optimize edilecek algoritmanın, belirlenen parametreler havuzu içindeki en iyi performansı veren değerleri belirler. KTS modelinde, rassal orman algoritmasının parametrelerinin optimizasyonu için GridSearch yöntemi kullanılmıştır (GS-RF).

Rassal orman algoritmasının parametreleri aşağıdaki gibi belirlenmiştir:

```
Parametreler = {'n_estimators': [2, 4, 6, 8, 9, 10],  
                'max_features': ['log2', 'sqrt', 'auto'],  
                'criterion': ['entropy', 'gini'],  
                'max_depth': [2, 3, 4, 5, 8, 10],  
                'min_samples_split': [2, 3, 5],  
                'min_samples_leaf': [1,2,4,5,8]}
```

}

GridSearch ile yapılan seçim sonucunda en iyi performansı veren parametreler, Tablo 5.13’de gösterilmiştir.

Tablo 5. 13: GridSearch ile Seçilen Rassal Orman Hiper Parametreleri

Parametre	Seçilen Değer
criterion	entropy
max_depth	10
max_features	log2
min_samples_leaf	5
min_samples_split	2
n_estimators	9

Rassal orman algoritması için, GridSearch ile parametre seçimi uygulandığında, modeldeki karmaşıklık matrisi değerlerinin değişimi Tablo 5.14’te verilmiştir. Sonuçlar değerlendirildiğinde, Rassal Orman algoritmasında GridSearch ile hiperparametre optimizasyonu ile doğruluk ve duyarlılık değerlerinin arttığı görülmektedir. Projede özellikle duyarlılık ve F1 Skoru değerlerine baktığımızda, duyarlılık değerinin %52,57’den %75,25’e çıktığını ve böylelikle kaçak kullananlar kümesinde, doğru tahmin edilen kaçak kullananlarının oranının arttığı görülmektedir. Projede F1 Skoru da, %57,98 den %71,54 değerine ulaşılmıştır. Aynı şekilde doğruluk değerinde de %58,97’den %68,60 başarımlı değerine ulaşılmıştır. Böylelikle GS-RF ile RF’e göre tüm değerlerde performans artışı gözlemlenmiştir.

Tablo 5.14: Rassal Orman - GridSearch Değerlendirme Sonuçları

Değer	RF	RF GridSearch
Doğruluk	0.5897	0.6860
Duyarlılık	0.5257	0.7525
Kesinlik	0.6463	0.7491
F1 Skoru	0.5798	0.7154

5.4.2.2 SVM Parametre Optimizasyonu

Rassal Orman algoritmasında, farklı çekirdek tipleri ve polinomial çekirdek fonksiyonu dereceleri ile model değerlendirilmiştir. Tablo 5.15, farklı çekirdek ve fonksiyon derecelerine ait karmaşıklık matrisleri değerlerini göstermektedir.

Tablo 5.15: SVM Parametre Seçimine Göre Karmaşıklık Matrisi

Kernel	Derece	Doğruluk	Duyarlılık	Kesinlik	F1 Skoru
Rbf	3	0.6793	0.8283	0.6614	0.7355
Sigmoid	3	0.6337	0.7435	0.6369	0.6861
Linear	3	0.6742	0.8166	0.6594	0.7296
Poly	3	0.6793	0.8241	0.6625	0.7345
Poly	4	0.6810	0.8129	0.6672	0.7329
Poly	5	0.6767	0.7912	0.6689	0.7249

Tablo incelendiğinde; 0,6810 değeri ile en iyi doğruluk sonucunu, 4. dereceden “poly” çekirdek tipi; 0,8243 değeri ile de en iyi duyarlılık sonucunu 3. dereceden “rbf” çekirdek tipi vermektedir. En iyi kesinlik değeri 5. dereceden “poly” olurken, F1 skorunda en iyi dğer “rbf” çekirdek tipinde 3. derece fonksiyonda elde edilmiştir.

5.4.3 Metasezgisel Yöntemler ile Öznitelik Seçimi

Metasezgisel algoritmalar, doğadaki fiziksel ve biyolojik olaylardan esinlenerek geliştirilmiş algoritmalar (Can, 2015). Son yıllarda bilimin birçok alanında, metasezgisel algoritmalar aktif olarak kullanılmaktadır (Kaveh, 2017). Bu algoritmalar, genellikle karmaşık problemlerin çözümünde daha başarılı performans göstermektedir (Talbi, 2009). Metasezgisel algoritmalar, çözüm uzayındaki optimum çözüme ulaşırken, verimli arama tekniklerini kullanarak, yerel optimum noktalara takılmadan global sonuca ulaşabilmektedir. Metasezgisel algoritmalar genellikle; evrimsel, fizik tabanlı, insan tabanlı ve sürü tabanlı olmak üzere dört sınıfa ayrılır (Seyyedabbasi & Kiani, 2022). Metasezgisel algoritmaların sınıflandırılması ve ilgili sınıfta kullanılan bazı algoritmalar, Şekil 35’te gösterilmiştir.

METASEZGİSEL ALGORİTMALAR



Şekil 5.24: Metasezgisel Algoritma Kategorileri

Evrimsel algoritmalar, Darwin'in evrim teorisinden esinlenen ve evrimsel konsepti taklit eden algoritmalarlardır. Genetik Algoritmalar (GA) (Goldberg D.E., 1988), evrimsel algoritmalarından en bilineni olup, Differansiyel Evrim (DE) (Storn R, 1997), Tabu Arama (TA) (Glover, 1990), Biyocoğrafya Tabanlı Optimizasyon (BTO) (Simon, 2008), Karadul Örümcek Optimizasyonu (KÖO) (Hayyolalam V., 2020), Evrimsel Programlama (EP) (Yao, 1999) algoritmaları da literatürde kullanılan bazı diğer evrimsel algoritmalarlardır.

Fizik tabanlı algoritmalar, eylemsizlik kanunu, yerçekimi kanunu ve elektromanyetik güç gibi doğadaki fizik kanunlarından esinlenerek ortaya çıkan algoritmalarlardır. Yerçekimsel Arama Algoritması (YAA) en bilinen fizik tabanlı algoritma olup (Rashedi, 2009), Büyük Patlama-Büyük Çöküş (BPBÇ) (Erol, 2006), Kara Delik Algoritması (KDA) (Hatamlou, 2013), Kavisli Uzay Optimizasyonu (KUO) (Moghaddam, 2012), Galaksi Tabanlı Arama Algoritması (GTAA) (Shah-Hosseini, 2011) da kullanılan diğer fizik tabanlı metasezgisel algoritmalarlardır.

İnsan tabanlı metasezgisel algoritmalar, optimizasyon problemlerinin çözümü için, insanların sosyal davranışlarını örnek alır. Futbol Ligi Yarışması Algoritması (FLYA) (Moosavian, 2014), Öğretme Öğrenme Bazlı Optimizasyon (ÖÖBO) (Rao, 2011), Politik Optimizasyon (PO) (Askari, 2020), Açgözlü Politika Optimizasyonu (APO) (Melvix J., 2014), Parlamentar Optimizasyon Algoritması (POA) (Borji, 2007), Ölüm Oyunu Optimizasyonu (ÖOO) (Rahkar Farshi, 2021) insan tabanlı algoritmalarlardır.

Sürü tabanlı algoritmalar, kuş, kurt, ateşböceği, kedi ve arı gibi canlı sürülerin hareketlerinden esinlenerek geliştirilen algoritmalar. Sürü içindeki bireyler; diğer bireylerin davranışları, kendi deneyimleri ve en iyi bireyin davranışlarını kullanarak bilgi edinmekte, karşılaşacakları problemlerin çözümünde de ileride bu bilgileri kullanmaktadır. Örneğin bir sürü içerisindeki bir bireyin, bir tehlike ile karşılaştığında verdiği tepki, sürü içerisindeki diğer tüm bireyleri de etkileyerek, sürünün ortak bir davranış sergilemesini sağlamaktadır. Sürüler içerisindeki bu hareketler gözlemlenerek, sürü tabanlı algoritmalar geliştirilmiştir. Parçacık Sürü Optimizasyonu (PSO) (Kennedy, 1995) ve Karınca Kolonisi Optimizasyon (KKO) (Dorigo, 2006), en bilinen sürü tabanlı algoritmalar olup, diğer sürü tabanlı algoritmalar ise Yapay Arı Kolonisi (YAK) (Karaboğa, 2017), Yarasa Algoritması (YA) (Yang, 2010), Ateş Böceği Algoritması (ABA) (Yang, Firefly algorithms for multimodal optimization, 2009), Gri Kurt Optimizasyon Algoritması (GKO) (Mirjalili S. M., 2014), Kelebek Optimizasyon Algoritması (KOA) (Arora S., 2019), Kedi Sürüsü Optimizasyon Algoritması (KSO) (Chu, 2006) ve Salp Sürü Algoritması (SSA) (Mirjalili S. G., 2017) algoritmalarıdır.

5.4.3.1 PSO – Parçacık Sürü Optimizasyonu

PSO – Parçacık Sürü Optimizasyonu, 1995 yılında Kennedy and Eberhart tarafından bulunan bir optimizasyon tekniğidir (J. Kennedy, 1995). Sürekli değişkenler için geliştirilmiş olup, ikili model için de çözüm üretmektedir (BPSO – Binary Particle Swarm Optimization). PSO algoritması, sezgisel optimizasyon tekniklerinden biri olup, kuş ve balık sürülerinin sosyal davranışlarından etkilenecek şekilde geliştirilmiştir (Hiram A). PSO algoritmasında, parçacık adı verilen potansiyel çözümler, problem uzayında uçuş yaparlar. PSO’da her çözüm, arama uzayında bir kuştur. Her bir parçacık, pozisyon ve hız bilgilerine sahiptir. Parçacıklar mevcut optimum parçacıkları takip ederek uçarlar.

Her iterasyonda her bir parçacık, gbest ve pbest olarak belirtilen değerlere göre güncellenir. “Pbest” her bir parçacığın en iyi değerini gösterirken, “gbest” ise global en iyi olarak bilinen herhangi bir parçacık tarafından o ana kadar elde edilen en iyi uygunluk değerini içerir.

D boyutlu bir uzayda, her bir parçacık için pozisyon vektörü:

$$X_i = (X_{i1}, \dots, X_{ij}, \dots, X_{iD}) \quad (5.1)$$

hız vektörü ise

$$V_i = (V_{i1}, \dots, V_{ij}, \dots, V_{iD}) \quad (5.2)$$

olarak ifade edilir.

Arama işlemi boyunca, her parçacık kendi pozisyonuna ve her iterasyon sonucundaki en iyi çözüme göre kendi hızını değiştirir.

Klasik PSO sürekli uzayda işlediğinden, öznitelik seçimi gibi ikili çözüm içeren problemlerde kullanılamaz. Bu işlem için geliştirilmiş PSO olan, DBPSO – Discrete Binary Particular Swarm Optimization kullanılmaktadır (R.C.Eberhart, 1997).

DBPSO algoritmasında hız değiştirme formülü:

$$v_{ij}^{t+1} = \omega \times v_{ij}^t + c_1 \times \text{rand}() \times (p_{ij} - x_{ij}^t) + c_2 \times \text{rand}() \times (g_j - x_{ij}^t) \quad (5.3)$$

Hız aşağıdaki gibi tekrar tanımlandığında:

$$\text{Sigmoid}(v_{ij}^{t+1}) = 1/(1 + e^{-v_{ij}^{t+1}}) \quad (5.4)$$

Sonuç olarak ikili bir vektör olarak aday çözüm aşağıdaki gibi ifade edilir.

$$x_{ij}^{t+1} = \begin{cases} 1 & \text{if } \text{rand}() < \text{Sigmoid}(v_{ij}^{t+1}), \\ 0 & \text{otherwise.} \end{cases} \quad (5.5)$$

$\text{rand}()$: 0 ile 1 arasında rastgele bir sayı

v_{ij}^t : t. iterasyondaki i. bireyin j. elemanı

ω : ağırlık

c_1, c_2 : öğrenme faktörleri

p_i : local en iyi pozisyon

g_j : global en iyi pozisyon

PSO algoritması için prosedür aşağıda verilmiştir (C.Karakuzu).

For her parçacık için başlangıç koşullamaları

End

Do For her parçacık için uygunluk degerini hesapla

eger uygunluk degeri, pbest ten daha iyi

ise; simdiki degeri yeni pbest olarak ayarla

End

Tüm parçacıkların buldugu pbest degerlerinin en

iyisini, tüm parçacıkların gbest'i olarak ayarla

For her parçacık için

(2) denklemine göre parçacık hızını

hesapla

(3)denklemine göre parçacık pozisyonunu

güncelle

End

While maksimum iterasyon sayısına veya minimum

hata kosulu saglanana kadar devam et

(C.Karakuzu)

BPSO ile Öznitelik Seçimi

Modeldeki 144 öznitelikten en iyi çözümü üreten öznitelik alt kümesi seçimi için BPSO – İkili parçacık sürü algoritması kullanılmıştır. BPSO ile seçilen alt küme öznitelikler, ikili parçacık sürü rassal orman (RF-BPSO) ve ikili parçacık sürü lojistik regresyon (LR-BPSO) algoritmaları ile oluşturulan modellere uygulanmıştır.

I : Maksimum İterasyon Sayısı, n = Parçacık Sayısı olmak üzere, BPSO algoritmasında kullanılan sabit parametreler aşağıdaki gibidir.

{ $\omega = 0,9$, $c_1 = 0,5$, $c_2 = 0,5$, $k=30$, $p=2$ }

İterasyon sayısı I = 100 için, parçacık sayısı $n=\{60, 70, 80, 90, 100\}$ ve $n=50$ olmak üzere, $I=\{50, 60, 70, 80, 90, 100\}$ değerleri için; RF-BPSO ve LR-BPSO ile öznitelik seçimi yapılarak, algoritmaların karmaşık matrisi performans değerleri hesaplanmıştır.

Tablo 5.16: RF-BPSO ile Öznitelik Seçimi

Değer k=10	I=50 n=100	I=50 n=90	I=50 n=80	I=50 n=70	I=50 n=60	I=50 n= 50	I=60 n=50	I=70 n=50	I=80 n=50	I=90 n=50	I= 100 n=50
Öznitelik Say.	117	113	117	114	112	117	118	113	114	113	116
Doğruluk	0.6046	0.5903	0.6317	0.6551	0.6223	0.6456	0.6088	0.6599	0.6511	0.6474	0.6251
Duyarlılık	0.8267	0.6990	0.6126	0.8272	0.8442	0.7266	0.5771	0.7960	0.7562	0.7350	0.7435
Kesinlik	0.5956	0.6031	0.6737	0.6387	0.6073	0.6538	0.6552	0.6505	0.6516	0.6533	0.6283
F1 Skoru	0.6924	0.6475	0.6417	0.7208	0.7064	0.6883	0.6137	0.7159	0.7000	0.6918	0.6811

Her bir parçacık değeri için, seçilen alt öznitelik sayısı ile k=10 katlamalı çapraz doğrulama kullanılarak elde edilen ortalama değerler, RF-BPSO ve LR-BPSO için sırasıyla Tablo 5.16 ve Tablo 5.17’de verilmiştir. Tablo 5.16 incelendiğinde, RF-BPSO için en iyi performans 114 alt öznitelik kullanılarak, I=50 iterasyon ve n=70 parçacık sayısı için, 0,6551 doğruluk, 0,7208 f1 değerleri ile elde edilmiştir. En iyi duyarlılık değerine I=50 n=60 için 112 parametre ile 0,8442 ile, en iyi kesinlik değerine ise 0,6737 olarak, 117 alt öznitelik ve I=50, n=80 parametreleri ile ulaşılmıştır. RF-BPSO ile seçilen 114 öznitelik Ek 2’de, verilmiştir.

Tablo 5.17: LR-BPSO ile Öznitelik Seçimi

Değer k=10	I=50 n=100	I=50 n=90	I=50 n=80	I=50 n=70	I=50 n=60	I=50 n= 50	I=60 n=50	I=70 n=50	I=80 n=50	I=90 n=50	I= 100 n=50
Öznitelik Say.	111	116	114	112	113	114	113	114	115	118	118
Doğruluk	0.6730	0.6539	0.6742	0.6805	0.6762	0.6194	0.6756	0.6294	0.6291	0.6762	0.6670
Duyarlılık	0.8241	0.7504	0.8251	0.8098	0.7944	0.8288	0.8145	0.7652	0.6513	0.8426	0.7848
Kesinlik	0.6564	0.6562	0.6572	0.6675	0.6674	0.6074	0.6614	0.6278	0.6569	0.6549	0.6606
F1 Skoru	0.7307	0.7001	0.7317	0.7318	0.7254	0.7010	0.7300	0.6898	0.6541	0.7370	0.7174

LR-BPSO algoritması için en iyi performanslar, I=50, n=70, 109 parçacık ile 0,6805 duyarlılık, 0,6675 kesinlik ile; I=90, n=50 parametreleri ve 118 alt öznitelik seçimi 0,8426 duyarlılık, 0,7370 f1 skoru değerleri ile elde edilmiştir. Tablo 5.18 en iyi performansı gösteren parametreleri göstermektedir. Ek 3’te, LR-BPSO ile seçilen 118 öznitelik gösterilmiştir.

Tablo 5.18: BPSO Öznitelik Seçimi Değerlendirme

Değer	RF-BPSO	LR-BPSO
Öznitelik Sayısı	114	118
i (iterasyon)	50	90
n (parçacık sayısı)	70	50
Doğruluk	0.6551	0.6762
Duyarlılık	0.8272	0.8426
Kesinlik	0.6387	0.6549
F1 Skoru	0.7208	0.7370

Sürü tabanlı metasezgisel yöntemlerden BPSO, Rassal Orman ve Lojistik Regresyon algoritmaları ile birlikte uygulandığında, BPSO-RF ve BPSO-LR karmaşıklık matrisi sonuçlarının, RF vr LR ile karşılaştırma tablosu, Tablo 5.18’de verilmiştir. Sonuçlar incelendiğinde, özellikle BPSO-RF ile ciddi performans kazanımı sağlanmıştır. Kesinlik dışındaki tüm değerlerde performans artışı gözlemlenmiştir. Doğruluk 0,5897’den 0,6551’e, duyarlılık 0,5257’den 0,8272’ye, f1 skoru da 0,5798’den 0,7208’e çıkmış olup, sadece kesinlik değeri 0,6463’ten 0,6387’ye düşmüştür. BPSO-LR tarafında ise sadece duyarlılıkta 0,8129’dan 0,8426’ya bir artış gözlemlenmiştir. Sonuç olarak, BPSO-LR başarılı bir performans göstermezken, BPSO-RF tüm değerlerde önemli ölçüde performans artışı sağlamıştır.

Tablo 5.19: BPSO Performans Karşılaştırma

Değer	RF	BPSO-RF	LR	BPSO-LR
Doğruluk	0.5897	0.6551	0.6819	0.6762
Duyarlılık	0.5257	0.8272	0.8129	0.8426
Kesinlik	0.6463	0.6387	0.6681	0.6549
F1 Skoru	0.5798	0.7208	0.7334	0.7370

ALTINCI BÖLÜM

SONUÇ VE ÖNERİLER

Enerjide dışa bağımlılığımızın yüksek olduğu ülkemizde kaçak elektrik kullanımı, ekonomimize de ciddi etkisi olan önemli sorunlarımızdan bir tanesidir. Özellikle Türkiye’de elektrik kurumlarının özelleştirilmesi sonrası, enerji sektöründeki regülatör kurum olan EPDK tarafından, kaçağın düşürülmesi noktasında yapılan çalışmalar ve verilen teşvikler, bu konuda önemli iyileştirmeler sağlamıştır. Bu çalışmalar sayesinde ülke genelinde kayıp kaçak oranı, 2014’te yüzde 14,6’ya, 2015’te yüzde 14’e, 2016’da yüzde 13,4’e, 2017’de yüzde 12,6’ya ve 2018 sonu itibarıyla da yüzde 11,8’e düşmüştür.

Ülkemizde kayıp kaçağın en yüksek olduğu bölgeler, güneydoğu bölgelerimizdir. Elektrik dağıtım sektörü, 21 bölgeye ayrılarak bölge bazında özelleştirilmiştir. Bu bölgeler içinde Dicle EDAŞ, Vangölü EDAŞ, Aras EDAŞ olmak üzere, bu üç bölgedeki kaçak kullanımı, Türkiye ortalamasının çok üzerindedir. Türkiye’de toplam kaçak oranını arttıran bu bölgeler içinde, 2019 için Aras EDAŞ %21,64 kaçak oranına sahipken, Vangölü EDAŞ %47,56, Dicle EDAŞ ise %51,32 gibi çok yüksek bir kaçak oranına sahiptir. Kaçak oranı yüksekliğinde dördüncü sırada olan Çamlıbel EDAŞ ise sadece %4,75 kaçak oranına sahiptir. Bu dağılımdan da görülebileceği üzere, özellikle bu üç elektrik dağıtım bölgesinde bu alanda yapılacak çalışmalar, Türkiye’de kaçak oranının düşürülmesinde önemli katkı sağlayacaktır.

Tez kapsamında yapılan bu çalışma, Türkiye’nin en büyük kaçak oranına sahip Dicle EDAŞ uygulama sistemleri verileri kullanılarak gerçekleştirilmiştir. Bu amaçla, birden fazla büyük canlı sistem verileri konsolide edilerek oluşturulan veri ambarı kullanılmıştır. İlgili uygulama sistemleri, bu uygulama sistemlerinde dijital olarak takip edilen iş süreçleri ve bu sistemlere ait veritabanlarında tutulan verilerin yapıları üzerinde çalışmalar yapılmıştır. Dicle EDAŞ’ta mevcut uzman sistem yaklaşımı ile yapılan önceki çalışmalar detaylı incelenmiştir. Veri ön işleme kapsamında, kirli verilerin temizlenmesi ve eksik verilerin doldurulması çalışmaları gerçekleştirilmiştir.

Oluşturulan KTS modelinde, LR, RF, SVM, KNN, DT, XGB makina öğrenmesi algoritmaları ile birlikte, K-katlamalı çapraz doğrulama tekniği kullanılmıştır. KTS performans ölçümü için karmaşıklık matrisi kullanılmıştır. Modelde tüm veriler içinde

kaçak abonelerin tespiti önemli olduğundan, duyarlılık metriği başta olmak üzere doğruluk, kesinlik ve f1 skoru değerleri izlenmiştir. Son bölümde model iyileştirme çalışmaları kapsamında, alt öznitelik seçimi, algoritma parametrelerinin optimizasyonu gibi performans artırma odaklı yöntemler geliştirilmiştir. Ek olarak meta sezgisel algoritmalar incelenmiş olup, bu algoritmalarından BPSO algoritması ile optimum öznitelik çıkarımı gerçekleştirilmiştir.

Çalışma sonucunda k=10 katlamalı çapraz doğrulama ile LR doğruluk ve duyarlılık değerleri sırasıyla %67,11 ve %81,70 olarak elde edilmiştir. SVM için bu doğruluk ve duyarlılık değerleri %66,70 ve %83,07 iken, KNN için %62,52 ve %62,23; DT için %63,54 ve %64, RF için %68,12 ve %67,34; XGB için %71,03 ve %73,51 olarak hesaplanmıştır. LR-RFE ile iyileştirme sonrası, 80 parametre için %67,38 doğruluk, %82,29 duyarlılık değerlerine ulaşılmıştır. RF-BPSO modeli ile öznitelik seçiminde 114 parametre ile %65,51 doğruluk ve %82,72 duyarlılık, LR-BPSO kullanılarak 118 parametre ile %67,62 doğruluk ve %84,26 duyarlılık değerleri elde edilmiştir. Sonuçta en yüksek doğruluk değerine %71,03 olarak XGB ile ulaşılırken, en yüksek duyarlılık değeri %84,26 olarak LR-BPSO ile ulaşılmıştır. LR-RFE 10 parametre ile elde edilen %93,41 duyarlılık değeri, doğruluk değeri %54,48 değeriyle düşük kaldığından dikkate alınmamıştır.

Verilerin kirli ve bozuk olması, çalışma performansını olumsuz yönde etkilemiştir. Dengesiz veri setinin dengeli hale getirilmesi noktasında, farklı yöntemler kullanılarak farklı veri setleri oluşturulmuştur. Bu veri setleri ayrı ayrı değerlendirilerek, en uygun veri seti seçilmiş ve dengeli veri seti ile çalışmalara devam edilmiştir. Sistemde yer alan kaçak tutanakları, saha personellerinin tespitlerine istinaden düzenlendiğinden, saha çalışmaları sonuçları doğru kabul edilmiştir.

Bu tezde yapılan çalışma, Türkiye’de makine öğrenmesi tekniklerinin kullanılarak kaçak tespiti noktasında gerçekleştirilen ilk çalışma olduğu için, bu alana yönelik örnek teşkil etmektedir. Çalışmada elde edilen sonuçların kaçak tespitinde kullanılabilir olması, ülkemizde kaçak elektrik kullanımı ile mücadeleye katkı sağlayacaktır.

KAYNAKÇA

- Akdemir, B. (2009). *Tahmin Uygulamalarında Performans Geliştirmek için Kullanılan Normalizasyon Metodlarına Yeni Bir Yaklaşım* (Doktora Tezi). Konya, Türkiye: Selçuk Üniversitesi Fen Bilimleri Enstitüsü.
- Amin, S., Schwartz, G., & Tembine, H. (2012). Incentives and Security in Electricity Distribution Networks. *GameSec 2012: Decision and Game Theory for Security*: 264-280.
- Angelos, E., Saavedra, O., Cortes, O., & Souza, A. (2011). Detection and identification of abnormalities in customer consumptions in power distribution systems. *IEEE Transactions on Power Delivery*, 26(4): 2436-2442.
- Arora S., S. S. (2019). Butterfly optimization algorithm: a novel approach for global optimization. *Soft Computing*, 23(3): 715-734.
- Askari, Q. Y. (2020). Political Optimizer: A novel socio-inspired meta-heuristic for global optimization. *Knowledge-Based Systems*: 195, 1-25.
- Borji, A. (2007). A New global optimization algorithm inspired by parliamentary political competitions. *MICAI 2007: Advances in Artificial Intelligence: 6th Mexican International Conference on Artificial Intelligence 6*: 61-71
- C.Karakuzu, S. T. (2006). Parçacık Sürüsü Optimizasyon Algoritması ve Benzetim Örnekleri. *ELECO`2006 Elektrik-Elektronik-Bilgisayar Mühendisliği Sempozyumu*.
- Cabral, J., Pinto, J., & Pinto, A. (2009). Fraud detection system for high and low voltage electricity consumers based on data mining. *IEEE Power & Energy Society General Meeting* (pp. 1-5).
- Can, U. A. (2015). Physics based metaheuristic algorithms for global optimization. *American Journal of Information Science and Computer Engineering*: 94,106.
- Chu, S. T. (2006). Cat swarm optimization. In Pacific Rim international conference on artificial intelligence . *International conference on artificial intelligence*: 854-858.Springer, Berlin, Heidelberg.

- CIREN. (2017). Reduction of Technical and Non-Technical Losses in Distribution Networks. *Congrès International des Réseaux Electriques de Distribution: CIREN Overview*, final report: 14-25.
- Costa, B. C., Alberto, B., Portela, A., Maduro, W., & Eler, E. (2013). Fraud detection in electric power distribution networks using an annbased knowledge-discovery process. *International Journal of Artificial Intelligence & Applications*, 4(6): 17.
- Çelik, Y. (2013). *Optimizasyon Problemlerinde Bal Arılarının Evlilik Optimizasyonu Algoritmasının Performansının Geliştirilmesi* (Doktora Tezi). Konya, Türkiye: Selçuk Üniversitesi Fen Bilimleri Enstitüsü.
- Davidson, E., Odubiyi, A., Kachienga, M. O., & Manhire, B. (2002). Techinal loss computation and economic dispatch model for T&D systems in deregulated ESI. *Power Engineering Journal*, 16(2): 55-60.
- Davidson, I. E. (2003). Evaluation and effective management of nontechnical losses in electrical power networks. *The Transactions of The South African Institute of Electrical Engineers* 94(3): 39-42.
- Depuru, S. S., Wang, L., & Devabhaktuni, V. (2011). Support vector machine based data classification for detection of electricity theft. *In Power Systems Conference and Exposition (PSCE)*: 1-8.
- Depuru, S. S., Wang, L., & Devabhaktuni, V. (2012). Enhanced encoding technique for identifying abnormal energy usage pattern. *North American Power Symposium (NAPS)*: 1-6.
- Dorigo, M. M. (2006). Ant colony optimization. *IEEE Computational Intelligence Magazine*, 1(4): 28-39.
- EPDK. (2019). *Elektrik Piyasası Gelişim Raporu*. Ankara: Strateji Gelişirme Dairesi Başkanlığı.
- Ergün, K. (2019). Veri Madenciliği (Veri Ön İşleme-2) Sunumu. *Endüstri Mühendisliği Bölümü Veri Madenciliği Dersi*. Balıkesir Üniversitesi.
- Erol, O. E. (2006). A new optimization method: big bang–big crunch. *Anv Eng Softw.*, 37(2): 106,111.

- Eskidere, Ö. (2012). Ses Ölçümlerinden Parkinson Hastalığının Teşhisi İçin Öznitelik Seçme Yöntemlerinin Karşılaştırılması. *Sigma* 30, 402-414.
- Fourie, J., & Calmeyer, J. (2005). A statistical method to minimize electrical energy losses in a local electricity distribution network. *IEEE 7th AFRICAON Conference*: 668.
- Glauner, P., Boechat, A., Dolberg, L., State, R., Bettinger, F., Rangoni, Y., & Duarte, D. (2016). Large-scale detection of non-technical losses in imbalanced data sets. In *Innovative Smart Grid Technologies Conference (ISGT)*. IEEE Power & Energy Society.
- Glauner, P., Meira, J. A., Valtchev, P., State, R., & Bettinger, F. (2017). The Challenge of Non-Technical Loss Detection using Artificial Intelligence: A Survey. *International Journal of Computational Intelligence Systems*, 10(1): 760-775.
- Glover, F. (1990). Tabu Search: A Tutorial. *Inform Journal on Applied Analytics*, 20(4): 75-94.
- Goldberg D.E., H. J. (1988). Genetic algorithms and machine learning. *Mach Learn*, 3: 95-9.
- Gontijo, J., Delaiba, A., & Pinto, E. (2004). Fraud identification in electricity company customers using decision tree. *International Conference on Systems, Man and Cybernetics*, 4: 3730-3734.
- Hatamlou, A. (2013). Black hole: A new heuristic optimization approach for data clustering. *Information Sciences*, 222: 175-184.
- Hayyolalam V., K. A. (2020). Black widow optimization algorithm: A novel metaheuristic approach for solving engineering optimization problems. *Engineering Applications of Artificial Intelligence*: 87: 103249, 1-28.
- Hiram A, F. E. (tarih yok). Swarmed Feature Selection. *Proceedings of the 33rd Applied Imagery Pattern Recognition Workshop*: 1550-5219.
- J. Kennedy, R. (1995). Particle Swarm Optimization. *Proceedings IEEE International Conference on Neural Networks*: 1942-1948.

- James, B., & Yoshua, B. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research* 13: 281-305.
- Jayalakshmi, T., & Santhakumaran, D. (2011, February). Statistical Normalization and Back Propagation for Classification. *International Journal of Computer Theory and Engineering*, 3(1): 1793-8201.
- Karaboğa, D. B. (2017). Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems. *International fuzzy systems association world congress: 789-798*. Berlin, Heidelberg.
- Kaveh, A. (2017). *Applications of metaheuristic optimization algorithms in civil engineering*. Basel, Switzerland: Springer International Publishing.
- Kennedy, J. E. (1995). Particle swarm optimization. *International Conference on Neural Networks 4*: 1942-1948.
- Melvix J. (2014). Greedy politics optimization: Metaheuristic inspired by political strategies adopted during state assembly elections. *IEEE International Advance Computing Conference*: 1157-1162. India.
- Mirjalili, S. G. (2017). Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. *Advances in Engineering Software: 114*: 163-191.
- Mirjalili, S. M. (2014). Grey wolf optimizer. *Advances in engineering software: 69*: 46-61.
- Moghaddam, F. M. (2012). Curved space optimization: a random search based on general relativity theory. *arXiv preprint*, arXiv:1208.2214.
- Moosavian, N. R. (2014). Soccer league competition algorithm: A novel metaheuristic algorithm for optimal design of water distribution networks. *Swarm Evol.Comput.*, 17: 14-24.
- Muniz , C., Marley, M., Tansheit, R., & Figueiredo, K. (2009). A neuro-fuzzy system for fraud detection in electricity distribution. *IFSA/EUSFLAT Conference*: 1096-1101.

- Nagi, J., Mohammad, M., Yap, K., Tiong, K., & Ahmed, S. (2008). Non-technical loss analysis for detection of electricity theft using support vector machines. *2nd IEEE International Conference on Power and Energy (PECon 08)*. Johor Bahru, Malaysia.
- Nagi, J., Yap, K. S., Tiong, S. K., Ahmed, S. K., & Mohamad, M. (2010). Nontechnical loss detection for metered customers in power utility using support vector machines. *Transactions on Power Delivery*, 25(2): 1162,1171.
- Nagi, J., Yap, K., Nagi, F., Tiong, S., Koh, S., & Ahmed, S. (2010). NTL Detection of Electricity Theft and Abnormalities for Large Power Consumers In TNB Malaysia. *Proceedings of 2010 IEEE Student Conference on Research and Development (SCOReD 2010)*. Putrajaya, Malaysia.
- Nagi, J., Yap, K., Tiong, S., Ahmed, S., & Mohammad, A. (2008). Detection of abnormalities and electricity theft using genetic support vector machines. *TENCON 2008-2008 IEEE Region 10 Conference*: 1-6.
- Nagi, J., Yap, K., Tiong, S., Ahmed, S., Mohammad, A., & Nagi, F. (2011). Improving svm-based nontechnical loss detection in power utility using the fuzzy inference system. *IEEE Transactions on power delivery*, 26(2): 1284-1285.
- Navani, J., Sharma, N., & Sonal, S. (2012). Technical and Non-Technical Losses in Power System and Its Economic Consequence in Indian Economy. *International Journal of Electronics and Computer Science Engineering*, 1(2): 757-761.
- Nizar, A. H., Zhao, J. H., & Dong, Z. Y. (2006). Customer Information System Data Pre-Processing with Feature Selection Techniques for Non-Technical Losses Prediction in an Electricity Market. *International Conference on Power System Technology*.
- R.C.Eberhart, J. a. (1997). A Discrete Binary Version of The Particle Swarm Algorithm. IEEE.
- Rahkar Farshi, T. (2021). Battle royale optimization algorithm. *Neural Computing & Applications*, 33(4): 1139-1157.

- Ramos, C., Souza, A., & Gastaldello, D. (2012). Identification and feature selection of non-technical losses for industrial consumers using the software WEKA. *In Industry Applications (INDUSCON) 10th IEEE/IAS International Conference*: 1-6.
- Ramos, C., Souza, A., Papa, J., & Falcao, A. (2009). Fast non-technical losses identification through optimum-path forest. *Intelligent System Applications to Power Systems*: 1-5.
- Rao, R. S. (2011). Teaching-learning-based optimization: A novel method for constrained mechanical design optimization problems. *Comput.Aided Des.*, 43(3): 303-315.
- Rashedi, E. N.-P. (2009). GSA: a gravitational search algorithm. *Information Sciences*, 179(13): 2232-2248.
- Sahoo, S., Nikovski, D., Muso, T., & Tsuru, K. (2015). Electricity theft detection using smart meter data. *Innovative Smart Grid Technologies Conference (ISGT)*: 1-5.
- Seyyedabbasi, A., & Kiani, F. (2022). Sand Cat swarm optimization: a nature-inspired algorithm to solve global optimization problems. *Enginnering with Computers*: 1-25.
- Shah-Hosseini, H. (2011). Principal components analysis by the galaxy-based search algorithm: a novel metaheuristic for continuous optimisation. *International Journal of Computational Science and Engineering*, 6(1-2): 312-140.
- Simon, D. (2008). Biogeography-based optimization. *IEEE Transactions on Evolutionary Computation*, 12(6): 702-713.
- Smith, T. (2004). Electricity theft: a comparative analysis. *Energy Policy* 32: 2067-2076.
- Spiric, J. V., Stankovic, S. S., Docic, M. B., & Popovic, T. D. (2014). Using the rough set theory to detect fraud committed by electricity customers. *International Journal of Electrical Power*, 62: 727-734.
- Storn R, P. K. (1997). Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. *J Glob Optim*, 11(4): 341-359.

- Talbi, E. (2009). *Metaheuristics: from design to implementation*. John Wiley & Sons.
- Xia, X., Liang, W., Xiao, Y., Zheng, M., & Xiao, Z. (2015). A Difference-Comparison-based Approach for Malicious Meter Inspection in Neighborhood Area Smart Grids. *IEEE ICC Communication for the Smart Grid*.
- Yang, X. (2009). Firefly algorithms for multimodal optimization. *International symposium on stochastic algorithms*: 169-178. Springer, Berlin, Heidelberg.
- Yang, X. (2010). A new metaheuristic bat-inspired algorithm. *Proceedings of the workshop on nature inspired cooperative strategies for optimization (NICSO 2010)*, Springer: 65-74.
- Yao, X. L. (1999). Evolutionary programming made faster. *IEEE Transactions on Evolutionary Computation*, 3(2): 82-102.

EKLER

Ek 1: Türkiye’de Elektrik Dağıtım Şirketleri ve Bölgeleri

Elektrik Dağıtım Kurumları	Kısa Adı	Bölge Kapsamındaki İller
Dicle Elektrik Dağıtım	Dicle EDAŞ	Diyarbakır, Şanlıurfa, Batman, Mardin, Siirt ve Şırnak
Vangölü Elektrik Dağıtım	VEDAŞ	Bitlis, Hakkari, Muş
Aras Elektrik Dağıtım	Aras EDAŞ	Erzurum, Ağrı, Ardahan, Bayburt, Erzincan, Iğdır, Kars
Çoruh Elektrik Dağıtım	Çoruh EDAŞ	Trabzon, Artvin, Giresun, Gümüşhane, Rize
Fırat Elektrik Dağıtım	Fırat EDAŞ	Elazığ, Malatya, Bingöl, Tunceli
Çamlıbel Elektrik Dağıtım	Çamlıbel EDAŞ	Sivas, Tokat, Yozgat
Toroslar Elektrik Dağıtım	Toroslar EDAŞ	Adana, Gaziantep, Hatay, Kilis, Mersin, Osmaniye
Meram Elektrik Dağıtım	MEDAŞ	Konya, Aksaray, Niğde, Kırşehir, Nevşehir, Karaman
Başkent Elektrik Dağıtım	Başkent EDAŞ	Ankara, Bartın, Çankırı, Karabük, Kastamonu, Kırıkkale, Zonguldak
Akdeniz Elektrik Dağıtım	AEDAŞ	Antalya, Burdur, Isparta
Gediz Elektrik Dağıtım	Gediz EDAŞ	İzmir, Manisa
Uludağ Elektrik Dağıtım	UADEŞ	Bursa, Balıkesir, Çanakkale, Yalova
Trakya Elektrik Dağıtım	TREDAŞ	Kırklareli, Tekirdağ, Edirne
Anadolu Yakası Elektrik Dağıtım	AYEDAŞ	İstanbul Anadolu Yakası
Sakarya Elektrik Dağıtım	SEDAŞ	Sakarya, Kocaeli, Bolu, Düzce
Osmangazi Elektrik Dağıtım	OEDAŞ	Eskişehir, Afyon, Bilecik, Kütahya, Uşak
Boğaziçi Elektrik Dağıtım	BEDAŞ	İstanbul Avrupa Yakası
Kayseri ve Civarı Elektrik	KCETAŞ	Kayseri ve Civarı
Aydem Elektrik Dağıtım	Aydem EDAŞ	Denizli, Muğla, Aydın
Akedaş Elektrik Dağıtım	AKEDAŞ	Kahramanmaraş, Adıyaman
Yeşilirmak Elektrik Dağıtım	YEDAŞ	Samsun, Amasya, Çorum, Ordu, Sinop

EK 2: RF-BPSO ile Seçilen Öznitelikler (114)

RF-BPSO En İyi Doğruluk Değerini Veren Öznitelikler		
Resmi Daireler	OrtalamaGunlukTuketimAy6	OrtalamaAkim2Ay8
Kitler	OrtalamaGunlukTuketimAy7	OrtalamaAkim2Ay9
Cami-İbadethane	OrtalamaGunlukTuketimAy8	OrtalamaAkim2Ay10
Hizmet Binaları	OrtalamaGunlukTuketimAy9	OrtalamaAkim2Ay11
Arıtma Tesisleri	OrtalamaGunlukTuketimAy10	OrtalamaAkim2Ay12
Ticarethane	OrtalamaGunlukTuketimAy11	OrtalamaAkim3Ay1
Belediyeler	OrtalamaGunlukTuketimAy12	OrtalamaAkim3Ay2
Üreticiler	MaxDemandAy1	OrtalamaAkim3Ay3
İçme Suyu	MaxDemandAy2	OrtalamaAkim3Ay4
Şantiye	MaxDemandAy3	OrtalamaAkim3Ay5
Hayır Kurumları	MaxDemandAy4	OrtalamaAkim3Ay6
Karayolları+Park Aydınlatma	MaxDemandAy5	OrtalamaAkim3Ay7
Mesken	MaxDemandAy6	OrtalamaAkim3Ay8
Tarımsal Sulama	MaxDemandAy7	OrtalamaAkim3Ay9
Diğer	MaxDemandAy8	OrtalamaAkim3Ay10
Sekonder	MaxDemandAy9	OrtalamaAkim3Ay11
Primer	MaxDemandAy10	OrtalamaAkim3Ay12
ŞIRNAK	MaxDemandAy11	OrtalamaGerilim1Ay1
DİYARBAKIR	MaxDemandAy12	OrtalamaGerilim1Ay2
ŞANLIURFA	OrtalamaAkim1Ay1	OrtalamaGerilim1Ay3
SIİRT	OrtalamaAkim1Ay2	OrtalamaGerilim1Ay4
MARDİN	OrtalamaAkim1Ay3	OrtalamaGerilim1Ay5
BATMAN	OrtalamaAkim1Ay4	OrtalamaGerilim1Ay6
Köy	OrtalamaAkim1Ay5	OrtalamaGerilim1Ay7
VIK	OrtalamaAkim1Ay6	OrtalamaGerilim1Ay8
LUN	OrtalamaAkim1Ay7	OrtalamaGerilim1Ay9
ELM	OrtalamaAkim1Ay8	OrtalamaGerilim1Ay10
LNO	OrtalamaAkim1Ay9	OrtalamaGerilim1Ay11
AEL	OrtalamaAkim1Ay10	OrtalamaGerilim1Ay12

EK 2: RF-BPSO ile Seçilen Öznitelikler (114)**(Devamı)**

MSY	OrtalamaAkim1Ay11	OrtalamaGerilim2Ay1
YOK	OrtalamaAkim1Ay12	OrtalamaGerilim2Ay2
KuruluGuc	OrtalamaAkim2Ay1	OrtalamaGerilim2Ay3
OrtalamaTuketim	OrtalamaAkim2Ay2	OrtalamaGerilim2Ay4
OrtalamaGunlukTuketimAy1	OrtalamaAkim2Ay3	OrtalamaGerilim2Ay5
OrtalamaGunlukTuketimAy2	OrtalamaAkim2Ay4	OrtalamaGerilim2Ay6
OrtalamaGunlukTuketimAy3	OrtalamaAkim2Ay5	OrtalamaGerilim2Ay7
OrtalamaGunlukTuketimAy4	OrtalamaAkim2Ay6	OrtalamaGerilim2Ay8
OrtalamaGunlukTuketimAy5	OrtalamaAkim2Ay7	OrtalamaGerilim2Ay9

Ek 3: LR-BPSO ile Seçilen Öznitelikler (112 Öznitelik)

LR-BPSO En İyi Doğruluk Değerini Veren Öznitelikler		
Resmi Daireler	Ticarethane	Hayır Kurumları
Kitler	Belediyeler	Karayolları + Park Aydınlatma
Cami-İbadethane	Üreticiler	Mesken
Hizmet Binaları	İçme Suyu	Tarımsal Sulama
Arıtma Tesisleri	Şantiye	Diğer
Primer	MaxDemandAy1	OrtalamaGunlukTuketimAy1
Sekonder	MaxDemandAy2	OrtalamaGunlukTuketimAy2
Diğer	MaxDemandAy3	OrtalamaGunlukTuketimAy3
ŞIRNAK	MaxDemandAy4	OrtalamaGunlukTuketimAy4
DİYARBAKIR	MaxDemandAy5	OrtalamaGunlukTuketimAy5
ŞANLIURFA	MaxDemandAy6	OrtalamaGunlukTuketimAy6
SİİRT	MaxDemandAy7	OrtalamaGunlukTuketimAy7
BATMAN	MaxDemandAy8	OrtalamaGunlukTuketimAy8
Köy	MaxDemandAy9	OrtalamaGunlukTuketimAy9
VIK	MaxDemandAy10	OrtalamaGunlukTuketimAy10
LUN	MaxDemandAy11	OrtalamaGunlukTuketimAy11
ELM	MaxDemandAy12	OrtalamaGunlukTuketimAy12
LNO	OrtalamaAkim2Ay1	OrtalamaAkim3Ay1
AEL	OrtalamaAkim2Ay11	OrtalamaAkim3Ay2
MSY	OrtalamaAkim2Ay12	OrtalamaAkim3Ay3
YOK	OrtalamaAkim2Ay2	OrtalamaAkim3Ay4
OrtalamaGerilim1Ay1	OrtalamaAkim2Ay3	OrtalamaAkim3Ay5
OrtalamaGerilim1Ay2	OrtalamaAkim2Ay4	OrtalamaAkim3Ay6
OrtalamaGerilim1Ay3	OrtalamaAkim2Ay5	OrtalamaAkim3Ay7
OrtalamaGerilim1Ay4	OrtalamaAkim2Ay6	OrtalamaAkim3Ay8
OrtalamaGerilim1Ay5	OrtalamaAkim2Ay7	OrtalamaAkim3Ay9
OrtalamaGerilim1Ay6	OrtalamaAkim2Ay8	OrtalamaAkim3Ay10
OrtalamaGerilim1Ay7	OrtalamaAkim2Ay9	OrtalamaAkim3Ay11
OrtalamaGerilim1Ay8	OrtalamaAkim2Ay10	OrtalamaAkim3Ay12

Ek 3: LR-BPSO ile Seçilen Öznitelikler (112 Öznitelik)**(Devamı)**

OrtalamaGerilim1Ay8	OrtalamaAkim2Ay10	OrtalamaAkim3Ay12
OrtalamaGerilim1Ay9	OrtalamaAkim1Ay1	OrtalamaAkim1Ay11
OrtalamaGerilim1Ay10	OrtalamaAkim1Ay2	OrtalamaAkim1Ay12
OrtalamaGerilim1Ay11	OrtalamaAkim1Ay3	OrtalamaGerilim2Ay7
OrtalamaGerilim1Ay12	OrtalamaAkim1Ay4	OrtalamaGerilim2Ay6
OrtalamaGerilim2Ay1	OrtalamaAkim1Ay5	OrtalamaAkim1Ay10
OrtalamaGerilim2Ay2	OrtalamaAkim1Ay6	OrtalamaAkim1Ay9
OrtalamaGerilim2Ay3	OrtalamaAkim1Ay7	OrtalamaGerilim2Ay5
OrtalamaGerilim2Ay4	OrtalamaAkim1Ay8	MARDİN
OrtalamaTuketim		

Ek 4: Uygulama Python Kodları

Çalışma kapsamında geliştirilen uygulama Python kodları, eğitim ve test veri paketleri, erişime açık bir şekilde <https://github.com/celikpence/KacakTakipSistemi> adresinde sunulmaktadır.



ÖZGEÇMİŞ

Kişisel Bilgiler

Adı-Soyadı: Mustafa ÇELİKPENÇE
Uyruğu: T.C.

Eğitim Bilgileri

Lisans: İstanbul Üniversitesi, Bilgisayar Mühendisliği, 2000

Yüksek Lisans: İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği, 2003

Yayınlar

Çelikpençe, M. (2017). Farklı Lokasyonlardaki RES ve HES SCADA Sistemlerinden Alınan Verilerin Merkezileştirilerek Ortak Platformdan Raporlanması. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 9(1): 19-28.
<https://dergipark.org.tr/tr/pub/tbbmd/issue/30338/327433>