

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



**KURAKLIK TAHMİNİNDE VERİ ÖN İŞLEME ADIMLARIYLA
GÜÇLENDİRİLMİŞ MAKİNE ÖĞRENMESİ YÖNTEMLERİ**

Serap ERÇEL

Yüksek Lisans Tezi

YAZILIM MÜHENDİSLİĞİ ANABİLİM DALI

ŞUBAT 2025

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Yazılım Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

**KURAKLIK TAHMİNİNDE VERİ ÖN İŞLEME ADIMLARIYLA
GÜÇLENDİRİLMİŞ MAKİNE ÖĞRENMESİ YÖNTEMLERİ**

Tez Yazarı
Serap ERÇEL

Danışman
Dr. Öğr. Üyesi Sinem AKYOL

ŞUBAT 2025
ELAZIĞ

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Yazılım Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Başlığı: Kuraklık Tahmininde Veri Ön İşleme Adımlarıyla Güçlendirilmiş Makine Öğrenmesi Yöntemleri

Yazarı: Serap ERÇEL

İlk Teslim Tarihi: 13.01.2025

Savunma Tarihi: 14.02.2025

TEZ ONAYI

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına göre hazırlanan bu tez aşağıda imzaları bulunan jüri üyeleri tarafından değerlendirilmiş ve akademik dinleyicilere açık yapılan savunma sonucunda OYBİRLİĞİ ile kabul edilmiştir.

Danışman:	Dr. Öğr. Üyesi Sinem AKYOL Fırat Üniversitesi, Mühendislik Fakültesi	<i>İmza</i> Onayladım
Başkan:	Doç. Dr. Bihter DAŞ Fırat Üniversitesi, Teknoloji Fakültesi	Onayladım
Üye:	Dr. Öğr. Üyesi Erkut ARICAN Bahçeşehir Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi	Onayladım

Bu tez, Enstitü Yönetim Kurulunun/...../20..... tarihli toplantısında tescillenmiştir.

İmza

Prof. Dr. Burhan ERGEN
Enstitü Müdürü

BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım “Kuraklık Tahmininde Veri Ön İşleme Adımlarıyla Güçlendirilmiş Makine Öğrenmesi Yöntemleri” Başlıklı Yüksek Lisans Tezimin içindeki bütün bilgilerin doğru olduğunu, bilgilerin üretilmesi ve sunulmasında bilimsel etik kurallarına uygun davrandığımı, kullandığım bütün kaynakları atıf yaparak belirttiğimi, maddi ve manevi desteği olan tüm kurum/kuruluş ve kişileri belirttiğimi, burada sunduğum veri ve bilgileri unvan almak amacıyla daha önce hiçbir şekilde kullanmadığımı beyan ederim.

14.02.2025

Serap ERÇEL



ÖNSÖZ

Bu çalışma, küresel ölçekte giderek artan bir tehdit olan kuraklık olgusunu anlamak, analiz etmek ve bu zorluğa bilimsel bir bakış açısı sunmak amacıyla hazırlanmıştır. Kuraklık, su kaynaklarının azalması, tarım verimliliğinin düşmesi ve ekosistem dengesinin bozulması gibi pek çok olumsuz etkiye neden olmaktadır. Bu çalışma, bu karmaşık doğa olayının kökenini anlamak ve gelecekteki mücadele stratejilerine bilimsel bir temel oluşturmak için önemli bir adım niteliğindedir.

Danışmanım Dr. Öğr. Üyesi SİNEM AKYOL'un bilge rehberliği ve değerli önerileri, bu çalışmanın kalitesini artırmamıza katkı sağlamıştır. Bütün bu katkıları ve desteği için danışmanıma minnettarlığımı sunarım. Ayrıca, hayatım boyunca her daim yanımda olup bana inanç ve destek veren canım aileme; sevgili meslektaşım ve hayat arkadaşıma sonsuz teşekkür ederim.

Serap ERÇEL
ELAZIĞ, 2025

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	iv
İÇİNDEKİLER	v
ÖZET	vii
ABSTRACT	viii
ŞEKİLLER LİSTESİ	ix
TABLolar LİSTESİ	x
1. Giriş	1
2. KURAKLIK.....	5
2.1. Meteorolojik Kuraklık Endeksleri	6
2.1.1. Standartlaştırılmış Yağış Endeksi (SPI)	6
2.1.2. Standartlaştırılmış Yağış-Buharlaştırma-Terleme endeksi (SPEI).....	7
2.1.3. Standart Streamflow Endeksi (SSFI).....	7
2.2. Kuraklığın Nedenleri	7
2.2.1. Yağış Eksiklikleri	7
2.2.2. Önceden Var Olan Meteorolojik Koşulların Rolü.....	7
2.2.3. Toprak Özellikleri	8
2.2.4. İklim Değişikliği	8
2.3. Türkiye’de Kuraklık	8
3. MATERYAL VE METOT	10
3.1. Materyal.....	10
3.1.1. Veri Seti	10
3.1.2. Ön İşleme Adımları	14
3.1.3. Veri Temizleme	15
3.1.4. Veri İşleme	15
3.2. Keşifsel Veri Analizi	16
3.2.1. Univariate Analiz	16
3.2.2. Aykırı Değer İşleme	18
3.2.3. Bivariate Analiz.....	20
3.2.4. Bağımsız Değişkenler Arasındaki Korelasyon.....	21
3.3. Metot	23
3.3.1. Veri Standardizasyonu	23
3.3.2. Özellik Seçimi	23
3.3.3. Sınıf Dengesizliği İle Başa Çıkma	24
3.3.4. Boyut Azaltma İşlemleri	24
3.4. Model Eğitimi ve Değerlendirmesi	25
3.4.1. Model Seçimi	25
3.4.2. Eğitim Süreci.....	25
4. BULGULAR.....	31
4.1. Meteorolojik Verilerden Hibrit Derin Öğrenme Modeli.....	34
4.2. Near Miss Downsampling ile Karar Ağacı.....	35
4.3. SMOTE Upsampling ile Karar Ağacı.....	35
4.4. Near Miss Downsampling ve PCA ile Karar Ağacı	36

4.5. SMOTE Upsampling ve PCA ile Karar Ağacı	36
4.6. Near Miss Downsampling ve LDA ile Karar Ağacı	37
4.7. SMOTE Upsampling ve LDA ile Karar Ağacı.....	37
4.8. Karar Ağacı (Herhangi Bir Dengesizlik İşlemi Uygulanmadan).....	38
4.9. Hiperparametre Ayarı ile Karar Ağacı (Herhangi Bir Dengesizlik İşlemi Uygulanmadan)	38
4.10.KNN (Herhangi Bir Dengesizlik İşlemi Uygulanmadan).....	39
4.11.Hiperparametre Ayarı İle KNN (Herhangi Bir Dengesizlik İşlemi Uygulanmadan).....	39
4.12.SMOTE Upsampling ile KNN.....	40
4.13.Near Miss Downsampling ile KNN.....	40
4.14.Naive Bayes (Herhangi Bir Dengesizlik İşlemi Uygulanmadan)	41
4.15.Random Forest (Herhangi Bir Dengesizlik İşlemi Uygulanmadan).....	41
4.16.Hiperparametre Ayarı İle Random Forest (Herhangi Bir Dengesizlik İşlemi Uygulanmadan)	42
5. TARTIŞMA	43
6. ÖNERİLER VE GELECEKTEKİ ÇALIŞMALAR	44
7. SONUÇLAR.....	45
KAYNAKLAR.....	47
ÖZGEÇMİŞ	

ÖZET

Kuraklık Tahmininde Veri Ön İşleme Adımlarıyla Güçlendirilmiş Makine Öğrenmesi Yöntemleri

Serap ERÇEL

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü

Yazılım Mühendisliği Anabilim Dalı

Şubat 2025, Sayfa: xi + 51

Kuraklık, dünya genelinde su kaynaklarının azalması, tarım verimliliğinin düşmesi ve ekosistem dengesinin bozulması gibi pek çok olumsuz etkiye neden olan karmaşık bir doğa olayıdır. Bu çalışma, kuraklık tespiti konusunda gelişmiş modellerin oluşturulmasını hedefleyerek, iklim değişiklikleriyle mücadelede bilimsel bir yaklaşım sunmayı amaçlamaktadır. Kuraklık üzerinde yapılan derinlemesine analiz, bilim dünyasına bu önemli konudaki anlayışımızı derinleştirmek ve sürdürülebilir çözümler geliştirmek adına önemli bir katkı sağlamayı amaçlamaktadır.

Çeşitli sınıf dengesizliği işlemleri ve boyut azaltma tekniklerinin uygulanmasıyla kuraklık tespiti modelleri uygulanmıştır. Temel odak, sentetik azınlık üzerine örnekleme tekniği ile örnekleme yapmanın ve Near Miss ile örnekleme yapmanın kuraklık tahmin modelinin performansına etkisini değerlendirmektir. Ayrıca, Temel Bileşen Analizi ve Lineer Diskriminant Analizi, özellik boyutlarını azaltmak ve tahmin modelinin etkinliğini artırmak için kullanılmaktadır.

Veri seti, sınıf dengesizliklerini ele almak için Sentetik Azınlık Üzerine Örnekleme Tekniği ve Near Miss yöntemleri kullanılarak ön işleme tabi tutulmaktadır. Daha sonra Temel Bileşen Analizi ve Lineer Diskriminant Analizi, özellik uzayını azaltmak ve kritik bilgileri korumak için uygulanmaktadır. Ön işlenmiş veri üzerinde makine öğrenmesi algoritmaları farklı kombinasyonlarla eğitilerek ve hibrit derin öğrenme modeli eğitilerek önerilen metodolojilerin etkinliği değerlendirilmektedir.

Sonuçlar, Sentetik Azınlık Üzerine Örnekleme Tekniği tabanlı örnekleme yönteminin özellikle doğruluk, hassasiyet, duyarlılık ve F1 skoru açısından model performansını önemli ölçüde artırdığını göstermektedir. Ayrıca, boyut azaltma tekniklerinin birleşiminin, kuraklık tahmin modellerinin genel güvenilirliğini artırmada etkili olduğunu göstermektedir.

Çalışmanın algoritmaları farklı kombinasyonlarla ve hazırlıklarla değerlendirmesine ek olarak bu algoritmaları hibrit kullanımıyla literatürden ayırmaktadır. Yapay zekanın etkin kullanımında gerçek dünya problemlerine karşı yaklaşımıyla çalışma öne çıkacaktır.

Bu çalışma, sınıf dengesizliği işlemleri ve boyut azaltma tekniklerinin sağlam kuraklık tespiti modellerinin geliştirilmesine katkı sağladığını göstermektedir. Bulgular, sınıf dengesizliklerini ele almak için uygun ön işleme adımlarının seçilmesi ve boyut azaltma tekniklerinin stratejik kullanımının, model etkinliğini artırmada önemli olduğunu vurgulamaktadır.

Anahtar Kelimeler: Kuraklık Tespiti, SMOTE, Near Miss, PCA-LDA, Hibrit Derin Öğrenme, Makine Öğrenmesi

ABSTRACT

Machine Learning Methods Enhanced With Data Preprocessing Steps for Drought Prediction

Serap ERÇEL

Master's Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences
Department of Software Engineering

February 2025, Pages: xi + 51

Drought is a complex natural phenomenon that causes numerous adverse effects worldwide, including the depletion of water resources, reduced agricultural productivity, and disruption of ecosystem balance. This study aims to develop advanced models for drought detection, offering a scientific approach to combating climate change. The in-depth analysis of drought seeks to enhance our understanding of this critical issue in the scientific community and contribute to the development of sustainable solutions.

Drought detection models have been implemented by applying various class imbalance processing techniques and dimensionality reduction methods. The primary focus is to evaluate the impact of the Synthetic Minority Over-sampling Technique and Near Miss sampling on the performance of the drought prediction model. Additionally, Principal Component Analysis and Linear Discriminant Analysis are employed to reduce feature dimensions and enhance model efficiency.

The dataset is preprocessed using the Synthetic Minority Over-sampling Technique and Near Miss methods to address class imbalances. Subsequently, Principal Component Analysis and Linear Discriminant Analysis are applied to reduce the feature space while preserving critical information. The effectiveness of the proposed methodologies is assessed by training machine learning algorithms in different combinations and developing a hybrid deep learning model on the preprocessed data.

The results indicate that SMOTE-based sampling significantly improves model performance, particularly in terms of accuracy, precision, recall, and F1 score. Furthermore, the combination of dimensionality reduction techniques proves to be effective in enhancing the overall reliability of drought prediction models.

In addition to evaluating algorithms with different combinations and preprocessing techniques, this study distinguishes itself from the existing literature by exploring the hybrid use of these algorithms. The study stands out in the field of artificial intelligence by addressing real-world problems through an effective methodological approach.

This research demonstrates that class imbalance processing and dimensionality reduction techniques contribute to the development of robust drought detection models. The findings emphasize the importance of selecting appropriate preprocessing steps to address class imbalances and strategically utilizing dimensionality reduction techniques to enhance model performance.

Keywords: Drought Detection, SMOTE, Near Miss, PCA-LDA, Hybrid Deep Learning, Machine Learning

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 3.1. Veri Dengesizliği.....	14
Şekil 3.2. Sürekli Değişkenlerin Dağılımı.....	18
Şekil 3.3. Aykırı Değerleri Belirleme.....	19
Şekil 3.4. Sıcaklık ve Özgül Nem Değişimi.....	20
Şekil 3.5. Çiğ Noktası Sıcaklığı ile Genel Sıcaklık Değişimi.....	20
Şekil 3.6. Rüzgar Hızının Farklı Uzaklıklarda Değişimi.....	21
Şekil 3.7. Özellik Seçimi İçin Bağımsız Değişkenler Arasındaki Korelasyon 1.....	22
Şekil 3.8. Özellik Seçimi İçin Bağımsız Değişkenler Arasındaki Korelasyon 2.....	22
Şekil 3.9. Kuraklık Tahmini Akışı.....	25
Şekil 4.1. Near Miss Downsampling ile Karar Ağacı için Çok Sınıflı ROC eğrisi.....	35
Şekil 4.2. SMOTE Upsampling ile Karar Ağacı için Çok Sınıflı ROC eğrisi.....	35
Şekil 4.3. Near Miss Downsampling ve PCA ile Karar Ağacı için Çok Sınıflı ROC eğrisi.....	36
Şekil 4.4. SMOTE Upsampling ve PCA ile Karar Ağacı için Çok Sınıflı ROC eğrisi.....	36
Şekil 4.5. Near Miss Downsampling ve LDA ile Karar Ağacı için Çok Sınıflı ROC eğrisi.....	37
Şekil 4.6. SMOTE Upsampling ve LDA ile Karar Ağacı için Çok Sınıflı ROC eğrisi.....	37
Şekil 4.7. Karar Ağacı için herhangi bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi.....	38
Şekil 4.8. Karar Ağacı Hiperparametre ayarı herhangi bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi.....	38
Şekil 4.9. KNN herhangi bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi.....	39
Şekil 4.10. KNN Hiperparametre ayarı ve bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi.....	39
Şekil 4.11. SMOTE Upsampling ile KNN için Çok Sınıflı ROC Eğrisi.....	40
Şekil 4.12. Near Miss Downsampling ile KNN için Çok Sınıflı ROC eğrisi.....	40
Şekil 4.13. Naive Bayes herhangi bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi.....	41
Şekil 4.14. Random Forest herhangi bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi.....	41
Şekil 4.15. Random Forest Hiperparametre ayarı ile herhangi bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi.....	42

TABLULAR LİSTESİ

	Sayfa
Tablo 3.1. Sınıfların Veri Sayısı	10
Tablo 3.2. Meteorolojik Göstergeler	11
Tablo 3.3. Veri Sayısı	12
Tablo 3.4. Kuraklık Kategorileri	13
Tablo 3.5. Tanımlayıcı İstatistikler	17
Tablo 4.1. Model Performans Kategorileri	31
Tablo 4.2. Kuraklık Tahmini: Upsampling Yaklaşımları	31
Tablo 4.3. Kuraklık Tahmini: Downsampling Stratejileri	32
Tablo 4.4. SPI Hibrit Modeli Sonuçları	33
Tablo 4.5. Yeniden Örnekleme Tekniklerinin Model Performansına Etkisi	33
Tablo 4.6. Algoritma Ailelerinin Performans Karşılaştırması	33
Tablo 4.7. Meteorolojik Hibrit Modelin Sonuçları	34

SİMGELER VE KISALTMALAR

Simgeler

m	: Metre
s	: Saniye
g	: Atmosfer Basıncı
kg	: Kilogram
C	: Sıcaklık
kPa	: Yüzey Basıncı

Kısaltmalar

SMOTE	: Sentetik Azınlık Aşırı Örnekleme Tekniği
PCA	: Temel Bileşen Analizi
LDA	: Doğrusal Diskriminant Analizi
SPI	: Standartlaştırılmış Yağış Endeksi
SPEI	: Standartlaştırılmış Evapotranspirasyon Endeksi
SSFI	: Standart Streamflow Endeksi
EDA	: Keşifsel Veri Analizi
DR	: Boyutsallık Azaltma
RFE	: Özyinelemeli Özellik Eliminasyonu

1. GİRİŞ

Dünya genelinde iklim değişikliği, su kaynakları, tarım ve ekosistemler üzerinde etkisini artırmaktadır. Özellikle kuraklık iklim olaylarına doğrudan sebep olan sorunları daha da derinleştirmektedir. Su kaynaklarının azalması, tarımın olumsuz etkilenmesi ve ekosistem dengesinin bozulması gibi sorunlar geri dönüşü olmayan zararlar meydana getirmektedir. Bu nedenle, kuraklığın önceden tahmin edilmesi, olumsuz etkilerle mücadele ve kaynak yönetimi gibi kritik konularda önemli bir stratejik konuma sahiptir.

Kuraklığın ekoloji, çevre, hidroloji ve tarım üzerindeki önemli etkileri nedeniyle kuraklık izleme ve erken uyarı sistemi geliştirme önemli araştırma alanları haline gelmiştir [1]. Tarımsal kuraklık afetlerinin risk değerlendirmeleri, kuraklığı niceliksel olarak anlamak ve önleme ve yardım çalışmalarına rehberlik etmek için gereklidir [2]. Ayrıca, kuraklık risk değerlendirmesi alanında, bu kritik alana artan ilgiyi yansıtan çok sayıda çalışma yapılmıştır [3]. Kuraklık azaltımı, kuraklık riskinin nicel değerlendirmesinin ayrılmaz bir parçasıdır ve etkileri kuraklık azaltımının yoğunlaştırılmış bir ifadesi olarak hizmet eder [4].

Kuraklıkların etkilerini önceden tahmin edebilmek, tarım alanında ürün kayıplarını minimize etme, su kaynakları üzerindeki baskıyı hafifletme ve ekosistemlerin sürdürülebilirliğini koruma açısından kritik bir öneme sahiptir. Kesin ve güvenilir kuraklık tahminleri, gerekli önlemlerin alınmasına yardımcı olarak tarım verimliliği, su kaynaklarının etkili yönetimi ve doğal ekosistemlerin muhafaza edilmesi adına stratejik kararlar alınmasını sağlamaktadır. Bu bağlamda, iklim değişikliği ile mücadelede ve sürdürülebilir kaynak yönetiminde bilimsel temelli kuraklık tahmin modellerine duyulan ihtiyaç daha da belirgin hale gelmiştir.

Kuraklık riski değerlendirmeleri, risk bilgisinin artırılmasına ve öncelikli yönetim alanlarının belirlenmesine yardımcı olduğu için proaktif ve uygun kuraklık riski yönetimi kararlarını bilgilendirmek için çok önemlidir [5]. Kuraklığın tarım üzerindeki etkisi önemlidir, iklim faktörleri yıl içi değişikliklere hakimdir ve doğru değerlendirme için kuraklık endekslerinin tarımsal anket verileriyle birleştirilmesinin önemini vurgulamaktadır [6]. Kuraklıkların insanların tüketim amaçlı kullanımları ve su ekosistemleri üzerinde geniş kapsamlı sonuçları vardır ve bu da ekolojik etkilerini ve insanların tepkilerini anlama ihtiyacını vurgulamaktadır [7]. Hayvancılık sektörü de dahil olmak üzere tarım, kuraklığın etkilerine karşı özellikle savunmasızdır ve tarımsal üretimde önemli kayıplara yol açmaktadır [8].

Kuraklık tahmini, makine öğrenimi tekniklerinin önemli bir potansiyel gösterdiği kritik bir alandır. Mishra ve Desai (2005), ARMA modellerinin, desen tanıma tekniklerinin ve fiziksel temelli

modellerin kuraklık tahmininde kullanılmasını tartışarak güvenilir tahminlerin önemini vurgulamışlardır [9]. Park ve arkadaşları (2020), kısa vadeli kuraklık tahmin modellerini ConvLSTM ve rastgele orman yaklaşımlarını kullanarak geliştirmiş ve makine öğreniminin zamansal desenleri yakalama konusundaki etkinliğini göstererek doğru tahminler sağlamıştır [10].

Sutanto ve arkadaşları (2019), hidro-meteorolojik kuraklık indekslerini rapor edilen etkilerle ilişkilendirerek kuraklık tehlikesinden etki tahminlerine geçişi makine öğrenimi kullanarak ele almış ve kuraklık etkilerinin tahmin edilebilirliğini göstermiştir. Bu çalışma, makine öğrenimi tekniklerinin sadece kuraklık tahmininde değil, aynı zamanda bu tahminlerin potansiyel etkilerini öngörmeye de kullanılabileceğini ortaya koymuştur [11]. Devanand (2024), Avustralya'da şiddetli bir kuraklığın makine öğrenimi aracılığıyla başarılı bir şekilde tahmin edilebileceğini vurgulayarak kuraklık tahmininin geliştirilme potansiyelini ortaya koymuştur [12]. Ayrıca, Dikshit ve arkadaşları (2021), uzun vadeli kuraklık tahminlerinde makine öğrenimi algoritmalarının faydalarını vurgulayarak bu tekniklerin etkinliğini desteklemiştir. Uzun vadeli tahminler, tarım planlaması ve su kaynakları yönetimi için kritik öneme sahiptir ve derin öğrenme teknikleri bu alanda önemli ilerlemeler sağlamıştır. [13].

Aghakouchak ve arkadaşları (2022), yapay zeka ve hibrit fiziksel-istatistiksel tahminlerin sunduğu fırsatları tartışarak makine öğrenimi ile doğru modeller geliştirme konusunda sağlanan ilerlemeleri vurgulamıştır [14]. Ayrıca, Aghelpour ve arkadaşları (2020), entropi teorisi ve makine öğrenimi yöntemlerini kullanarak farklı türde kuraklıkların aynı anda tahmin edilmesi için teorik bir yaklaşım önermiş ve makine öğreniminin çeşitli kuraklık tahminlerine yönelik esnekliğini ortaya koymuştur. Entropi teorisi, belirsizliği ölçme ve bilgi kazancını değerlendirmede etkilidir. Makine öğrenimi ile desteklendiğinde, kuraklık tahmininde yeni bir yaklaşım sağlar [15].

Khan ve arkadaşları (2021), Convolutional Neural Network (CNN) ve LSTM'yi birleştirerek hibrit modeller geliştirmişlerdir. Mekansal ve zamansal verileri aynı anda işleyerek daha yüksek doğruluk elde etmişlerdir. CNN, mekansal verileri iyi bir şekilde işlerken, LSTM, zamansal bağımlılıkları öğrenmede etkili olmuştur [16].

Zhang ve arkadaşları (2020), uzun vadeli kuraklık tahmini için derin öğrenme modelleri kullanarak, özellikle yıllık ve mevsimsel tahminlerde yüksek doğruluk oranları elde etmişlerdir. Bu modeller, büyük veri setlerinden öğrenerek iklim değişikliklerini ve bunların kuraklık üzerindeki etkilerini daha iyi anlamaktadır [17]. Benzer şekilde kuraklığın izlenmesinde ve Liu arkadaşları (2020) çalışmasında ilişkileri analiz ederek başarılı sonuçları almıştır [18].

Sonuç olarak, bu referansların sentezi, makine öğreniminin kuraklık tahmininde etkinliğine dair genel bir fikir birliği olduğunu ortaya koymaktadır. Kısa vadeli tahminlerden etki değerlendirmelerine ve farklı kuraklık türlerinin eşzamanlı tahminine kadar makine öğrenimi teknikleri, kuraklık tahminlerinin doğruluğunu ve güvenilirliğini artırmada değerli araçlar olduğunu kanıtlamıştır.

Bu literatür özetleri, kuraklık tahmininde kullanılan çeşitli makine öğrenmesi ve veri ön işleme tekniklerinin etkinliğini vurgulamaktadır. Özellikle sınıf dengesizliği ile başa çıkmak için kullanılan SMOTE ve boyut azaltma teknikleri (PCA, LDA), modellerin performansını artırmada kritik bir rol oynamaktadır. Bu çalışmalardan elde edilen bulgular, kuraklık tahmin modellerinin geliştirilmesi ve doğruluklarının artırılması için önemli bir temel oluşturmaktadır.

Bu çalışma, hava ve toprak verilerini kullanarak kuraklık tahminlerinde daha kesin ve doğru sonuçlar elde edilmesini hedeflemektedir. Meteorolojik veriler, toprak nem içeriği, iklim değişkenleri ve diğer faktörlerin bir araya getirilmesiyle oluşturulan modeller, gelecekteki kuraklık risklerini tahmin etmek için kullanılmaktadır.

Modellerin geliştirilmesi, kuraklıkla mücadelede daha etkili stratejiler oluşturulmasına katkı sağlamaktadır. Modelin doğruluğunun artırılması ve tahmin sürekliliğinin sağlanması, kuraklık risklerine karşı toplumların hazırlıklı olmasını ve sürdürülebilir bir gelecek için önlemlerin alınmasını desteklemektedir.

Elde edilen sonuçlar, tarım sektöründe ürün kayıplarının azaltılmasında, su kaynaklarının etkin bir şekilde yönetilmesinde ve ekosistemlerin korunmasında stratejik kararların alınmasına yardımcı olacaktır. Bu çalışma, iklim değişikliğine bağlı kuraklık risklerine karşı toplumları bilimsel bir temel üzerinde hazırlıklı hale getirmeyi amaçlamaktadır.

Bu çalışmanın özgünlüğü, kuraklık tespitine yönelik literatürde yaygın olarak kullanılan tek yönlü modellerin ötesine geçerek, veri dengesizliği ve yüksek boyutluluk gibi temel problemlere bütüncül bir yaklaşım getirmesidir. Çalışma, SMOTE ve Near Miss gibi gelişmiş örnekleme yöntemleri ile sınıf dengesizliğini giderirken, PCA ve LDA gibi boyut indirgeme tekniklerini entegre ederek verinin en kritik özelliklerini korumayı başarmıştır. Buna ek olarak, hibrit derin öğrenme mimarisi kullanılarak, zaman serisi verilerindeki karmaşık ilişkiler daha etkin bir şekilde modellenmiş ve geleneksel istatistiksel yöntemlere kıyasla daha üstün performans elde edilmiştir. Bu kapsamlı ve yenilikçi metodoloji, sadece model doğruluğunu artırmakla kalmayıp, kuraklık tahminine dair bilimsel anlayışı derinleştirerek, iklim değişikliğiyle mücadelede pratik ve uygulanabilir çözümler sunması bakımından literatüre önemli bir katkı sağlamaktadır.

Giriş bölümü, araştırmanın temel amacını ve kapsamını belirleyerek başlamaktadır. İlk olarak, kuraklık olgusunun ne olduğu ve neden bu konunun önemli olduğu açıklanmaktadır. Ardından, ikinci bölümde meteorolojik kuraklık endekslerine odaklanılarak, Standartlaştırılmış Yağış Endeksi (SPI), Standartlaştırılmış Yağış-Buharlaşma-Terleme endeksi (SPEI) ve Standart Streamflow Endeksi (SSFI) gibi önemli kavramlar detaylı bir şekilde ele alınmıştır. Üçüncü bölüm, Türkiye'deki kuraklık durumunu ele alarak, bu bağlamda literatürdeki mevcut bilgileri sunmaktadır. Ardından, dördüncü bölümde kullanılan metodoloji, veri seti, ön işleme adımları, keşifsel veri analizi ve model eğitimi adımları detaylı bir şekilde açıklanmaktadır. Beşinci bölümde ise, farklı dengeleme teknikleri kullanılarak yapılan modelleme çalışmalarının bulguları sunulmuş ve önerilerde bulunulmuştur. Son olarak, Sonuç ve Tartışma bölümü, elde edilen bulguların özetlenmesi, literatürle karşılaştırılması ve çalışmanın genel katkılarına vurgu yaparak sona ermektedir.

2. KURAKLIK

Kuraklık, su kaynaklarının yetersiz olduğu veya yağış miktarının normalden uzun süre boyunca azaldığı bir iklim olayıdır. Kuraklık genellikle yavaş gelişen bir süreçtir ve uzun bir zaman dilimini kapsar. Bu süreç boyunca, bitki ve hayvanlar kuraklık nedeniyle su kıtlığından ve düşük nem seviyelerinden olumsuz etkilenirler.

Kuraklığın tespit edilmesi gerçekten zor bir süreçtir. Çünkü kuraklık, yağış miktarı ve su kaynaklarındaki değişiklikler gibi birçok faktörün bir arada değerlendirilmesini gerektirir. Bu yüzden, kuraklık tahminleri ve izleme sistemleri geliştirilmekte ve meteorolojik veriler, su kaynakları, bitki örtüsü ve toprak nem verileri gibi bilgilerin entegrasyonunu içeren modeller kullanılmaktadır. Bu şekilde, kuraklığın tespit edilmesi ve önceden tahmin edilmesi daha sağlam ve güvenilir bir şekilde gerçekleştirilmeye çalışılmaktadır.

Literatürde genellikle üç temel kuraklık türü tanımlanmaktadır: meteorolojik, tarımsal ve hidrolojik kuraklık [19]. Bunun yanı sıra, kuraklığın ekonomik ürünlerin arz ve talebini olumsuz etkilemesiyle ortaya çıkan ve bazı kaynaklarda "sosyo-ekonomik kuraklık" olarak adlandırılan bir kavram da mevcuttur [20]. Meteorolojik kuraklık, belirli bir zaman diliminde ortalama yağış miktarının azalmasıyla ortaya çıkar. Eğer bu kuraklık uzun süre devam ederse, toprak neminin azalması sonucu bitkiler fizyolojik su ihtiyaçlarını karşılayamaz hale gelir ve bu durum tarımsal kuraklık olarak tanımlanır. Yağış eksikliği daha da belirgin hale geldiğinde, toprak neminin yanı sıra yüzey ve yeraltı su kaynaklarında da ciddi azalmalar meydana gelir ve hidrolojik kuraklık ortaya çıkar. Sosyo-ekonomik kuraklık ise bu üç kuraklık türüyle doğrudan ilişkilidir ve su kaynaklarının talebi karşılama yetersiz kalması ya da insan sağlığı üzerindeki olumsuz etkileriyle kendini gösterir [19]. Tüm kuraklık süreçleri meteorolojik kuraklıkla başlar ve zamanla diğer kuraklık türlerine dönüşebilir.

1951-2010 yılları arasında küresel ölçekte yapılan bir araştırma, kuraklığın sıklığı, süresi ve şiddetini incelemiştir. Bulgulara göre, 1951-1970 yılları arasında Kuzey Amerika ve Avustralya, 1971-1990 döneminde Ekvatorial bölge, 1991-2010 yılları arasında ise Akdeniz Havzası kuraklıktan en çok etkilenen bölgeler olmuştur [21]. Doğu Akdeniz Havzası ve Türkiye için yapılan çalışmalar değerlendirildiğinde özellikle 1970'li yıllardan sonra toplam yağış miktarlarında azalma ve kurak koşulların etkisinde artış ve hatta çölleşmeye eğilimli alanların varlığı ortaya konmuştur [22]. Ceylan vd. çalışmasında Türkiye'de çölleşmeye meyilli alanların Konya Ovasının doğusundan Doğu Akdeniz bölümüne doğru ilerlediği ifade etmiştir.

2.1. Meteorolojik Kuraklık Endeksleri

Meteorolojik kuraklık endeksleri, atmosferdeki nem durumunu ve yağış eksikliğini değerlendirmek için kullanılan ölçüm ve değerlendirme sistemleridir. Bu endeksler genellikle tarım, su kaynakları yönetimi, orman yangınları ve su talepleri gibi çeşitli sektörlerde kullanılarak kuraklık durumlarını izlemeye ve değerlendirmeye yardımcı olurlar.

2.1.1. Standartlaştırılmış Yağış Endeksi (SPI)

Standartlaştırılmış yağış endeksi (SPI), çeşitli zaman ölçeklerinde meteorolojik kuraklığı karakterize etmek için yaygın olarak kullanılan bir indekstir. Kısa zaman ölçeklerinde SPI toprak nemi ile yakından ilişkiliyken, daha uzun zaman ölçeklerinde SPI yeraltı suyu ve rezervuar depolamasıyla ilişkilendirilmektedir. SPI, önemli ölçüde farklı iklimlere sahip bölgeler arasında karşılaştırılabilir. Gözlemlenen yağışı, ham yağış verilerini modelleyen seçilmiş bir olasılık dağılım fonksiyonundan standartlaştırılmış bir sapma olarak ölçer. Ham yağış verileri tipik olarak bir gama veya Pearson tip III dağılımına uyarlanır ve daha sonra normal bir dağılıma dönüştürülür. SPI değerleri, gözlemlenen anomalinin uzun vadeli ortalamadan saptığı standart sapmaların sayısı olarak yorumlanabilir. SPI, aylık yağış verileri kullanılarak 1-36 aylık farklı dönemler için hesaplanabilir. Operasyonel toplulukta SPI, meteorolojik kuraklığın ölçülmesi ve raporlanması için dünya çapında mevcut olması gereken standart endeks olarak kabul edilmiştir. Evapotranspirasyondaki değişikliklerle ilgilenmediğinden, SPI'nin iklim değişikliğiyle ilişkili kuraklıktaki değişikliklerin bir ölçüsü olarak kullanılması konusunda endişeler ortaya çıkmıştır. Buharlaşma ve terlemeyle ilgilenen alternatif endeksler önerilmiştir. Sonuçta Gama dağılımına dayalı kümülatif olasılıkların standart normal dağılıma dönüştürülmesiyle SPI değeri hesaplanır [23].

SPI şu şekilde hesaplanabilir:

$$\text{SPI} = s \frac{t - (c_2 t + c_1)t + c_0}{((d_3 t + d_2) + d_1) + 1} t = \sqrt{\ln \frac{1}{H(X)^2}} \quad (2.1)$$
$$G(x) = \frac{1}{\beta^\gamma \Gamma(\gamma)_0} \int_0^x X^{\gamma-1} e^{-\frac{x}{\beta}} dx, X > 0(\gamma) = \int_0^\infty X^{\gamma-1} e^{-x} dx$$

Burada x yağış değerini belirtir;

b ve c, C fonksiyonunun ölçek ve şekil parametrelerini temsil eder;

S pozitif ve negatif katsayılarıdır ve c0, c1, c2 ve d1, d2, d3 hesaplanan parametrelerdir [24].

2.1.2. Standartlaştırılmış Yağış-Buharlaşıma-Terleme endeksi (SPEI)

Standartlaştırılmış Yağış Endeksi (SPI) ve Standartlaştırılmış Yağış Evapotranspirasyon Endeksi (SPEI), kuraklık koşullarını değerlendirmek için yaygın olarak kullanılan kuraklık endeksleridir. Her iki endeks de kuraklık olaylarını izlemek ve analiz etmek için çeşitli coğrafi bölgelerde uygulanmıştır. Çalışmalar, özellikle kuraklığa bağlı etkilerin daha belirgin olduğu yaz gibi kritik mevsimlerde, SPEI'nin kuraklığa verilen tepkileri yakalamada SPI'den daha iyi performans gösterdiğini göstermiştir [25].

2.1.3. Standart Streamflow Endeksi (SSFI)

Araştırmacılar, meteorolojik kuraklık indekslerine benzer standart hidrolojik kuraklık indeksleri geliştirdiler. Standart akarsu akış endeksi ve akarsu akış kuraklık endeksini içeren bilinen iki hidrolojik kuraklık endeksi aynı teorik altyapıya sahiptir. Her iki endeks de aylık akışı standart normal dağılıma (SPI yöntemine benzer şekilde ortalama sıfır ve birin varyansı) dönüştürmeye ve hidrolojik kuraklık endeksini hesaplamaya çalışır. SSFI parametrik olmayan yaklaşıma göre hesaplanır. Dolayısıyla bu endeksi hesaplamak için aynı hesaplama işlemi tekrarlanır; tek fark yağış verileri yerine akış verilerinin girdi verileri olarak kullanılmasıdır. Bu indeks için aynı SPI gibi gama dağılımı uygundur. Bu yöntem basit ve SPI yöntemine benzer. Her aya ait kümülatif akım değerleri ayrı ayrı değerlendirilerek her ay için SGK hesaplanır [24].

2.2. Kuraklığın Nedenleri

Genellikle iklimsel değişimler ve meteorolojik etmenlerle ilişkili olup atmosferik koşulların kompleks etkileşimi sonucu meydana gelen kuraklığın temel nedenleri yağış, evapotranspirasyon etkisi, toprak özellikleri gibi sıralanabilir.

2.2.1. Yağış Eksiklikleri

Kuraklık olaylarının ana nedeni, belirli bir coğrafi alanda beklenen yağışların olmaması veya önemli ölçüde azalmasıdır. Yetersiz yağış, su kaynaklarını doğrudan etkileyerek toprak neminde hızlı bir düşüşe yol açar. Kuraklığın nedenleri çok yönlüdür ve çeşitli faktörlere bağlanabilir. Yetersiz yağışla karakterize edilen meteorolojik kuraklık, kuraklığın birincil nedenidir ve su arzı açıkları nedeniyle daha sonra tarımsal kuraklığa yol açar [26].

2.2.2. Önceden Var Olan Meteorolojik Koşulların Rolü

Kuraklık genellikle uzun süreli meteorolojik koşulların bir araya gelmesi sonucu gelişir. Alçak basınç sistemleri, uzun süreli yüksek sıcaklıklar ve düzensiz yağış dağılımı gibi faktörler bölgesel

kuraklık oluşumları için gerekli koşulları yaratır. Normalin üzerindeki hava sıcaklıkları gibi bölgesel atmosferik anomaliler, toprak nemi veya hidrolojik kuraklık gelişimine katkıda bulunarak kuraklık etkilerini artırabilir [26].

2.2.3. Toprak Özellikleri

Su emme kapasitesi, tutma ve iletim özellikleri de dahil olmak üzere toprak özellikleri, kuraklık süreçlerini etkileyen önemli faktörlerdir. Düşük geçirgenliğe sahip topraklar yağışın verimli bir şekilde tutulmasını engelleyerek kuraklığın şiddetini artırır.

2.2.4. İklim Değişikliği

Atmosferdeki sera gazı birikiminin yol açtığı küresel iklim değişikliği, kuraklık riskini artırma potansiyeline sahiptir. Artan sıcaklıklar, düzensiz yağış modelleri ve değişen iklim dinamikleri kuraklık olaylarının sıklığını ve yoğunluğunu etkilemektedir. Orman ölümleri ve yaprak alanının azalması gibi büyük ölçekli bozulmalar, sürekli karbondioksit akışına katkıda bulunabilir, net ekosistem değişimini ve karbon akışlarını bozarak hidrolojik döngüyü etkilemektedir.

2.3. Türkiye’de Kuraklık

Türkiye, iklim sınıflandırmasına göre yarı kurak bir bölgede yer almaktadır. Bu durum, kuraklığın düzenli olarak izlenmesini ve gerekli önlemlerin önceden alınmasını zorunlu kılmaktadır. Geniş alanları etkisi altına alabilen kuraklık, açlık, kıtlık ve işsizlik gibi ciddi toplumsal sorunlara yol açarak kalıcı etkiler bırakabilen meteorolojik kökenli bir doğal afettir. Ancak, kuraklık düzenli olarak takip edildiğinde, etkileri minimize edilebilir ve alansal ile zamansal değişimleri belirlenerek gerekli müdahaleler zamanında gerçekleştirilebilir [27].

Türkiye’nin yarı kurak iklim kuşağında bulunması, kuraklığın ülkemiz için sürekli bir tehdit unsuru olduğunu ortaya koymaktadır. Bu tehdidin yönetilebilmesi için en önemli adımlardan biri, kuraklığın sürekli olarak izlenmesi ve analiz edilmesidir. Bu bağlamda, farklı disiplinlerden uzmanların yer alacağı Ulusal Kuraklık Önleme Merkezi’nin acilen kurulması gerekmektedir. Böyle bir merkez, kuraklıkla mücadelede bilimsel ve stratejik yaklaşımlar geliştirerek ülke genelinde daha etkin önlemler alınmasını sağlayacaktır [27].

İklim etmenleri ve bitki örtüsü dikkate alındığında, Türkiye’deki çölleşmeye eğilimli kurak araziler, karasal iç ve doğu bölgelerinin önemli bir bölümünü ve Güneydoğu Anadolu Bölgesini içerir. Akdeniz ve Ege bölgelerinin geniş bir bölümü ise, yüksek ve parçalı yeryüzü şekilleri, tarım

arazilerinin yaklaşık son 40 yıldaki tarım dışı ve sürdürülebilir olmayan kullanımı, kentsel ve turizm getirisi yüksek olan tarım ve orman arazilerinin tarım ve orman rejimi dışına çıkarılmasına yönelik girişimler ve yasal düzenlemeler, sanayi, turizm ve orman yangınları gibi iyi bilinen öteki doğal ve insan kaynaklı etmenler dikkate alındığında, gelecekte çölleşme süreçlerinden daha fazla

etkilenebilecek yarı nemli alanlar olarak kabul edilmelidir. Uzun süreli ve şiddetli yaz kuraklıklarının ve yüksek hava sıcaklıklarının yanı sıra, yağış ve kuraklık indisi dizilerinde gözlenen kurak koşullara yönelik değişme eğilimlerinin, Akdeniz ve Ege bölgelerinde iklim etmenlerinin çölleştirme kuvvetini arttırmakta oluşu, bu düşünceyi destekler niteliktedir [28].

Kuraklığın Türkiye'deki etkisini anlamak için, ülkenin küresel ısınmaya karşı hassasiyetini ve kuraklık olaylarının artan sıklık ve şiddetini dikkate almak büyük önem taşıyor. Türkiye yarı kurak bir bölgede yer aldığından kuraklığa karşı özellikle hassastır [29]. Araştırmalar, Türkiye'de kuraklığın etkilerinin hem mekânsal hem de zamansal olarak ekonomik, çevresel ve sosyal açılardan farklılık gösterebileceğini göstermiştir [30]. Palmer Kuraklık Şiddet Endeksi (PDSI), Türkiye'de kuraklığın etkilerini izlemek ve azaltmak için yararlı bir araç olarak önerilmiştir [31], [32]. Ayrıca Türkiye'de yaşanan kuraklığın, insanların alternatif yaşam koşulları aramaya zorlanması nedeniyle göçlere yol açabileceği ileri sürülmektedir [33]. Türkiye'nin de içinde bulunduğu Doğu Akdeniz havzası, düşük yağışlardan kaynaklanan kuraklık olayları nedeniyle çeşitli sorunlarla karşı karşıyadır [34]. Kuraklığın etkisi su kaynaklarına da uzanıyor ve çalışmalar belirli bölgelerde hidrolojik kuraklık eğilimlerinin arttığını gösteriyor [35]. Küresel ısınmanın bir sonucu olarak Türkiye'de kuraklığın ekolojik bozulmalara, çölleşmeye ve su kaynaklarının zayıflamasına yol açma potansiyeli vurgulanmıştır [36]. Türkiye'de kuraklığın tarım, ormancılık, su yönetimi gibi çeşitli sektörleri etkileyen çok yönlü bir sorun olduğu açıktır [37]. Özellikle azalan su kaynakları ve ekolojik bozulma potansiyeli karşısında etkili su yönetimi stratejilerine duyulan ihtiyaç, ülkenin kuraklığa karşı duyarlılığını daha da vurgulamaktadır [38].

3. MATERYAL VE METOT

Bu çalışmada, kuraklık tespiti amacıyla geniş veri setleri kullanılarak çeşitli makine öğrenimi yöntemleri uygulanmıştır. Veri ön işleme sürecinde sınıf dengesizliklerini giderme ve boyut indirgeme teknikleriyle verinin kalitesi artırılmış; ardından, farklı algoritmalar eğitilerek model performansı uygun metriklerle değerlendirilmiştir.

3.1. Materyal

Meteorolojik göstergeler, hava ve toprak verilerinin ölçümlerini içerir. Birçok alanda kullanılan bu veriler kuraklık tahmininde de etkili bir kullanım sağlamaktadır. Rüzgar hızı, sıcaklık, nem, basınç, yağış ve diğer meteorolojik değerler veri setinde yer almaktadır. ABD Kuraklık Monitörü'nden alınan veriler uzmanlar tarafından geniş bir veri yelpazesi kullanılarak manuel olarak oluşturulan kuraklığın analizini sağlamaktadır.

3.1.1. Veri Seti

Bu veri setinin amacı, kuraklıkların yalnızca meteorolojik veriler kullanılarak tahmin edilemeyeceğinin araştırılmasına yardımcı olmak ve potansiyel olarak ABD tahminlerinin dünyanın diğer bölgelerine genelleştirilmesine yol açmaktır. İlk olarak U.S [Global Drought Monitor](#)'den alınan, paylaşılmış veri seti kullanılmıştır.

Veri setinde olmasa dahi kuraklık olmayan durumu ve altı farklı kuraklık düzeyini içeren bir sınıflandırma modeli için kullanılmak üzere oluşturulmuştur. Her bir giriş, belirli bir Amerika Birleşik Devletleri eyaletinde belirli bir zaman noktasındaki kuraklık seviyesini temsil eder ve bu seviyeler, aşağıda belirtilen 18 meteorolojik göstergeye dayanmaktadır. Girişler, gözlem süresi olarak son 90 günü içerir ve her bir gösterge, ilgili eyaletteki meteorolojik koşulları ifade eder. Bu göstergeler, kuraklık düzeylerini belirlemek ve sınıflandırmak için kullanılan temel veri noktalarını oluşturur. Bu veri setinde 26 özellik bulunmaktadır. Kuraklık seviyeleri 0 ile 4 arasında değişen 5 sınıf olarak etiketlenmiştir. Bu sınıfların dağılımı Tablo 3.1'de belirtildiği üzeredir.

Tablo 3.1. Sınıfların Veri Sayısı

Sınıf	Veri Sayısı
0	1652230
1	466944
2	295331
3	196802

4	106265
5	39224

Veri setinden değerlendirilen sütunlar Tablo 3.2. ile gösterilmiştir.

Tablo 3.2. Meteorolojik Göstergeler

Gösterge	Tanım
WS10M_MIN	10 Metrede Minimum Rüzgar Hızı (<i>m/s</i>)
QV2M	2 Metrede Özgül Nem (<i>g/kg</i>)
T2M_RANGE	2 Metrede Sıcaklık Aralığı (<i>C</i>)
WS10M	10 Metrede Rüzgar Hızı (<i>m/s</i>)
T2M	2 Metrede Sıcaklık (<i>C</i>)
WS50M_MIN	50 Metrede Minimum Rüzgar Hızı (<i>m/s</i>)
T2M_MAX	2 Metrede Maksimum Sıcaklık (<i>C</i>)
WS50M	50 Metrede Rüzgar Hızı (<i>m/s</i>)
TS	Dünya Cilt Sıcaklığı (<i>C</i>)
WS50M_RANGE	50 Metrede Rüzgar Hızı Aralığı (<i>m/s</i>)
WS50M_MAX	50 Metrede Maksimum Rüzgar Hızı (<i>m/s</i>)
WS10M_MAX	10 Metrede Maksimum Rüzgar Hızı (<i>m/s</i>)
WS10M_RANGE	10 Metrede Rüzgar Hızı Aralığı (<i>m/s</i>)
PS	Yüzey Basıncı (<i>kPa</i>)
T2MDEW	2 Metrede Çiy/Don Noktası (<i>C</i>)
T2M_MIN	2 Metrede Minimum Sıcaklık (<i>C</i>)
T2MWET	2 Metrede Yaş Termometre Sıcaklığı (<i>C</i>)
PREKTOT	Yağış (<i>mm gün-1</i>)

Bu meteorolojik göstergelerin kombinasyonu, her bir girişin kuraklık durumunu objektif bir şekilde değerlendirmek için kullanılır. Sınıflandırma modeli, bu göstergelerin karmaşıklığını analiz ederek ilçedeki kuraklık düzeyini belirler. Veri kümesi, bu bilgileri içererek belirli bir zamandaki meteorolojik koşulların kuraklık açısından değerlendirilmesi için bir temel sağlamaktadır.

Kuraklık Tahmini İçin SPEI Global Drought Monitor Veri Seti

Bu çalışma kapsamında kullanılan ikinci veri seti SPEI veri setidir. SPEI [Global Drought Monitor](#)'den alınan, paylaşılmış veri seti, Ocak 1950 - Aralık 2010 arasındaki kalibrasyon periyoduna dayanarak hesaplanmış ve 1955 yılından itibaren tutarlı bilgi sağlamak amacıyla oluşturulmuştur. Veriler her ayın ilk günlerinde güncellenmekte ve en güncel iklim verilerine dayanmaktadır. Kuraklık değerlendirmesi için kullanılan SPEI endeksi, Thorthwaite yöntemi ile potansiyel evapotranspirasyon (PET) tahmini yapmaktadır. Tablo 3.3'de veri dağılımı

görülmektedir. Bu tabloda, veri kümesinin 360 boylamsal, 180 enlemsel konumda ve 1332 zaman noktasında ölçülen toplam 86,313,600 veri noktası içerdiği görülmektedir.

Tablo 3.3. Veri Sayısı

Özellik	Değer
Uzunluk (Longitude) boyutları	-179.75° ile 179.75° arasında, 1° çözünürlükte (toplam 360 nokta)
Enlem (Latitude) boyutları	-89.75° ile 89.75° arasında, 1° çözünürlükte (toplam 180 nokta)
Zaman boyutu	Ocak 1901 - Aralık 2011 (toplam 1332 ay)
Toplam grid hücre sayısı	$360 \times 180 = 64,800$ konum
Toplam veri noktası sayısı	$64,800 \text{ konum} \times 1332 \text{ ay} = 86,313,600$ veri noktası
SPEI değer aralığı	-3.0E32 ile 3.0E32 arasında (float)
Veri birimi	z değerleri

Aşağıdaki boyutlar ve değişkenlerden oluşmaktadır:

- **Boyutlar:**
 - **Uzunluk (Longitude):** -179.75° ile 179.75° arasında değişen 1° çözünürlükte değerler.
 - **Enlem (Latitude):** -89.75° ile 89.75° arasında değişen 1° çözünürlükte değerler.
 - **Zaman (Time):** Ocak 1901 ile Aralık 2011 arasında aylık zaman ölçeği.
- **Değişkenler:**
 - **SPEI:** Standart Yağış-Evapotranspirasyon İndeksi (SPEI), "z değerleri" cinsinden ölçülmektedir ve -3.0E32 ile 3.0E32 arasında değişen float veri tipine sahiptir.
 - **Uzunluk ve Enlem Değerleri:** Derece (degrees_E, degrees_N) cinsinden ifade edilmiştir.
 - **Zaman:** "1901-01'den itibaren aylar" biriminde ve Gregoryen takvimi kullanılarak hesaplanmıştır.

Kuraklığın Kategorilendirilmesi

Kuraklık durumlarını ve olası etkilerini sınıflandırılmıştır. D0'dan D4'e kadar olan kategoriler, "Anormal Kurak"tan "Olağanüstü Kurak" durumlarına kadar değişen şiddetteki kuraklık koşullarını tanımlar. Her kategori, tarım ve su kaynakları üzerindeki spesifik etkileri açıklar. Tablo 3.4, kuraklık kategorileri ve olası etkileri hakkında ayrıntılı bilgi vermektedir.

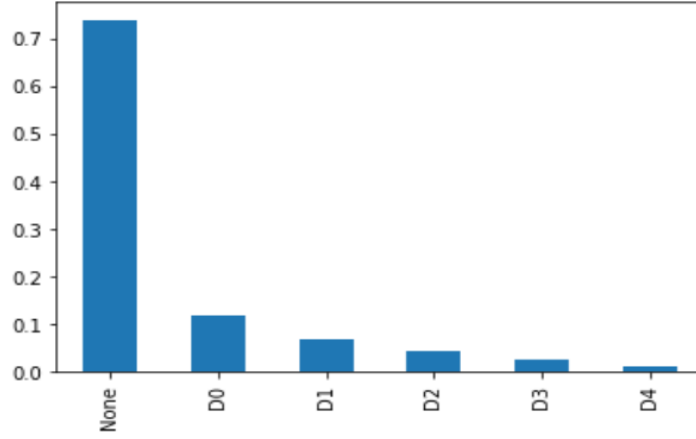
Tablo 3.4. Kuraklık Kategorileri

Kategori	Açıklama	Olası Etkileri
D0	Anormal Kuruluk	- Kısa süreli kuraklık, bitki yetiştirme veya otlakların büyümesini yavaşlatır. - Kuraklıktan çıkarken bazı kalıcı su eksiklikleri olabilir. - Otlaklar veya bitkiler tamamen iyileşmez.
D1	Orta Kurak	Bazı bitki hasarları, otlaklar - Akarsular, rezervuarlar veya kuyular düşük, bazı su kıtlıkları geliyor veya yakın zamanda oluşacak. - Gönüllü su kullanımı kısıtlamaları istendi.
D2	Şiddetli Kurak	- Tarım veya otlak kayıpları muhtemeldir. - Su kıtlıkları yaygındır. - Su kısıtlamaları uygulanır.
D3	Aşırı Kurak	- Önemli bitki/otlak kayıpları - Yaygın su kıtlıkları veya kısıtlamaları
D4	Olağanüstü Kurak	- Olağanüstü ve yaygın bitki/otlak kayıpları - Rezervuarlardaki su kıtlıkları, akarsular ve kuyulardaki su kıtlıkları acil durumlar yaratıyor.

Veri Dengesi

Dengesizlik, farklı sınıflardaki örnek sayılarının belirgin bir şekilde farklı olduğu durumu ifade etmektedir. Bu durum, analiz veya sınıflandırma modelleri geliştirilirken belirli bir sınıftaki örneklerin diğer sınıflara göre daha baskın olduğunu veya az olduğunu göstermektedir. Dengesiz veri setleri, modelin eğitimini etkilemektedir ve genellikle daha fazla örnek içeren sınıfların daha iyi performans göstermesine neden olmaktadır.

Dengesizlik durumuyla başa çıkmak için örnekleme teknikleri veya farklı sınıflardan gelen veri noktalarının ağırlıklarını düzenleme gibi çeşitli yöntemler kullanılarak, modelin dengesizlikten etkilenme olasılığı azaltılabilir. Bu tür dengesizlik analizleri, bilgisayar bilimleri, veri madenciliği ve makine öğrenimi gibi alanlarda önemli bir konudur. Şekil 3.1'deki grafik incelendiğinde, veri setinin dengesiz olduğu gözlemlenmektedir. Bu durum ön işleme adımlarının ve keşifsel veri analizinin kritik öneme sahip olduğunu göstermektedir.



Şekil 3.1. Veri Dengesizliği

3.1.2. Ön İşleme Adımları

Verileri etkili bir şekilde temizlemek ve ön işlemek için, verilerin doğruluğunu ve güvenilirliğini sağlamak için sistematik bir yaklaşım izlemek esastır. Süreç, eksik değerlerin işlenmesi, kopyaların kaldırılması, veri biçimlerinin standartlaştırılması ve aykırı değerlerle başa çıkma dahil olmak üzere çeşitli adımları içerir.

Eksik Değerlerin İşlenmesi

İstatistiksel yöntemler veya görselleştirme teknikleri kullanarak veri kümesindeki eksik değerler belirlenir. Ortalama, medyan veya mod değerlerini kullanarak atama veya daha karmaşık veri kümeleri için çoklu atama gibi gelişmiş teknikler gibi eksik verileri işlemek için en uygun yönteme karar verilir.

Kopyaları Kaldırma

Fazlalığı önlemek ve veri bütünlüğünü sağlamak için veri kümesindeki yinelenen kayıtlar belirlenir ve kaldırılır. Belirli anahtar sütunlara veya niteliklere dayalı olarak yinelenen girişleri belirlemek ve ortadan kaldırmak için programlama araçlarını veya işlevlerini kullanılır.

Veri Formatlarının Standartlaştırılması

Tutarlılığı sağlamak ve analizi kolaylaştırmak için tarihler, saatler ve kategorik değişkenler gibi veri alanlarının biçimi standartlaştırılır. Tarihleri standart bir biçime (ör. YYYY-AA-GG) dönüştürmek ve tek tip ölçü birimleri sağlamak gibi veriler tutarlı bir biçime dönüştürülür.

Aykırı Değerlerle Başa Çıkma

Z-skorları, kutu grafikleri veya dağılım grafikleri gibi istatistiksel yöntemler kullanarak aykırı değerleri belirlenir. Veri girişi hatalarıysa bunları kaldırmak, verileri dönüştürmek veya aykırı

değerlere karşı daha az duyarlı olan sağlam istatistiksel teknikler kullanmak gibi aykırı değerleri ele almak için en uygun yaklaşıma karar verilir.

Kategorik Değişkenlerin Kodlanması

Verileri makine öğrenimi algoritmalarına uygun hale getirmek için tek yönlü kodlama, etiket kodlama veya ikili kodlama gibi teknikleri kullanarak kategorik değişkenleri sayısal bir biçime dönüştürün.

Ölçeklendirme ve Normalleştirme

Sayısal özellikler standart bir aralığa ölçeklendirilir ve özellikle makine öğrenimi modellerinde tüm değişkenlerin analize eşit şekilde katkıda bulunmasını sağlamak için bunlar normalleştirilir.

Özellik Seçimi

Boyutsallığı azaltmak ve model performansını artırmak için alan bilgisine, istatistiksel testlere veya özellik önemi tekniklerine dayalı olarak ilgili özellikler seçilir.

3.1.3. Veri Temizleme

Bu çalışmada kuraklık tahmini için meteorolojik veri setinin temizlenmesi aşamasında aşağıdaki adımlar atılmıştır:

Çiftler, Tutarsızlıklar ve Anomalilerin İncelenmesi

Veri setindeki çiftler, tutarsızlıklar ve diğer anormal durumlar gözden geçirilmiştir. Bu adım, veri setinin güvenilirliğini artırmak ve yanıltıcı bilgileri engellemek için önemlidir.

Hedef Değişkenindeki Null Değerlerin İşlenmesi

Hedef değişken olan kuraklık skoru yalnızca haftada bir ölçüldüğünden, bu değerlerin eksik olması anlamlıdır. Null değerler, analizin doğruluğunu etkileyebileceği için bu durum ele alınmıştır.

Veri Setinin Genel Yapısına Yönelik Düzenlemeler

Veri setinin genel yapısında gerekli görülen düzenlemeler yapılmıştır. Bu, veri setini analiz için daha uygun hale getirme amacını taşır.

3.1.4. Veri İşleme

Çeşitli veri işleme adımları uygulanarak veri seti optimize edilmiştir. Aşağıda adım adım anlatılmıştır.

Eksik Veri İşleme

Hedef değişken olan kuraklık skorundaki eksik değerler, skorların sadece haftada bir ölçüldüğü göz önüne alınarak uygun bir strateji ile ele alınmıştır. Bu, analizin hedefine daha uygun ve güvenilir veri seti elde etmeyi sağlamaktadır. Kuraklık puanı yalnızca 7 günde bir kez geçerli olduğundan hedef değişkendeki boş değerlerin kaldırılması işlemi uygulanmıştır.

Veri Setinin Yeniden Biçimlendirilmesi

Veri setinin yapısındaki değişiklikler, analizin gereksinimlerine uyacak şekilde düzenlenmiştir. Tarih ve score bilgisi reformat edilmiştir. Bu, veri setinin daha anlamlı ve etkili bir şekilde kullanılmasına olanak tanır.

3.2. Keşifsel Veri Analizi

Keşifsel Veri Analizi (EDA), bir veri kümesinin özelliklerini anlamada ve daha fazla analize rehberlik edebilecek kalıpları ortaya çıkarmada çok önemli bir adımdır. EDA'nın birincil amacı, nihai sonuçlara ulaşmaktan ziyade hipotezler üretmek ve veriler hakkında içgörü elde etmektir [39]. Bu aşama, veri madenciliği ve istatistiksel analiz süreçlerinin ilk adımını oluşturur ve araştırmacılara veri setinin genel yapısı hakkında önemli içgörüler sağlamaktadır.

Keşifsel veri analizinde etkili olabilmek için, boyutsallık azaltma (DR) tekniklerinin, ideal olarak etkileşimli ve anında yaklaşımlar yoluyla insan ihtiyaçlarına ve alana özgü sorunlara uyarlanabilir olması gerekir [40]. Ayrıca, EDA'nın amacı, verilerdeki kalıpları keşfetmek ve sonraki istatistiksel modelleme veya iş raporlaması için veri özelliklerini özetlemektir [41].

İstatistiksel hipotezlerin çok az olduğu veya hiç olmadığı durumlarda veya belirli hipotezler mevcut olduğunda, ancak istatistiksel sonuçların yorumlanabilirliğini sağlamak için ek temsillere ihtiyaç duyulduğunda, araştırmacıların verileri anlamalarına yardımcı olmayı amaçlayan nicel bir gelenektir [42]. Özetle, EDA, özellikle karmaşık ve çok boyutlu veri kümelerinde verilerin anlaşılmasında ve yorumlanmasında hayati bir rol oynar.

3.2.1. Univariate Analiz

Univariate analiz, tek bir değişkenin dağılımını inceleyerek veri setinin temel özelliklerini tanımlamayı amaçlar. Bu aşama, hem görselleştirmelerle hem de istatistiksel ölçütlerle gerçekleştirilir. Temel istatistiksel ölçütler arasında ortalama, medyan, standart sapma ve çeyrekler bulunmaktadır. Görselleştirmeler arasında histogramlar, kutu grafikleri ve yoğunluk grafiği gibi araçlar yer alır.

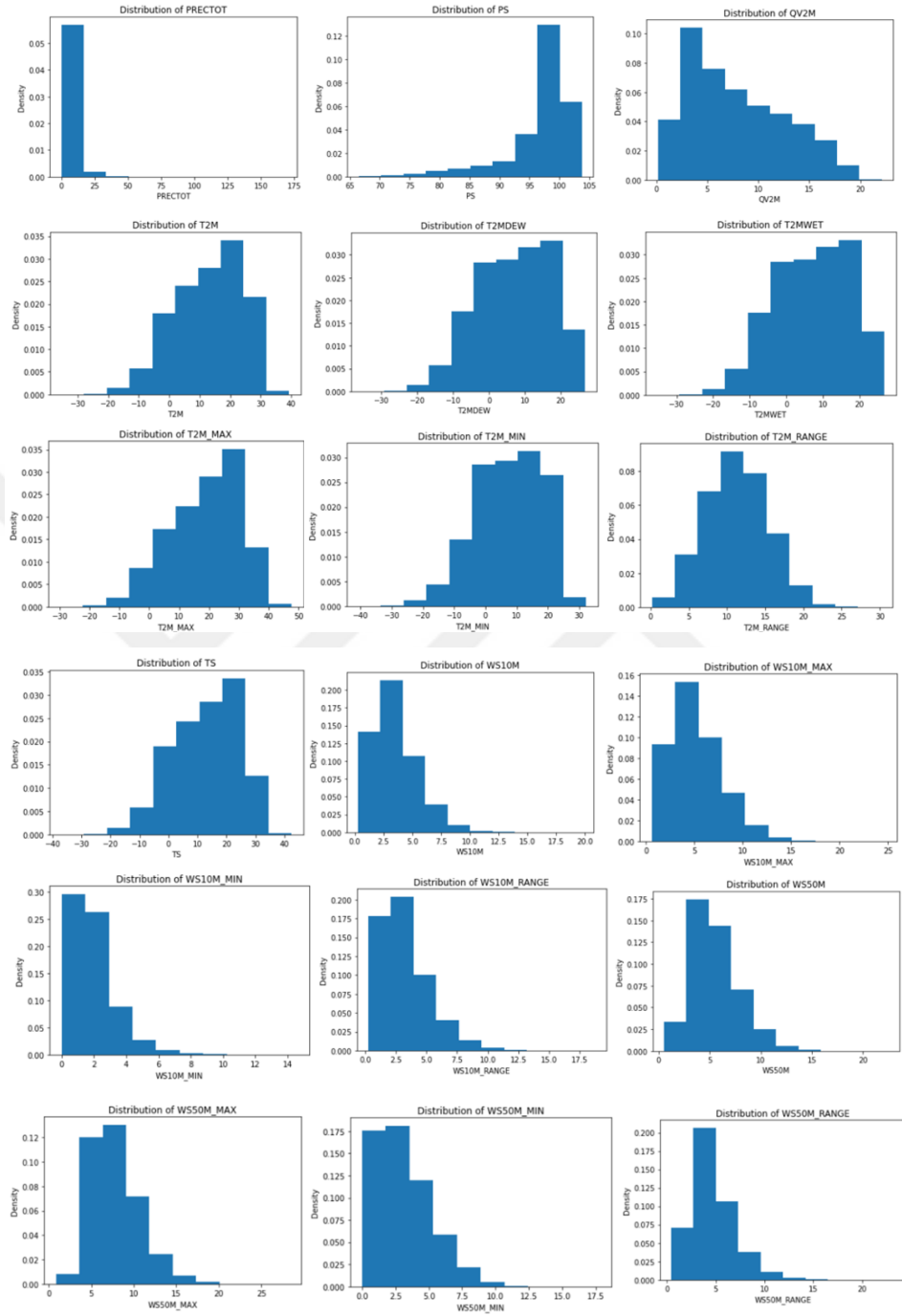
Tanımlayıcı İstatistikler (Descriptive Statistics)

Temel değişkenlerin tanımlayıcı istatistikleri incelenmiştir. Ortalama, medyan, standart sapma gibi istatistiksel ölçümlerle değişkenlerin genel özellikleri belirlenmiştir. Verilerin tanımlayıcı istatistikleri Tablo 3.5'te belirtilmiştir.

	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
FİPS	2,76E+12	3,07E+10	1,50E+10	1,00E+09	1,90E+10	2,92E+10	4,60E+10	5,60E+10
PRECTOT	2,76E+12	2,71E+06	6,25E+06	0,000000e+00	0,000000e+00	1,90E+05	2,26E+06	1,69E+08
PS	2,76E+12	9,66E+07	5,44E+06	6,65E+07	9,58E+07	9,83E+07	9,99E+07	1,04E+08
QV2M	2,76E+12	7,88E+06	4,72E+06	1,40E+05	3,81E+06	6,94E+06	1,15E+07	2,21E+07
T2M	2,76E+12	1,29E+07	1,10E+07	-3,54E+07	4,58E+06	1,42E+07	2,20E+07	3,93E+07
T2MDEW	2,76E+12	7,05E+06	1,02E+07	-3,54E+07	-8,80E+05	7,81E+06	1,57E+07	2,69E+07
T2MWET	2,76E+12	7,08E+06	1,01E+07	-3,55E+07	-8,40E+05	7,81E+06	1,57E+07	2,69E+07
T2M_MAX	2,76E+12	1,88E+07	1,16E+07	-3,00E+07	1,04E+07	2,06E+07	2,80E+07	4,78E+07
T2M_MIN	2,76E+12	7,52E+06	1,06E+07	-4,09E+07	-5,70E+05	8,26E+06	1,63E+07	3,23E+07
T2M_RANGE	2,76E+12	1,12E+07	4,04E+06	1,60E+05	8,37E+06	1,12E+07	1,41E+07	3,02E+07
...
WS10M_MIN	2,76E+12	1,92E+06	1,34E+06	0,000000e+00	9,60E+05	1,66E+06	2,57E+06	1,46E+07
WS10M_RANGE	2,76E+12	3,28E+06	1,94E+06	2,60E+05	1,82E+06	2,82E+06	4,28E+06	1,87E+07
WS50M	2,76E+12	5,53E+06	2,30E+06	5,10E+05	3,82E+06	5,15E+06	6,88E+06	2,25E+07
WS50M_MAX	2,76E+12	7,83E+06	2,84E+06	8,50E+05	5,70E+06	7,35E+06	9,49E+06	2,83E+07
WS50M_MIN	2,76E+12	3,12E+06	2,11E+06	0,000000e+00	1,45E+06	2,77E+06	4,39E+06	1,78E+07
WS50M_RANGE	2,76E+12	4,71E+06	2,18E+06	4,20E+05	3,14E+06	4,27E+06	5,81E+06	2,34E+07
SCORE	2,76E+12	8,23E+05	1,24E+06	0,000000e+00	0,000000e+00	0,000000e+00	1,00E+06	5,00E+06
YEAR	2,76E+12	2,01E+09	4,90E+06	2,00E+09	2,00E+09	2,01E+09	2,01E+09	2,02E+09
MONTH	2,76E+12	6,52E+06	3,45E+06	1,00E+06	4,00E+06	7,00E+06	1,00E+07	1,20E+07
DAY	2,76E+12	1,57E+07	8,80E+06	1,00E+06	8,00E+06	1,60E+07	2,30E+07	3,10E+07

Sürekli Değişkenlerin Dağılımı (Distribution of Continuous Variables)

Sürekli değişkenlerin histogramları çizilerek, veri setindeki dağılımlar görsel olarak incelenmiştir. Bu adım, değişkenlerin merkezi eğilimleri ve dağılım karakteristikleri hakkında fikir sahibi olmayı sağlamaktadır. Sürekli değişkenlerin dağılım grafikleri Şekil 3.2'de gösterilmiştir.



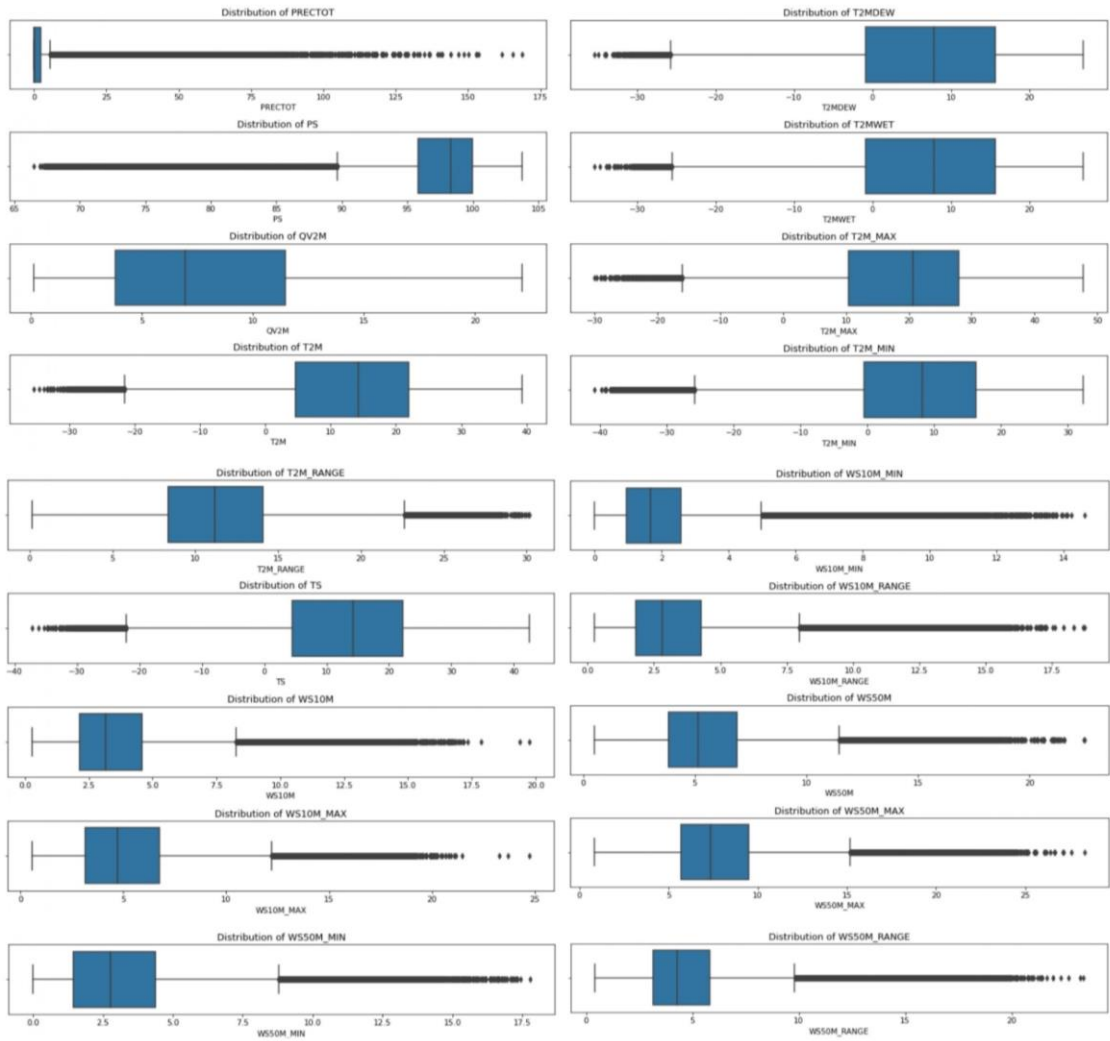
Şekil 3.2. Sürekli Değişkenlerin Dağılımı

3.2.2. Aykırı Değer İşleme

Aykırı değerler, genellikle veri setinde diğer gözlemlerden önemli ölçüde farklı olan noktalar. Bu aşamada, aykırı değerleri tanımlama yöntemleri ve bu değerlere nasıl müdahale edileceği ele

alınır. Aykırı değerleri belirleme adımları, genellikle standart sapma veya çeyrekler arası mesafe gibi istatistiksel metotlara dayanır.

Standart aykırı değer sınırının ötesindeki değerlerin sayısı belirlenmiştir. Aykırı değerler, belirlenen standart aykırılık sınırları dâhilinde kalmayanlar kaldırılmıştır. Bu, analizin istatistiksel sağlamlığını artırmak ve yanıltıcı etkileri önlemek için önemlidir. Aykırı değerlere karşı alınabilecek önlemler arasında kaldırma, değiştirme veya dönüştürme bulunabilir. Şekil 3.3'te aykırı değerlere ilişkin analiz bulunmaktadır.

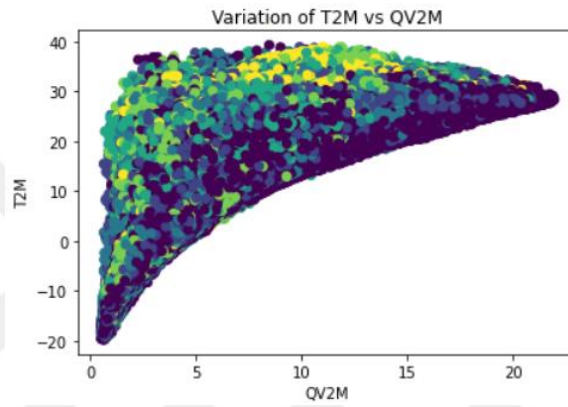


Şekil 3.3. Aykırı Değerleri Belirleme

3.2.3. Bivariate Analiz

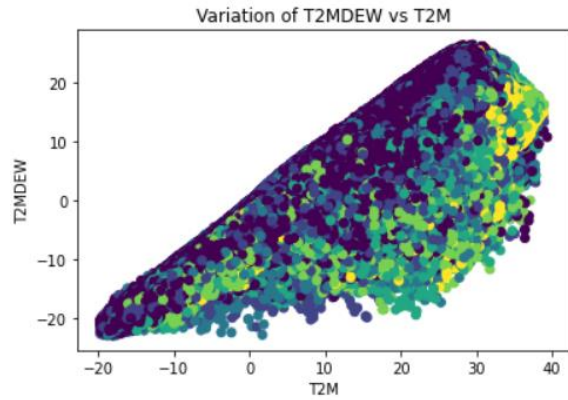
Bivariate analiz, iki deęişken arasındaki ilişkileri ve muhtemel korelasyonları inceleyerek daha derinlemesine içgörüler elde etmeyi amaçlar. Bu aşamada, deęişkenler arasındaki ilişkileri açıklamak için scatter plot, korelasyon katsayısı ve regresyon analizi gibi yöntemler kullanılır.

T2M ve QV2M'nin Deęiřimi: Sıcaklık (T2M) ile özgül nem (QV2M) arasındaki ilişki incelenmiştir. Şekil 3.4'te belirtilmiştir.



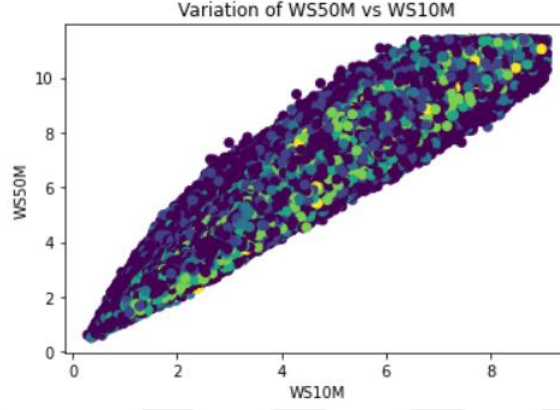
Şekil 3.4. Sıcaklık ve Özgül Nem Deęiřimi

T2MDEW ve T2M'nin Deęiřimi: Çiğ noktası sıcaklığı (T2MDEW) ile genel sıcaklık (T2M) arasındaki ilişki analiz edilmiştir. Şekil 3.5'te görülmektedir.



Şekil 3.5. Çiğ Noktası Sıcaklığı ile Genel Sıcaklık Deęiřimi

WS50M ve WS10M'nin Değişimi: Rüzgar hızı (WS50M) ile rüzgar hızının 10 metre üzerindeki ölçümü (WS10M) arasındaki ilişki incelenmiştir. Şekil 3.6'da görülmektedir.



Şekil 3.6. Rüzgar Hızının Farklı Uzaklıklarda Değişimi

3.2.4. Bağımsız Değişkenler Arasındaki Korelasyon

Analiz için bağımlı ve bağımsız değişkenler belirlenmiştir. Bağımlı değişken genellikle analizin odak noktasını oluştururken, bağımsız değişkenler arasındaki ilişkilerin anlaşılmasına ve modelleme sürecine yardımcı olur.

Veri setindeki QV2M, T2M, T2MDEW, T2MWET, T2M_MAX, T2M_MIN ve TS değişkenleri arasında güçlü pozitif korelasyonlar tespit edilmiştir. Benzer şekilde, WS10M, WS10M_MAX ve WS10M_MIN ile WS50M, WS50M_MAX ve WS50M_MIN değişken grupları arasında da güçlü pozitif korelasyonlar gözlemlenmiştir. Bu durum, bu değişken grupları arasında birbiriyle yakından ilişkili bir değişkenlik olduğunu göstermektedir.

Ancak, scatter plotlarda gözlemlenen belirgin varyans, bu değişken grupları arasında güçlü bir pozitif korelasyon olmasına rağmen, veri noktaları arasında önemli farklılıklar olduğunu işaret etmektedir. Bu durum, değişkenler arasındaki ilişkinin karmaşıklığını ve sadece korelasyon katsayılarına dayanmanın yeterli olmadığını gösterir. Belirli değişkenler arasındaki bu varyans, muhtemelen başka faktörlerin etkisi altında olduklarını veya modelleme sürecinde dikkate alınması gereken diğer önemli değişkenlerin bulunduğunu göstermektedir. Şekil 3.7'de ve Şekil 3.8'de korelasyon görselleştirilmiştir.

	PRECTOT	PS	QV2M	T2M	T2MDEW	T2MWET	T2M_MAX	T2M_MIN	T2M_RANGE	TS
PRECTOT	1.000.000	0.068775	0.245081	0.093258	0.231035	0.230975	0.026773	0.144929	-0.304171	0.089598
PS	0.068775	1.000.000	0.282412	0.164160	0.341234	0.341252	0.111979	0.208285	-0.225935	0.163830
QV2M	0.245081	0.282412	1.000.000	0.870242	0.959385	0.960434	0.804338	0.906144	-0.071547	0.862559
T2M	0.093258	0.164160	0.870242	1.000.000	0.913530	0.914218	0.983356	0.981629	0.244357	0.997515
T2MDEW	0.231035	0.341234	0.959385	0.913530	1.000.000	0.999970	0.854716	0.939934	-0.015643	0.905184
T2MWET	0.230975	0.341252	0.960434	0.914218	0.999970	1.000.000	0.855401	0.940629	-0.015500	0.905911
T2M_MAX	0.026773	0.111979	0.804338	0.983356	0.854716	0.855401	1.000.000	0.937762	0.407534	0.980101
T2M_MIN	0.144929	0.208285	0.906144	0.981629	0.939934	0.940629	0.937762	1.000.000	0.065037	0.979134
T2M_RANGE	-0.304171	-0.225935	-0.071547	0.244357	-0.015643	-0.015500	0.407534	0.065037	1.000.000	0.241564
TS	0.089598	0.163830	0.862559	0.997515	0.905184	0.905911	0.980101	0.979134	0.241564	1.000.000
WS10M	0.049730	-0.080747	-0.225449	-0.207874	-0.238299	-0.237971	-0.216764	-0.206382	-0.080163	-0.189823
WS10M_MAX	0.060981	-0.135905	-0.256452	-0.220192	-0.268686	-0.268292	-0.221671	-0.225829	-0.043127	-0.202713
WS10M_MIN	0.023346	0.022932	-0.108789	-0.125407	-0.115920	-0.115882	-0.141911	-0.112878	-0.110952	-0.110273
WS10M_RANGE	0.065755	-0.198332	-0.269203	-0.209030	-0.280702	-0.280199	-0.199614	-0.225256	0.018746	-0.196015
WS50M	0.069057	-0.043315	-0.205971	-0.193196	-0.204238	-0.204143	-0.195727	-0.197991	-0.041778	-0.180665
WS50M_MAX	0.079508	-0.091821	-0.249961	-0.206444	-0.245323	-0.245147	-0.196236	-0.225744	0.029737	-0.193347
WS50M_MIN	0.057816	0.036238	-0.081554	-0.112579	-0.082416	-0.082497	-0.133234	-0.096593	-0.128844	-0.102367
WS50M_RANGE	0.047477	-0.154479	-0.246203	-0.159589	-0.239335	-0.239029	-0.126331	-0.200157	0.163320	-0.152434

Şekil 3.7. Özellik Seçimi İçin Bağımsız Değişkenler Arasındaki Korelasyon 1

	WS10M	WS10M_MAX	WS10M_MIN	WS10M_RANGE	WS50M	WS50M_MAX	WS50M_MIN	WS50M_RANGE
PRECTOT	0.049730	0.060981	0.023346	0.065755	0.069057	0.079508	0.057816	0.047477
PS	-0.080747	-0.135905	0.022932	-0.198332	-0.043315	-0.091821	0.036238	-0.154479
QV2M	-0.225449	-0.256452	-0.108789	-0.269203	-0.205971	-0.249961	-0.081554	-0.246203
T2M	-0.207874	-0.220192	-0.125407	-0.209030	-0.193196	-0.206444	-0.112579	-0.159589
T2MDEW	-0.238299	-0.268686	-0.115920	-0.280702	-0.204238	-0.245323	-0.082416	-0.239335
T2MWET	-0.237971	-0.268292	-0.115882	-0.280199	-0.204143	-0.245147	-0.082497	-0.239029
T2M_MAX	-0.216764	-0.221671	-0.141911	-0.199614	-0.195727	-0.196236	-0.133234	-0.126331
T2M_MIN	-0.206382	-0.225829	-0.112878	-0.225256	-0.197991	-0.225744	-0.096593	-0.200157
T2M_RANGE	-0.080163	-0.043127	-0.110952	0.018746	-0.041778	0.029737	-0.128844	0.163320
TS	-0.189823	-0.202713	-0.110273	-0.196015	-0.180665	-0.193347	-0.102367	-0.152434
WS10M	1.000.000	0.952217	0.833340	0.702896	0.966275	0.908750	0.795424	0.412412
WS10M_MAX	0.952217	1.000.000	0.690087	0.866026	0.910717	0.946710	0.660428	0.592380
WS10M_MIN	0.833340	0.690087	1.000.000	0.235775	0.839187	0.666629	0.943983	-0.046209
WS10M_RANGE	0.702896	0.866026	0.235775	1.000.000	0.643131	0.810677	0.234645	0.827364
WS50M	0.966275	0.910717	0.839187	0.643131	1.000.000	0.917883	0.847885	0.373539
WS50M_MAX	0.908750	0.946710	0.666629	0.810677	0.917883	1.000.000	0.646726	0.674944
WS50M_MIN	0.795424	0.660428	0.943983	0.234645	0.847885	0.646726	1.000.000	-0.126283
WS50M_RANGE	0.412412	0.592380	-0.046209	0.827364	0.373539	0.674944	-0.126283	1.000.000

Şekil 3.8. Özellik Seçimi İçin Bağımsız Değişkenler Arasındaki Korelasyon 2

Bu nedenle, güçlü pozitif korelasyona rağmen, tüm bu değişkenleri koruma kararı alınmış ve başka özellik seçme yöntemleri denenmiştir. Bu, modelin daha etkili ve güvenilir olması için veri setindeki önemli özellikleri belirleme çabalarını yansıtmaktadır.

3.3. Metot

3.3.1. Veri Standardizasyonu

Veri standardizasyonu, farklı özelliklerin ölçeklerini birbirine uygun hale getirme işlemidir. Veri setindeki özelliklerin farklı birimlerde olması durumunda, bu özelliklerin aynı ölçekte olması, makine öğrenimi modellerinin daha iyi performans göstermesine katkı sağlamaktadır.

Z-Skoru (Z-Score): Veri setindeki her bir özelliğin ortalamasını sıfır, standart sapmasını ise bir yapacak şekilde dönüştürülmesiyle elde edilir. Matematiksel olarak:

$$Z = \frac{X - \text{Ortalama}}{\text{Standart Sapma}} \quad (3.3)$$

Bu işlem, her özelliğin dağılımını standart bir normal dağılıma çevirir.

3.3.2. Özellik Seçimi

Özellik seçimi, model performansını artırmak ve aşırı uyum (overfitting) problemlerini önlemek için kullanılan bir tekniktir. Burada RFE (Recursive Feature Elimination) ve Random Forest algoritmaları kullanılmıştır. RFE ve Random Forest algoritması kullanılarak 21 özellikten 15 tanesi seçilmiştir.

Seçilen Özellikler:

'QV2M', 'T2M', 'T2MWET', 'T2M_MAX', 'T2M_MIN', 'T2M_RANGE', 'TS', 'WS10M', 'WS10M_RANGE', 'WS50M', 'WS50M_MAX', 'WS50M_RANGE', 'year', 'day'.

RFE (Recursive Feature Elimination)

Bu yöntem, başlangıçta tüm özellikleri içeren bir model oluşturur ve en az katkı yapan özellikleri sırasıyla kaldırarak modeli günceller. Bu işlem, model performansını etkilemeyen özellikleri elemine ederek daha etkili bir model elde etmeyi amaçlar.

Random Forest Algoritması

Random Forest, birden çok karar ağacını birleştirerek daha güçlü bir model oluşturan bir ensemble (ansambl) algoritmasıdır. Özellik önem sıralamalarını kullanarak, veri setindeki özelliklerin model üzerindeki etkisini değerlendirir.

3.3.3. Sınıf Dengesizliği İle Başa Çıkma

Bu işlemler, modelin azınlık sınıfa ait örnekleri daha iyi öğrenmesini sağlayarak, sınıf dengesizliği problemini ele almayı amaçlar.

SMOTE ile Aşırı Örnekleme (Upsampling SMOTE)

SMOTE, özellikle azınlık sınıfa ait verilerin sayısını artırmak amacıyla kullanılan bir yöntemdir. Bu, azınlık sınıfındaki veri sayısını artırarak sınıflar arasındaki dengesizliği azaltmayı hedefler. Yani, SMOTE kullanılarak her bir azınlık sınıfı için yeni sentetik örnekler üretilmiş ve böylece sınıf dağılımları dengelenmiştir.

Near Miss ile Azaltma (Downsampling Near Miss)

Near Miss, özellikle çoğunluk sınıfına ait örneklerin sayısını azaltarak sınıflar arasındaki dengesizliği azaltmayı amaçlar. Near Miss yöntemi, her sınıftan çoğunluk sınıfına ait olan örnekleri, o sınıfa ait azınlık sınıfına en yakın örneklerle değiştirerek sınıf dengesizliğini azaltır.

Near Miss ile Azaltma (Downsampling Near Miss) Tekrarı

Aynı Near Miss yöntemi bir kez daha uygulanarak sınıf dengesizliği daha fazla azaltılmıştır. Bu adımda, önceki adımda azaltılmış olan örneklerin sayısı daha da düşürülmüştür. Bu adımda, Near Miss yöntemiyle önceki adımda azaltılmış olan örneklerin sayısı daha da düşürülerek sınıf dengesizliği daha da iyileştirilmiştir.

3.3.4. Boyut Azaltma İşlemleri

Bu işlemler, azınlık sınıfa ait özellikleri daha belirgin hale getirerek ve gereksiz bilgi kaybını minimize ederek veri setini daha etkili hale getirmeyi amaçlar. Ayrıca, bu sayede modelin genelleme yeteneği artırılır ve overfitting riski azaltılır. Sınıf dengesizliği işlemleri sonrasında elde edilen veri setlerinin boyutu azaltılarak, modelin daha hızlı öğrenmesi ve daha az hesaplama maliyetiyle çalışması sağlanmıştır.

PCA (Principal Component Analysis)

PCA, özellik matrisinin temel bileşenlere dönüştürülmesini sağlayan bir boyut indirgeme tekniğidir. Bu, veri setindeki varyansı korurken özellik sayısını azaltmaya yardımcı olur.

Near Miss yöntemi uygulandıktan sonra, elde edilen veri seti üzerine PCA uygulanmış ve özellik sayısı 5'e düşürülmüştür. Aynı şekilde SMOTE yöntemi uygulandıktan sonra, elde edilen veri seti üzerine PCA uygulanmış ve özellik sayısı 5'e düşürülmüştür.

LDA (Linear Discriminant Analysis)

LDA, sınıflar arasındaki ayrımı artırmak amacıyla özellik matrisini dönüştüren bir boyut indirgeme tekniğidir. Bu, sınıflar arasındaki farkları daha iyi öne çıkarmayı amaçlar.

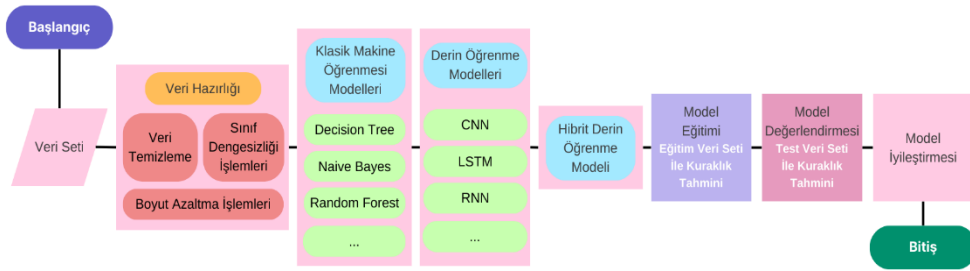
Near Miss yöntemi uygulandıktan sonra, elde edilen veri seti üzerine LDA uygulanmış ve özellik sayısı 5'e düşürülmüştür. Aynı şekilde SMOTE yöntemi uygulandıktan sonra, elde edilen veri seti üzerine LDA uygulanmış ve özellik sayısı 5'e düşürülmüştür.

3.4. Model Eğitimi ve Değerlendirmesi

3.4.1. Model Seçimi

Kuraklık tahmini için birden fazla makine öğrenmesi algoritması tercih edilmiştir. Bu seçimler, yüksek derecede özelleştirilebilirlik ve yorumlanabilirlik, veri setindeki karmaşıklığı anlama yeteneği ve sınıflandırma problemleri için etkili bir çözüm aranması nedeniyle yapılmıştır. Ayrıca, sınıf dengesizliği ile başa çıkma yeteneği ve işlem hızı avantajları da dikkate alınmıştır.

Veri setinde 1979470 satır ve 15 sütun bulunan eğitim özellikleri ve 1979470 satır ve 1 sütun bulunan eğitim hedefleri alınmıştır. Ayrıca, veri setinde 494868 satır ve 15 sütun bulunan test özellikleri ve 494868 satır ve 1 sütun bulunan test hedefleri alınmıştır. Bu veriler, makine öğrenimi modelinin eğitimi ve testi için kullanılır. Şekil 3.9'da kuraklık tahmini akışı yer almaktadır.



Şekil 3.9. Kuraklık Tahmini Akışı

3.4.2. Eğitim Süreci

Kuraklık tahmini çalışmasında kullanılan eğitim süreci, veri hazırlama ve model eğitimi aşamalarını içermektedir. Bu süreç, veri setinin hazırlanması, özelliklerin seçimi ve sınıf dengesizliği probleminin giderilmesine yönelik adımları kapsamaktadır.

Veri Seti Ayrımı ve Standardizasyon: Çalışmanın ilk aşamasında, veri seti eğitim (train) ve test olmak üzere ikiye ayrılmıştır. Bu ayırım, modelin eğitimi sırasında aşırı öğrenmeyi önlemek ve

performansının bağımsız bir veri seti üzerinde değerlendirilmesini sağlamak için kritik bir adımdır. Ardından, veri seti üzerinde standardizasyon işlemi gerçekleştirilmiştir. Bu işlem, eğitim setinden elde edilen ortalama ve standart sapma değerleri kullanılarak, tüm veri setinin ortak bir ölçüğe getirilmesini sağlamıştır.

Özellik Seçimi: Modelin karmaşıklığını azaltmak ve tahmin performansını artırmak amacıyla, Recursive Feature Elimination (RFE) algoritmalar kullanılarak özellik seçimi yapılmıştır. Bu yöntemlerle, veri setindeki en önemli 15 özellik belirlenmiş ve model eğitimi bu özellikler üzerinden gerçekleştirilmiştir. Bu adım, özellikle yüksek boyutlu veri setlerinde modelin daha hızlı ve daha etkili öğrenmesini sağlamaktadır.

Sınıf Dengesizliği İşlemleri: Sınıf dengesizliği problemini çözmek amacıyla çeşitli yöntemler kullanılmıştır. Bu kapsamda, Near Miss ve SMOTE gibi dengesiz veri setleri üzerinde etkili olan yöntemler uygulanmıştır.

Ayrıca, Principal Component Analysis (PCA) ve Linear Discriminant Analysis (LDA) gibi boyut indirgeme yöntemleri kullanılarak veri setindeki varyansın korunması ve sınıf ayrımının optimize edilmesi sağlanmıştır. Bu yöntemler, modelin performansını artırmak ve veri setinin daha anlaşılır hale gelmesini sağlamak amacıyla uygulanmıştır.

Performans Metrikleri: Her bir model, accuracy, precision, recall, F1 score, Cohen Kappa skoru gibi çoklu sınıf sınıflandırma metrikleri ile değerlendirilmiştir.

Makine öğrenmesi alanında, doğruluk, hassasiyet, geri çağırma, F1 skoru ve Cohen Kappa gibi performans metrikleri, model etkinliğini değerlendirmek için kritik öneme sahiptir. Doğruluk, toplam gözlemlenen vakalar arasında doğru sonuçların (hem doğru pozitiflerin hem de doğru negatiflerin) oranını ölçer. Ancak, sınıf dengesizliğinin olduğu durumlarda doğruluk yanıltıcı olabilir ve bu nedenle hassasiyet ve geri çağırma gibi metriklerin kullanılması gereklidir. Hassasiyet, doğru pozitiflerin sayısının, tahmin edilen pozitiflerin toplam sayısına bölünmesiyle hesaplanırken, geri çağırma (veya duyarlılık) gerçek pozitiflerin, veri setindeki gerçek pozitiflere oranını ölçer [43, 44].

F1 skoru, hassasiyet ve geri çağırmanın harmonik ortalamasıdır ve her iki metrik arasında bir denge sağlayan tek bir ölçüt sunar. Bu metrik, birinin diğerine göre daha fazla öncelik taşması durumunda özellikle faydalıdır [43, 45]. Örneğin, farklı makine öğrenmesi modellerinin karşılaştırıldığı bir çalışmada, F1 skoru dengesiz veri setleri üzerinde performansı değerlendirmek için kullanılmış ve Gradient Boosting Classifier (GBC), 0.658 F1 skoru ile CatBoost sınıflandırıcısını 0.654 skorluk bir performansla geride bırakmıştır [43]. Bu, F1 skorunun model performansına daha ince bir bakış

açısı sunduğunu, özellikle yanlış negatiflerin maliyetli olduğu uygulamalarda önemli olduğunu vurgulamaktadır.

Cohen Kappa, tahmin edilen ve gözlemlenen sınıflandırmalar arasındaki uyumu ölçen bir başka önemli metriktir ve rastgele gerçekleşebilecek uyum olasılığını göz önünde bulundurur. Çok sınıflı sınıflandırma problemlerinde özellikle kullanışlıdır ve modelin yaptığı tahminlerin güvenilirliği hakkında önemli bilgiler sunabilir [46]. Örneğin, prediktif modelleme çalışmalarında Kappa değerleri, modelin rastgele tahminlerin ötesinde ne kadar iyi performans gösterdiğini gösterebilir ve böylece model performansının daha sağlam bir şekilde değerlendirilmesine olanak tanır [46].

Pratik uygulamalarda, bu metrikler için yüksek değerlere sahip çeşitli çalışmalar rapor edilmiştir. Örneğin, bir öğretimi performansı tahmin etmek için geliştirilen makine öğrenmesi modeli, %98.89 doğruluk elde etmiş ve hassasiyet, geri çağırma ve F1 skorları %98'in üzerinde olmuştur [45]. Benzer şekilde, endüstriyel siber güvenlik bağlamında bir model, %97.50 doğruluk ve %97.60 hassasiyet ile geri çağırma değerleri sergileyerek, makine öğrenmesi algoritmalarının gerçek dünya uygulamalarındaki etkinliğini vurgulamaktadır [47]. Bu örnekler, birden fazla metriğin kullanıldığı durumlarda, modelin güçlü ve zayıf yönlerinin kapsamlı bir şekilde anlaşılmasını sağladığını ve bu doğrultuda daha ileri düzeyde iyileştirme ve rafine etme çabalarına rehberlik ettiğini göstermektedir.

Sınıflandırma Metrikleri

1. **Doğruluk (Accuracy):** Doğruluk, doğru sınıflandırılan örneklerin toplam örneğe oranını ifade eder.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.4.1)$$

- **TP (True Positive):** Gerçek pozitiflerin sayısı
- **TN (True Negative):** Gerçek negatiflerin sayısı
- **FP (False Positive):** Yanlış pozitiflerin sayısı
- **FN (False Negative):** Yanlış negatiflerin sayısı

2. **Hassasiyet (Precision):** Hassasiyet, pozitif olarak tahmin edilen örneklerin ne kadarının gerçekten pozitif olduğunu gösterir.

$$Precision = \frac{TP}{TP + FP} \quad (3.4.2)$$

3. **Geri Çağırma (Recall):** Geri çağırma, pozitif örneklerin ne kadarının doğru şekilde tespit edildiğini belirtir.

$$Recall = \frac{TP}{TP + FN} \quad (3.4.3)$$

4. **F1 Skoru:** F1 skoru, hassasiyet ve geri çağırmanın harmonik ortalamasıdır.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3.4.4)$$

5. **Cohen Kappa:** Cohen Kappa, modelin rastlantısal sınıflandırma performansına göre ne kadar iyi olduğunu ölçen bir istatistiktir.

$$\kappa = \frac{Po - Pe}{1 - Pe} \quad (3.4.5)$$

- Po: Gözlemlenen (gerçek) doğruluk oranı
- Pe: Rastlantısal doğruluk oranı (beklenen doğruluk)

Regresyon Metrikleri

1. **MAE (Mean Absolute Error – Ortalama Mutlak Hata):** MAE, tahmin edilen değerler ile gerçek değerler arasındaki mutlak farkların ortalamasıdır.

$$MAE = \left(\frac{1}{n}\right) \sum_{i=1}^n |Y_i - Y_{ii}| \quad (3.4.6)$$

- Y_i : Gerçek değer
- Y_{ii} : Tahmin edilen değer
- n: Örnek sayısı

2. **RMSE (Root Mean Squared Error – Kök Ortalama Kare Hata):** RMSE, hata karelerinin ortalamasının karekökünü alarak hesaplanır.

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (Y_i - Y_{ii})^2} \quad (3.4.7)$$

3. **R^2 (Determination Coefficient – Belirleme Katsayısı):** R^2 , modelin açıklayabildiği varyans oranını gösterir.

$$R^2 = 1 - \left(\frac{SS_{res}}{SS_{tot}}\right) \quad (3.4.8)$$

- SSres: Kalan kareler toplamı
- SStot: Toplam kareler toplamı

Uygulanan Modeller: Eğitim sürecinde farklı makine öğrenmesi algoritmaları kullanılmıştır. Bu algoritmalar şunlardır:

- Random Forest: Resampling uygulanmadan kullanılmış ve ayrıca SMOTE ve PCA ile kombinasyonlar halinde değerlendirilmiştir.
- K-Nearest Neighbors (KNN): Hem resampling olmadan hem de SMOTE ve Near Miss yöntemleri ile birlikte uygulanmıştır.
- Karar Ağacı: Resampling uygulanmadan, SMOTE ile, ve PCA ile birlikte kullanılmıştır. Ayrıca, Near Miss ile çeşitli kombinasyonları da değerlendirilmiştir.
- Naive Bayes: Resampling olmadan uygulanmıştır.
- Destek Vektör Makineleri (SVM): Near Miss yöntemi ile birlikte kullanılmıştır.

Bu süreçte, her bir modelin performansı, farklı sınıf dengesizliği giderme teknikleri ile birlikte test edilerek en uygun yöntemler belirlenmiştir. Bu kapsamlı eğitim süreci, modelin genel performansını artırmak ve kuraklık tahmini konusunda daha isabetli sonuçlar elde etmek amacıyla tasarlanmıştır.

Meteorolojik Veriler İçin Hibrit Derin Öğrenme Modeli

Kuraklık tahmini çalışmasında, farklı veri türlerinden en iyi şekilde faydalanmak amacıyla optimize edilmiş bir hibrit derin öğrenme modeli geliştirilmiştir. Bu model, hem hızlı özellik çıkarımı hem de zaman serisi analizine uygun olacak şekilde tasarlanmıştır. Model, birleştirilmiş CNN ve LSTM dalları ile karmaşık veri örüntülerini öğrenme yeteneğine sahiptir.

Model, giriş katmanında yer alan Convolutional Neural Network (CNN) ve Long Short-Term Memory (LSTM) ağlarının birleşiminden oluşmaktadır. CNN dalı, giriş verilerinden hızlı ve etkili bir şekilde özellik çıkarımı yaparken, LSTM dalı, zaman serisi verilerinin temporal özelliklerini öğrenme yeteneği ile katkı sağlamaktadır. Bu iki dal, birleştirilerek daha güçlü bir temsil oluşturulmuş ve nihai tahmin katmanına iletilmiştir.

Modelin eğitimi sırasında GPU bellek optimizasyonu sağlanmış ve TensorFlow'un bellek yönetim özellikleri kullanılarak modelin verimli bir şekilde çalışması sağlanmıştır. Ayrıca, modelin daha hızlı ve düşük bellek kullanımına sahip olması amacıyla karma hassasiyetli eğitim (mixed precision

training) uygulanmıştır. Eğitim sürecinde, erken durdurma (Early Stopping) ve öğrenme oranı azaltma (ReduceLRonPlateau) gibi callback'ler kullanılarak modelin aşırı öğrenme yapması engellenmiş ve eğitim süreci optimize edilmiştir.

SPEI İçin Hibrit Derin Öğrenme Modeli

Model, geçmiş SPEI zaman serisini ve mekânsal koordinatları girdi olarak alarak gelecekteki SPEI değerlerini tahmin etmektedir. Hibrit model, birden fazla LSTM katmanı kullanılarak geliştirilmiştir.

Modelin yapısı aşağıdaki gibidir:

- Giriş Katmanı: Zaman serisi verilerini alır.
- LSTM Katmanları: Uzun vadeli bağımlılıkları öğrenmek için iki katmanlı LSTM kullanılmıştır (64 ve 32 nöron).
- Tam Bağlantılı (Dense) Katmanlar: Öğrenilen temsilleri işleyerek çıkışı üretir.
- Çıkış Katmanı: Kuraklık endeksini tahmin eder.

Modelin eğitimi sırasında GPU bellek optimizasyonu sağlanmış ve TensorFlow'un bellek yönetim özellikleri kullanılarak modelin verimli bir şekilde çalışması sağlanmıştır.

Model, Adam optimizasyon algoritması ve Huber kayıp fonksiyonu kullanılarak eğitilmiştir. Eğitim sırasında erken durdurma (EarlyStopping) mekanizması eklenmiş ve modelin aşırı öğrenme yapması önlenmiştir.

4. BULGULAR

Bu çalışmanın temel bulguları şu şekildedir: SMOTE ile yapılan aşırı örnekleme, modellerin performansını belirgin bir şekilde artırmıştır. Aynı zamanda, boyut azaltma yöntemleri olan LDA ve PCA'nın kullanımıyla modelin genel performansında olumlu bir etki gözlemlenmektedir. Hiperparametre optimizasyonu da modelin daha yüksek bir performans sergileme potansiyeline katkı sağlamıştır.

Gerçekleştirilen hibrit derin öğrenme modelinin performansı, doğruluk ve kayıp değerleri üzerinden değerlendirilmiştir. Bu sonuç, modelin kuraklık tahmini konusunda ne derece etkili olduğunu göstermektedir. Hibrit derin öğrenme yaklaşımı, farklı veri türlerinin avantajlarını bir araya getirerek tahmin performansını önemli ölçüde artırmıştır. Tablo 4.1'de modellerin performanslarına göre kategorileri belirlenmiştir.

Tablo 4.1. Model Performans Kategorileri

Performans Düzeyi	Modeller	Ortalama F1 Skoru	Temel Özellikler
Yüksek Performans	Random Forest, KNN (tüm varyantlar)	0.798	Metrikler arası tutarlı performans, yüksek κ skorları
Orta Performans	Karar Ağacı varyantları (Near Miss hariç)	0.745	Metrikler arası iyi denge, orta κ skorları
Düşük Performans	Naive Bayes, Near Miss varyantları	0.308	Düşük duyarlılık, düşük κ skorları

Elde edilen sonuçlar, modellerin mevcut durumunu güçlendirmenin yanı sıra, gelecekteki çalışmalar ve geliştirmeler için bir temel oluşturmaktadır. Bu bağlamda, modellerin karmaşıklığını artırmak ve daha geniş bir perspektife taşımak için yapılabilecek potansiyel iyileştirmeler üzerinde durulmalıdır.

Tablo 4.2'de ve Tablo 4.3'de bulgular yer almaktadır.

Tablo 4.2. Kuraklık Tahmini: Upsampling Yaklaşımları

Algoritma	Accuracy	Precision	Recall	F1 Score	Cohen Kappa Score
Resampling Olmadan Random Forest	0.808959	0.796925	0.808959	0.798690	0.654981

Resampling Olmadan KNN	0.798651	0.798294	0.798651	0.798471	0.657498
SMOTE Upsampling ile KNN	0.795267	0.801758	0.795267	0.798198	0.657827
SMOTE Upsampling ile Karar Ağacı	0.764228	0.772588	0.764228	0.767987	0.607222
Resampling Olmadan Karar Ağacı	0.763337	0.762305	0.763337	0.762809	0.596681
SMOTE Upsampling ve PCA ile Karar Ağacı	0.691158	0.721815	0.691158	0.703299	0.504504
CNN ve LSTM ile Hibrit Derin Öğrenme Modeli	0.6056	-	-	-	-

SPI değerleri için oluşturulan hibrit modelin performansı farklı metrikler ile değerlendirilmiştir. Elde edilen sonuçlar, modelin kuraklık tahmini için yüksek doğruluk sunduğunu göstermektedir.

Tablo 4.3. Kuraklık Tahmini: Downsampling Stratejileri

Algoritma	Accuracy	Precision	Recall	F1 Score	Cohen Kappa Score
SMOTE Upsampling ve LDA ile Karar Ağacı	0.602809	0.674628	0.602809	0.628327	0.394723
Naive Bayes without resampling	0.585144	0.449910	0.585144	0.480441	0.080746
Near Miss Downsampling ile SVM	0.299534	0.512324	0.299534	0.362867	0.078111
Near Miss Upsampling ile KNN	0.232508	0.566489	0.232508	0.268879	0.093552
Near Miss Downsampling ile Karar Ağacı	0.224805	0.543185	0.224805	0.262600	0.078760
Near Miss Downsampling ve LDA ile Karar Ağacı	0.204748	0.514249	0.204748	0.249713	0.057772
Near Miss Downsampling ve PCA ile Karar Ağacı	0.189016	0.520894	0.189016	0.224072	0.059722

Özellikle LSTM tabanlı yaklaşımın, zaman serisi bağımlılıklarını başarılı bir şekilde öğrendiği gözlemlenmiştir. Tablo 4.4'de sonuçlar belirtilmiştir.

Tablo 4.4. SPI Hibrit Modeli Sonuçları

Metrik	Değer
Ortalama Mutlak Hata (MAE)	0.12
Kök Ortalama Kare Hatası (RMSE)	0.15
R² Skoru	0.89

Yeniden örnekleme tekniklerinin modellerin performansı üzerindeki etkisini görebilmek için Tablo 4.5 incelenmelidir. Δ , yeniden örnekleme tekniği uygulandığındaki metrik değişimini gösterir (pozitif değerler iyileşmeyi gösterir).

Tablo 4.5. Yeniden Örnekleme Tekniklerinin Model Performansına Etkisi

Yeniden Örnekleme Yöntemi	Model	Δ Doğruluk	Δ Kesinlik	Δ Duyarlılık	Δ F1 Skoru	Δ Cohen's κ
Örneklemez vs SMOTE	KNN	-0.003	+0.003	-0.003	-0.000	+0.000
Örneklemez vs SMOTE	Karar Ağacı	+0.001	+0.010	+0.001	+0.005	+0.011
Örneklemez vs Near Miss	KNN	-0.566	-0.232	-0.566	-0.530	-0.564
Örneklemez vs Near Miss	Karar Ağacı	-0.538	-0.219	-0.538	-0.500	-0.518

Yapay zeka algoritma ailelerinin performanslar bazında karşılaştırması Tablo 4.6'da yapılmıştır. Kararlılık, farklı konfigürasyonlardaki performans metriklerinin standart sapması ile ölçülmüştür.

Tablo 4.6. Algoritma Ailelerinin Performans Karşılaştırması

Algoritma Ailesi	Test Edilen Modeller	Ort. Doğruluk	Ort. F1 Skoru	Ort. κ	Kararlılık*
Ağaç-tabanlı	Random Forest, Karar Ağaçları	0.757	0.758	0.591	Yüksek
Mesafe-tabanlı	KNN varyantları	0.797	0.798	0.658	Yüksek
Olasılıksal	Naive Bayes	0.585	0.480	0.081	Düşük
Derin Öğrenme	CNN-LSTM Hibrit	0.606	-	-	-
SVM-tabanlı	Near Miss SVM	0.300	0.363	0.078	Düşük

4.1. Meteorolojik Verilerden Hibrit Derin Öğrenme Modeli

Çalışmada, hibrit bir derin öğrenme modeli kullanılarak çoklu sınıf sınıflandırma problemi üzerine deneyler gerçekleştirilmiştir. Model, bir Convolutional Neural Network (CNN) ve Long Short-Term Memory (LSTM) ağlarının birleşiminden oluşarak hem uzaysal hem de zamansal özellikleri çıkarmada etkin bir yaklaşım sunmuştur.

Eğitim Süreci ve Sonuçlar: Model, 30 epoch boyunca eğitilmiş ve her epoch sonunda modelin eğitim ve doğrulama doğrulukları ile kayıpları gözlemlenmiştir. İlk epoch sonunda doğruluk (accuracy) %64,15 olarak hesaplanmışken, kayıp değeri (loss) 0,9202 olarak gözlemlenmiştir.

Eğitim ilerledikçe modelin doğruluk değeri artmış ve kayıp değeri azalmıştır. 30. epoch sonunda, eğitim doğruluğu %77,81'e ulaşırken, doğrulama doğruluğu %78,47'ye yükselmiştir. Bu süreçte, modelin doğrulama seti üzerindeki kayıp değeri 0,5111 olarak ölçülmüştür.

Modelin test seti üzerinde yapılan değerlendirmeler sonucunda, aşağıdaki performansı hesaplanmıştır:

- **Accuracy (Doğruluk):** 0.6056

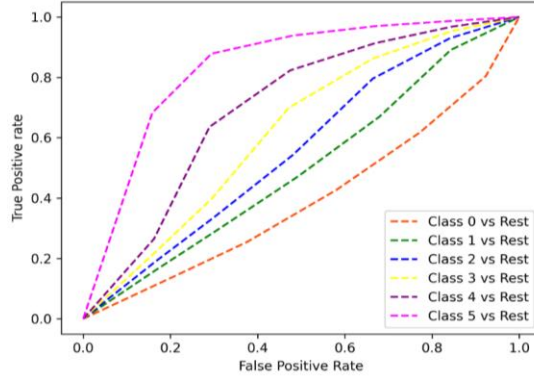
Meteorolojik verilerin sonuçları Tablo 4.7'de verilmiştir.

Tablo 4.7. Meteorolojik Hibrit Modelin Sonuçları

Aşama	Doğruluk (Accuracy)	Kayıp (Loss)
1. Epoch (Eğitim)	64,15%	0.9202
30. Epoch (Eğitim)	77,81%	-
30. Epoch (Doğrulama)	78,47%	0.5111
Test Seti	0.6056	-

4.2. Near Miss Downsampling ile Karar Ağacı

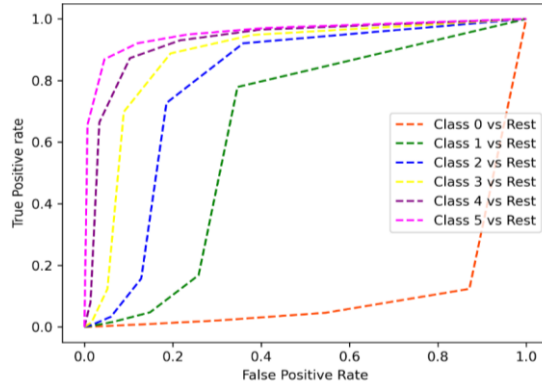
Performans ölçütleri: Accuracy: 0.2263, Precision: 0.5431, Recall: 0.2263, F1 Score: 0.2646, Cohen Kappa Score: 0.0794 olarak hesaplanmıştır. Düşük accuracy ve recall değerleri, sınıf dengesizliği azaltılmış olsa da modelin performansını düşürmüştür. Şekil 4.1'de ROC eğrisi bulunmaktadır.



Şekil 4.1. Near Miss Downsampling ile Karar Ağacı için Çok Sınıflı ROC eğrisi

4.3. SMOTE Upsampling ile Karar Ağacı

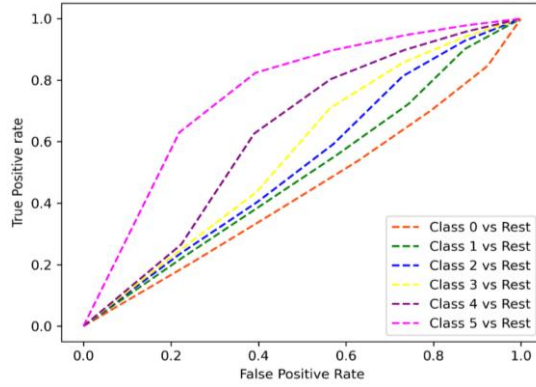
Performans ölçütleri: Accuracy: 0.7642, Precision: 0.7728, Recall: 0.7642, F1 Score: 0.7680, Cohen Kappa Score: 0.6073 olarak hesaplanmıştır. Yüksek accuracy, precision ve recall değerleri, sınıf dengesizliğinin etkili bir şekilde giderildiğini göstermektedir. Şekil 4.2'de ROC eğrisi bulunmaktadır.



Şekil 4.2. SMOTE Upsampling ile Karar Ağacı için Çok Sınıflı ROC eğrisi

4.4. Near Miss Downsampling ve PCA ile Karar Ağacı

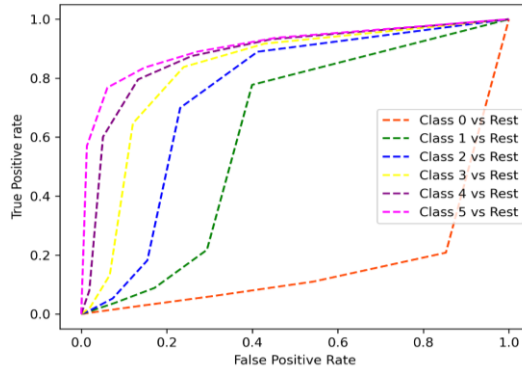
Performans ölçütleri: Accuracy: 0.1884, Precision: 0.5195, Recall: 0.1884, F1 Score: 0.2234, Cohen Kappa Score: 0.0589 olarak hesaplanmıştır. Düşük accuracy ve recall değerleri, sınıf dengesizliği ve boyut indirgeme işlemlerinin kombinasyonunun model performansını olumsuz etkilediğini göstermektedir. Şekil 4.3’de ROC eğrisi bulunmaktadır.



Şekil 4.3. Near Miss Downsampling ve PCA ile Karar Ağacı için Çok Sınıflı ROC eğrisi

4.5. SMOTE Upsampling ve PCA ile Karar Ağacı

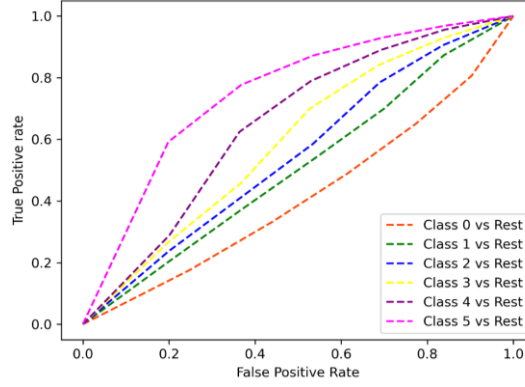
Performans ölçütleri: Accuracy: 0.6912, Precision: 0.7219, Recall: 0.6912, F1 Score: 0.7033, Cohen Kappa Score: 0.5046 olarak hesaplanmıştır. Yüksek accuracy, precision ve recall değerleri, sınıf dengesizliği ve boyut indirgeme işlemlerinin etkili bir şekilde uygulandığını göstermektedir. Şekil 4.4’de ROC eğrisi bulunmaktadır.



Şekil 4.4. SMOTE Upsampling ve PCA ile Karar Ağacı için Çok Sınıflı ROC eğrisi

4.6. Near Miss Downsampling ve LDA ile Karar Ağacı

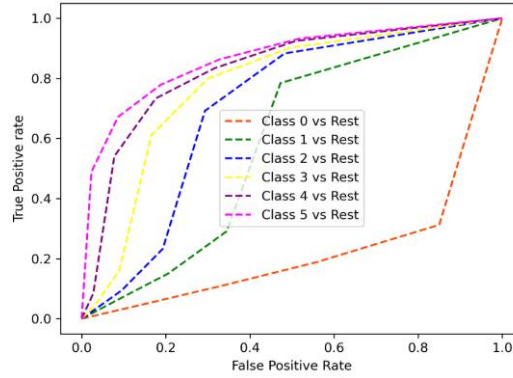
Performans ölçütleri: Accuracy: 0.2057, Precision: 0.5153, Recall: 0.2057, F1 Score: 0.2511, Cohen Kappa Score: 0.0586 olarak hesaplanmıştır. Düşük değerler, sınıf dengesizliği ve boyut indirgeme işlemlerinin kombinasyonunun model performansını olumsuz etkilediğini göstermektedir. Şekil 4.5’de ROC eğrisi bulunmaktadır.



Şekil 4.5. Near Miss Downsampling ve LDA ile Karar Ağacı için Çok Sınıflı ROC eğrisi

4.7. SMOTE Upsampling ve LDA ile Karar Ağacı

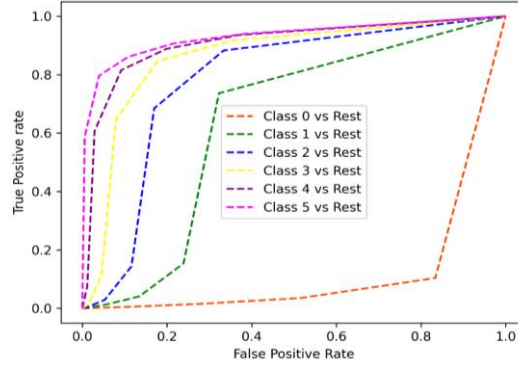
Performans ölçütleri: Accuracy: 0.6031, Precision: 0.6747, Recall: 0.6031, F1 Score: 0.6286, Cohen Kappa Score: 0.3951 olarak hesaplanmıştır. Yüksek accuracy, precision ve recall değerleri, sınıf dengesizliği ve boyut indirgeme işlemlerinin etkili bir şekilde uygulandığını göstermektedir. Şekil 4.6’de ROC eğrisi bulunmaktadır.



Şekil 4.6. SMOTE Upsampling ve LDA ile Karar Ağacı için Çok Sınıflı ROC eğrisi

4.8. Karar Ağacı (Herhangi Bir Dengesizlik İşlemi Uygulanmadan)

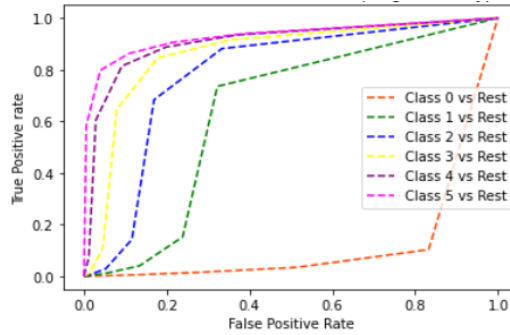
Performans ölçütleri: Accuracy: 0.7628, Precision: 0.7617, Recall: 0.7628, F1 Score: 0.7622, Cohen Kappa Score: 0.5957 olarak hesaplanmıştır. Yüksek değerler, sınıf dengesizliğini göz önüne almadan eğitilen modelin iyi bir performans sergilediğini göstermektedir. Şekil 4.7’de ROC eğrisi bulunmaktadır.



Şekil 4.7. Karar Ağacı için herhangi bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi

4.9. Hiperparametre Ayarı ile Karar Ağacı (Herhangi Bir Dengesizlik İşlemi Uygulanmadan)

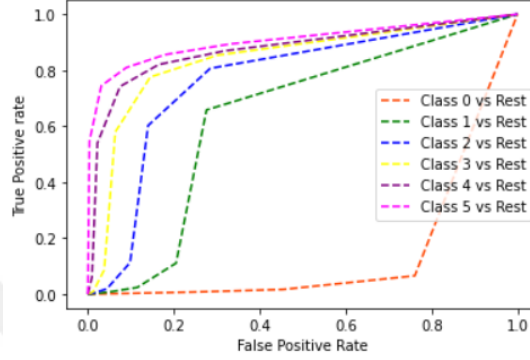
Performans ölçütleri: Accuracy: 0.7628, Precision: 0.7617, Recall: 0.7628, F1 Score: 0.7622, Cohen Kappa Score: 0.5957 olarak hesaplanmıştır. Yüksek değerler, sınıf dengesizliğini göz önüne almadan eğitilen modelin iyi bir performans sergilediğini göstermektedir. Şekil 4.8’de ROC eğrisi bulunmaktadır.



Şekil 4.8. Karar Ağacı Hiperparametre ayarı herhangi bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi

4.10. KNN (Herhangi Bir Dengesizlik İşlemi Uygulanmadan)

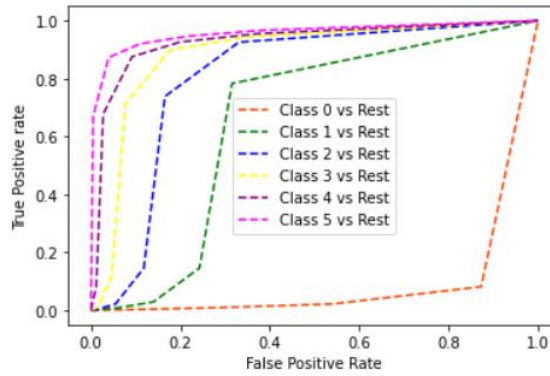
Performans ölçütleri: Accuracy: 0.7628, Precision: 0.7617, Recall: 0.7628, F1 Score: 0.7622, Cohen Kappa Score: 0.5957 olarak hesaplanmıştır. Yüksek, sınıf dengesizliğini göz önüne almadan eğitilen modelin iyi bir performans sergilediğini göstermektedir. Şekil 4.9'da ROC eğrisi bulunmaktadır.



Şekil 4.9. KNN herhangi bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi

4.11. Hiperparametre Ayarı İle KNN (Herhangi Bir Dengesizlik İşlemi Uygulanmadan)

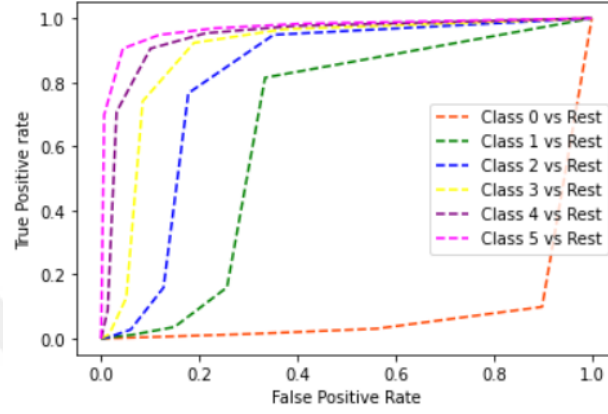
Performans ölçütleri: Accuracy: 0.7988, Precision: 0.7985, Recall: 0.7988, F1 Score: 0.7986, Cohen Kappa Score: 0.6578 olarak hesaplanmıştır. Yüksek değerler, sınıf dengesizliğini göz önüne almadan eğitilen modelin iyi bir performans sergilediğini göstermektedir. Şekil 4.10'da ROC eğrisi bulunmaktadır.



Şekil 4.10. KNN Hiperparametre ayarı ve bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi

4.12. SMOTE Upsampling ile KNN

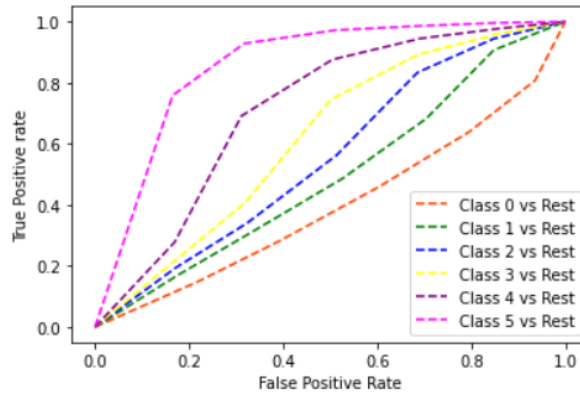
Performans ölçütleri: Accuracy: 0.7955, Precision: 0.8019, Recall: 0.7955, F1 Score: 0.7984, Cohen Kappa Score: 0.6582 olarak hesaplanmıştır. Yüksek değerler, sınıf dengesizliğinin etkili bir şekilde giderildiğini göstermektedir. Şekil 4.11’de ROC eğrisi bulunmaktadır.



Şekil 4.11. SMOTE Upsampling ile KNN için Çok Sınıflı ROC Eğrisi

4.13. Near Miss Downsampling ile KNN

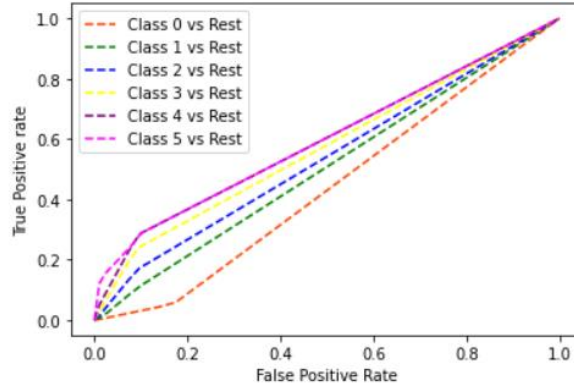
Performans ölçütleri: Accuracy: 0.2328, Precision: 0.5664, Recall: 0.2328, F1 Score: 0.2691, Cohen Kappa Score: 0.0939 olarak hesaplanmıştır. Düşük değerler boyut indirgeme işleminin KNN model performansını olumsuz etkilediğini göstermektedir. Şekil 4.12’de ROC eğrisi bulunmaktadır.



Şekil 4.12. Near Miss Downsampling ile KNN için Çok Sınıflı ROC eğrisi

4.14. Naive Bayes (Herhangi Bir Dengesizlik İşlemi Uygulanmadan)

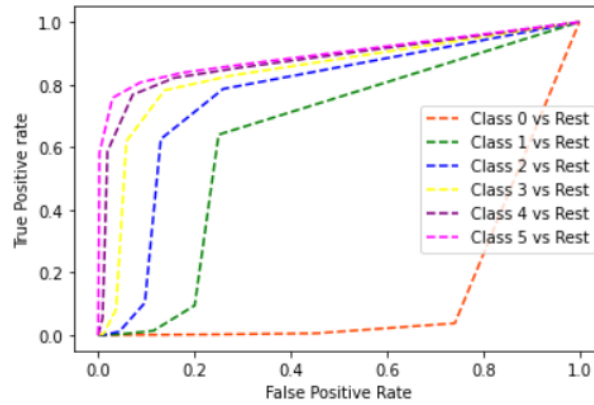
Performans ölçütleri: Accuracy: 0.5851, Precision: 0.4499, Recall: 0.5851, F1 Score: 0.4804, Cohen Kappa Score: 0.0807 olarak hesaplanmıştır. Düşük değerler herhangi bir dengesizlik işlemi uygulanmadan Naive Bayes model performansının olumsuz olduğunu göstermektedir. Şekil 4.13’de ROC eğrisi bulunmaktadır.



Şekil 4.13. Naive Bayes herhangi bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi

4.15. Random Forest (Herhangi Bir Dengesizlik İşlemi Uygulanmadan)

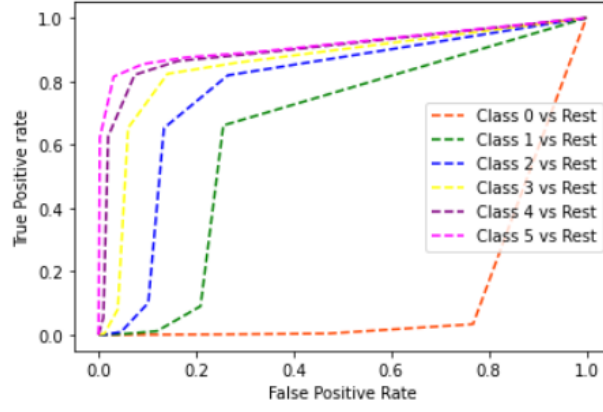
Performans ölçütleri: Accuracy: 0.7947, Precision: 0.7808, Recall: 0.7947, F1 Score: 0.7826, Cohen Kappa Score: 0.6265 olarak hesaplanmıştır. Yüksek değerler, sınıf dengesizliğini göz önüne almadan eğitilen modelin iyi bir performans sergilediğini göstermektedir. Şekil 4.14’de ROC eğrisi bulunmaktadır.



Şekil 4.14. Random Forest herhangi bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi

4.16. Hiperparametre Ayarı İle Random Forest (Herhangi Bir Dengesizlik İşlemi Uygulanmadan)

Performans ölçütleri: Accuracy: 0.8089, Precision: 0.7969, Recall: 0.8089, F1 Score: 0.7986, Cohen Kappa Score: 0.6549 olarak hesaplanmıştır. Yüksek accuracy, precision, recall ve F1 Score değerleri, sınıf dengesizliğini göz önüne almadan, hiperparametre ayarı ile eğitilen modelin iyi bir performans sergilediğini göstermektedir. Şekil 4.15’de ROC eğrisi bulunmaktadır.



Şekil 4.15. Random Forest Hiperparametre ayarı ile herhangi bir dengesizlik işlemi uygulanmadan Çok Sınıflı ROC eğrisi

5. TARTIŞMA

Bu çalışma, kuraklık tespiti için çeşitli makine öğrenimi tekniklerinin performansını değerlendirerek, sınıf dengesizliği ve boyut indirgeme yöntemlerinin model başarımına etkisini araştırmıştır. Karar ağacı modelinin dengeli veri setlerinde yüksek performans gösterdiğini ancak sınıf dengesizliği durumunda tahmin doğruluğunun önemli ölçüde düştüğünü ortaya koyulmuştur. Bu durum, gerçek dünya uygulamalarında sıkça karşılaşılan dengesiz veri dağılımının, doğrudan model performansını etkileyen kritik bir faktör olduğunu göstermektedir.

Çalışmamızda kullanılan SMOTE yöntemi, azınlık sınıfına yönelik örnekleme artırarak sınıflar arasındaki dengesizliği gidermiş ve model performansında genel bir iyileşme sağlamıştır. Özellikle, accuracy, precision, recall ve F1 score gibi metriklerde belirgin artışlar kaydedilmiştir. Buna karşılık, Near Miss yöntemiyle gerçekleştirilen downsampling işleminin model performansını düşürdüğü görülmüştür. Bu durum, sınıf dengesizliğini azaltma amacıyla yapılan aşırı veri kaybının, modelin önemli bilgileri kaçırmaya neden olabileceğini göstermektedir.

Boyut indirgeme teknikleri olan PCA ve LDA'nın etkisi ise, verinin boyutunu azaltarak modelin daha verimli hale getirilmesine katkı sağlamıştır. Ancak, bu yöntemlerin tek başına değil, sınıf dengesizliği işlemleriyle birlikte kullanıldığında daha etkili olduğu görülmüştür. Özellikle SMOTE ile birleştirildiğinde, modelin tahmin başarımının arttığı gözlemlenmiştir.

Önerilen hibrit derin öğrenme modeli, özellikle eğitim sürecinde istikrarlı bir doğruluk artışı sergilemiş, ancak test doğruluğu eğitim doğruluğuna kıyasla daha düşük kalmıştır. Bu durum, modelin belirli bir ölçüde overfitting yaptığını göstermektedir. Dolayısıyla, modelin genelleştirilebilirliğini artırmak amacıyla düzenli hale getirme (regularization) teknikleri, daha büyük ve çeşitli veri setleri kullanımı gibi önlemler alınmalıdır. Ayrıca, Cohen Kappa skoru analizi, modelin tahminlerinin rastgele olasılıklara kıyasla anlamlı olduğunu doğrulamış, ancak modelin iyileştirilebilecek yönleri olduğunu da ortaya koymuştur.

Kuraklık tahmini için geliştirilen hibrit LSTM modeli, geleneksel istatistiksel yöntemlere kıyasla daha yüksek doğruluk oranları sunarak, zaman serisi analizlerinde derin öğrenme modellerinin etkinliğini bir kez daha ortaya koymuştur. Gelecek çalışmalarda, GRU ve Transformer tabanlı modellerin karşılaştırmalı analizi yapılarak performans değerlendirmeleri genişletilebilir. Ayrıca, farklı iklim bölgeleri ve coğrafi konumlarda testler gerçekleştirilerek modelin genelleme kabiliyeti artırılabilir.

6. ÖNERİLER VE GELECEKTEKİ ÇALIŞMALAR

Bu çalışmada, kuraklık tahmini için geliştirilen modellerde sınıf dengesizliği ve boyut indirgeme tekniklerinin performansa olumlu etkisi gözlemlenmiştir. Gelecekte model doğruluğunu artırmak ve genelleme kabiliyetini geliştirmek için öneriler sunulmaktadır.

Farklı coğrafi bölgelerden ve zaman dilimlerinden elde edilen verilerle, modelin genelleme gücünü artırarak veri seti genişletilebilir. SMOTE'un yanı sıra ADASYN gibi yöntemler, sınıf dengesizliği sorununa daha etkili çözümler sunabilir. Gelişmiş veri ön işleme adımları alternatifleri kullanılabilir. SHAP gibi teknikler kullanılarak model karar mekanizması daha şeffaf hale getirilebilir. Açıklanabilirlik yöntemlerine katkıda bulunulabilir.

Son olarak, iklim değişikliği senaryoları ile model entegrasyonu yapılması, uzun vadeli kuraklık tahminlerinin daha gerçekçi hale gelmesine katkı sağlayabilir. Uygulama alanlarının genişletilmesi için erken uyarı sistemleri ve tarımsal destek programları gibi pratik uygulamalar, model sonuçlarının gerçek dünya faydasını artıracaktır.

Bu öneriler, kuraklık tahmini alanında daha kapsamlı ve güvenilir modellerin geliştirilmesine katkı sağlayacaktır.

7. SONUÇLAR

Bu araştırma, kuraklık tespiti için çeşitli makine öğrenimi tekniklerinin uygulanmasını ve sınıf dengesizliği ile başa çıkma stratejilerini incelemeyi amaçlamıştır. Çalışmamız, geniş bir özellik setini içeren meteorolojik verileri kullanarak, farklı algoritmaların performansının karşılaştırması üzerinde yoğunlaşmıştır.

Analiz sonuçları, sınıf dengesizliği olmadığında karar ağacı modelinin yüksek performans gösterdiğini ortaya koymaktadır. Ancak, gerçek dünya uygulamalarında sınıf dengesizliği sıkça karşılaşılan bir sorundur. Bu bağlamda, sınıf dengesizliği düzelten yöntemlerin uygulanması gerekmektedir.

SMOTE (Synthetic Minority Over-sampling Technique) kullanılarak yapılan upsampling işlemi, sınıflar arasındaki dengesizliği gidermekte başarılı olmuş ve modellerin genel performansını artırmıştır. Özellikle accuracy, precision, recall ve F1 score değerlerinde belirgin bir iyileşme görülmüştür.

Öte yandan, Near Miss yöntemiyle yapılan downsampling, modellerin performansını olumsuz yönde etkilemiştir. Bu durum, veri kaybının fazla olması nedeniyle sınıf temsilinin yetersiz kaldığını göstermektedir.

Ayrıca, boyut indirgeme teknikleri olan PCA (Principal Component Analysis) ve LDA (Linear Discriminant Analysis) kullanılarak elde edilen özellik alt kümesinin, modellerin performansına olan etkisi değerlendirilmiştir. Ancak, bu yöntemlerin kullanımıyla elde edilen sonuçlar genellikle, sınıf dengesizliği işlemleri sonrasında kullanıldığında daha etkili olmuştur.

Hibrit derin öğrenme modelinin doğruluk değeri, özellikle eğitim süreci boyunca istikrarlı bir şekilde artış göstermiştir. Ancak test doğruluğunun eğitim doğruluğuna göre daha düşük olması, modelin belirli bir düzeyde overfitting yaptığını göstermektedir. Precision, recall ve F1 score gibi metriklerin kullanılması, modelin sınıf dengesizliği karşısındaki performansını daha ayrıntılı bir şekilde değerlendirmeyi sağlamıştır. Cohen Kappa skoru ise modelin genel tutarlılığını ve sınıf etiketleri arasındaki anlaşmayı yansıtmakta, modelin rastgele tahminlerin ötesinde anlamlı bir performans sergilediğini göstermektedir.

Bu çalışmada SPI için, büyük ölçekli NetCDF verileri kullanılarak hibrit bir LSTM modeli geliştirilmiş ve kuraklık tahmini yapılmıştır. Model, geleneksel istatistiksel yöntemlere kıyasla daha başarılı sonuçlar vermiştir. Gelecek çalışmalar kapsamında, GRU ve Transformer tabanlı

modellerin performans karşılaştırmaları yapılabilir ve farklı coğrafi bölgelerde test edilerek modelin geliştirilebilirliği artırılabilir.

Sonuç olarak, bu çalışma, kuraklık tespiti için makine öğrenimi modellerinin performansını artırmak amacıyla sınıf dengesizliği işlemlerini ve boyut indirgeme tekniklerini içermiştir. Elde edilen sonuçlar, özellikle sınıf dengesizliği düzeltilmiş veri seti üzerinde SMOTE ve boyut indirgeme tekniklerinin kullanılmasının, karar ağacı modelinin güvenilir ve etkili bir şekilde kuraklık tespiti yapabilmesini sağladığını göstermektedir. Bu bulgular, gelecekteki kuraklık tahmini çalışmaları ve uygulamaları için önemli bir temel oluşturmaktadır.



KAYNAKLAR

- [1] B. R. Nikam, P. K. Das, and S. B. Jadhav “Assessment Of Early Season Agricultural Drought Using Remote Sensing”, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, c. XLIII-B3-2020, ss. 1691-1695, Ağu. 2020, doi: 10.5194/isprs-archives-XLIII-B3-2020-1691-2020.
- [2] H. Xu, K. Xu, and Y. Yang, “Risk assessment model of agricultural drought disaster based on grey matter-element analysis theory”, *Nat Hazards*, c. 107, sy 3, ss. 2693-2707, Tem. 2021, doi: 10.1007/s11069-021-04681-1.
- [3] L. Hao, X. Zhang, and S. Liu, “Risk assessment to China’s agricultural drought disaster in county unit”, *Nat Hazards*, c. 61, sy 2, ss. 785-801, Mar. 2012, doi: 10.1007/s11069-011-0066-4.
- [4] Y. Zhang, J. Jin, S. Jiang, S. Ning, Y. Zhou, and Z. Wu, “Quantitative Assessment Model for the Effects of Drought Mitigation on Regional Agriculture Based on an Expectation Index of Drought Mitigation Effects”, *Water*, c. 11, sy 3, Art. sy 3, Mar. 2019, doi: 10.3390/w11030464.
- [5] I. Aitkenhead, Y. Kuleshov, J. Bhardwaj, Z.-W. Chua, C. Sun, and S. Choy, “Validating a tailored drought risk assessment methodology: drought risk assessment in local Papua New Guinea regions”, *Natural Hazards and Earth System Sciences*, c. 23, sy 2, ss. 553-586, Şub. 2023, doi: 10.5194/nhess-23-553-2023.
- [6] X. Yang, H. Wang, and X. Jin, “Numerical Analysis of a Train-Bridge System Subjected to Earthquake and Running Safety Evaluation of Moving Train”, *Shock and Vibration*, c. 2016, ss. 1-15, 2016, doi: 10.1155/2016/9027054.
- [7] N. R. Bond, P. S. Lake, and A. H. Arthington, “The impacts of drought on freshwater ecosystems: an Australian perspective”, *Hydrobiologia*, c. 600, sy 1, ss. 3-16, Mar. 2008, doi: 10.1007/s10750-008-9326-z.
- [8] V. A. Myeki and Y. T. Bahta, “Determinants of Smallholder Livestock Farmers’ Household Resilience to Food Insecurity in South Africa”, *Climate*, c. 9, sy 7, Art. sy 7, Tem. 2021, doi: 10.3390/cli9070117.
- [9] A. K. Mishra and V. R. Desai, “Drought forecasting using stochastic models”, *Stoch Environ Res Ris Assess*, c. 19, sy 5, ss. 326-339, Kas. 2005, doi: 10.1007/s00477-005-0238-4.
- [10] S. Park, J. Im, D. Han, and J. Rhee, “Short-Term Forecasting of Satellite-Based Drought Indices Using Their Temporal Patterns and Numerical Model Output”, *Remote Sensing*, c. 12, sy 21, Art. sy 21, Oca. 2020, doi: 10.3390/rs12213499.

- [11] S. J. Sutanto, M. van der Weert, N. Wanders, V. Blauhut, and H. A. J. Van Lanen, “Moving from drought hazard to impact forecasts”, *Nat Commun*, c. 10, sy 1, s. 4945, Eki. 2019, doi: 10.1038/s41467-019-12840-z.
- [12] A. Devanand, J. T. Fasullo, A. S. Phillips, and R. Pincus, “Australia’s Tinderbox Drought: An extreme natural event likely worsened by human-caused climate change”, *Sci. Adv.*, c. 10, sy 10, s. eadj3460, Mar. 2024, doi: 10.1126/sciadv.adj3460.
- [13] A. Dikshit, B. Pradhan, and A. M. Alamri, “Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model”, *Science of The Total Environment*, c. 755, s. 142638, Şub. 2021, doi: 10.1016/j.scitotenv.2020.142638.
- [14] A. AghaKouchak, L. Cheng, E. M. Sadegh, A. Farahmand, and J. Teixeira, “Status and prospects for drought forecasting: opportunities in artificial intelligence and hybrid physical–statistical forecasting”, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, c. 380, sy 2238, s. 20210288, Eki. 2022, doi: 10.1098/rsta.2021.0288.
- [15] P. Aghelpour, B. Mohammadi, S. M. Biazar, O. Kisi, and Z. Sourmirinezhad, “A Theoretical Approach for Forecasting Different Types of Drought Simultaneously, Using Entropy Theory and Machine-Learning Methods”, *ISPRS International Journal of Geo-Information*, c. 9, sy 12, Art. sy 12, Ara. 2020, doi: 10.3390/ijgi9120701.
- [16] B. Li, Y. Zhang, J. Chen, and X. Wang, “Improving LSTM hydrological modeling with spatiotemporal deep learning and multi-task learning: A case study of three mountainous areas on the Tibetan Plateau”, *Journal of Hydrology*, c. 620, s. 129401, May. 2023, doi: 10.1016/j.jhydrol.2023.129401.
- [17] L. Zhang, T. Zhou, X. Chen, P. Wu, N. Christidis, and F. C. Lott, “The Late Spring Drought of 2018 in South China”, *Bulletin of the American Meteorological Society*, c. 101, sy 1, ss. S59-S64, 2020.
- [18] Q. Liu, S. Zhang, H. Zhang, Y. Bai, and J. Zhang, “Monitoring drought using composite drought indices based on remote sensing”, *Science of The Total Environment*, c. 711, s. 134585, Nis. 2020, doi: 10.1016/j.scitotenv.2019.134585.
- [19] A. F. Van Loon, “Hydrological drought explained”, *WIREs Water*, c. 2, sy 4, ss. 359-392, 2015, doi: 10.1002/wat2.1085.
- [20] G. P. Mengü, S. Anaç, ve E. Özçakal, “Kuraklık Yönetim Stratejileri”, *Ege Üniversitesi Ziraat Fakültesi Dergisi*, c. 48, sy 2, Art. sy 2, Haz. 2011.
- [21] J. Spinoni, G. Naumann, J. V. Vogt, and P. Barbosa, “The biggest drought events in Europe from 1950 to 2012”, *Journal of Hydrology: Regional Studies*, c. 3, ss. 509-524, Mar. 2015, doi: 10.1016/j.ejrh.2015.01.001.

- [22] A. Ceylan, E. Turgu, İ. İnal, A. Mollamahmutođlu, ve A. Aydođan, “Türkiye’de Son Yıllarda Gözlenen Kuraklık Hadiselerinin Deđerlendirilmesi”, Su Kaynakları, c. 2, sy 1, Art. sy 1, Mar. 2009.
- [23] T. B. McKee, N. J. Doesken, and J. Kleist, “The Relationship Of Drought Frequency And Duration To Time Scales”.
- [24] H. Salimi, E. Asadi, and S. Darbandi, “Meteorological and hydrological drought monitoring using several drought indices”, Appl Water Sci, c. 11, sy 2, s. 11, Oca. 2021, doi: 10.1007/s13201-020-01345-6.
- [25] S. M. Vicente-Serrano, S. Beguería, J. I. López-Moreno, C. Azorin-Molina, J. Peña-Gallardo, and M. Tomás-Burguera, “Performance of Drought Indices for Ecological, Agricultural, and Hydrological Applications”, Eyl. 2012, doi: 10.1175/2012EI000434.1.
- [26] Z. Zhang, W. Xu, Z. Shi, and Q. Qin, “Establishment of a Comprehensive Drought Monitoring Index Based on Multisource Remote Sensing Data and Agricultural Drought Monitoring”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, c. 14, ss. 2113-2126, 2021, doi: 10.1109/JSTARS.2021.3052194.
- [27] S. Sirdaş and Z. Şen, “Meteorolojik kuraklık modellemesi ve Türkiye uygulaması”, İTÜDERGİSİ/d, c. 2, sy 2, Art. sy 2, May. 2010, Erişim: 25 Ocak 2024. [Çevrimiçi]. Erişim adresi: http://160.75.25.161/index.php/itudergisi_d/article/view/766
- [28] M. Türkes, “Türkiye’de gözlenen ve öngörülen iklim deđişikliği, kuraklık ve çölleşme”, Ankara Üniversitesi Çevrebilimleri Dergisi, c. 4, sy 2, Art. sy 2, Eki. 2012, doi: 10.1501/Csaum_0000000063.
- [29] G. Aktürk ve O. Yildiz, “The Effect Of Precipitation Deficits On Hydrological Systems In The Çatalan Dam Basin, Turkey”, Uluslararası Muhendislik Arastırma ve Gelistirme Dergisi, ss. 10-28, Haz. 2018, doi: 10.29137/umagd.441389.
- [30] C. Ayva, A. Atalay Dutucu, ve B. Ustaoglu, “İklim Deđişikliğinin Su Kaynaklarına Etkisi Ve Uyum Önerileri: Kirazdere Havzası Örneđi”, Fırat Üniversitesi Sosyal Bilimler Dergisi, c. 33, sy 1, ss. 47-64, Oca. 2023, doi: 10.18069/firatsbed.1131015.
- [31] A. Abdullah, “Türkiye üzerindeki önemli kurak yıllar”, Co, c. 12, sy 2, ss. 101-118, 2014, doi: 10.1501/Cogbil_0000000155.
- [32] M. İ. Yüce, B. Bayram, ve A. Çakmak, “SPEI ve SPI ile Samsun İli Kuraklık Analizi”, Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi, c. 25, sy 3, ss. 285-295, Eyl. 2022, doi: 10.17780/ksujes.1108663.
- [33] N. An, M. T. Turp, ve L. Kurnaz, “İklim Deđişikliğine Bağlı Çevresel Bozulmanın Göç Kararına Etkisi: Genel Bir Bakış”, Ege Coğrafya Dergisi, c. 30, sy 2, ss. 383-403, Ara. 2021, doi: 10.51800/ecd.932879.

- [34] K. Oğuz, M. A. Pekin, H. Gürkan, E. Oğuz, and M. Coşkun, “Analyses of drought in Eastern Mediterranean basin with era-interim data”, *Anadolu Journal of Agricultural Sciences*, c. 32, sy 2, ss. 229-229, Haz. 2017, doi: 10.7161/omuanajas.321080.
- [35] O. M. KatiPoğlu, S. N. Yeşilyurt, ve H. Y. Dalkiliç, “Yeşilirmak havzasındaki hidrolojik kuraklıkların Mann-Kendall ve Yenilikçi Şen yöntemi ile trend analizi”, *Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, Oca. 2022, doi: 10.17714/gumusfenbil.1026893.
- [36] M. Ekinci, S. Örs, M. Turan, ve E. Yıldırım, “Nitrik Oksit Uygulamalarının Bitkilerde Abiyotik Stres Şartlarına Toleransı Arttırmadaki Etkileri”, *YYU J AGR SCI*, c. 28, sy 2, Art. sy 2, Haz. 2018, doi: 10.29133/yyutbd.427960.
- [37] İ. Koç, “Küresel İklim Değişikliğinin Bolu’da Bazı İklim Parametreleri ve İklim Tiplerine Etkisi”, *Bartın Orman Fakültesi Dergisi*, c. 23, sy 2, ss. 706-719, Ağu. 2021, doi: 10.24011/barofd.947981.
- [38] A. Özdemir, “İklim Değişikliğinin Havza Ölçeğinde Akım ve Sediman Miktarına Etkilerinin Değerlendirilmesi: Yuvacık Baraj Gölü Havzası”, *Jeoloji Mühendisliği Dergisi*, c. 45, sy 1, ss. 129-154, Haz. 2021, doi: 10.24232/jmd.941528.
- [39] A. K., “Descriptive Data Analysis: A Concept between Confirmatory and Exploratory Data Analysis”, *Methods of Information in Medicine*, c. 26, sy 02, 1987, doi: 10.1055/s-0038-1635488.
- [40] D. Sacha, M. Sedlmair, L. Zhang, M. A. T. F. Crnovrsanin, and D. A. Keim, “Visual Interaction with Dimensionality Reduction: A Structured Literature Analysis”, *IEEE Transactions on Visualization and Computer Graphics*, c. 23, sy 1, 2017, doi: 10.1109/tvcg.2016.2598495.
- [41] T.-H. Huang, M. L. Huang, Q. V. Nguyễn, L. Zhao, W. Huang, and J. Chen, “A Space-Filling Multidimensional Visualization (SFMDVis) for Exploratory Data Analysis”, *Information Sciences*, c. 390, 2017, doi: 10.1016/j.ins.2015.06.031.
- [42] J. T. Behrens and C.-H. Yu, “Exploratory Data Analysis”, *Handbook of Psychology*, 2003, doi: 10.1002/0471264385.wei0202.
- [43] P. Jia, Q. Zhao, X. Wu, F. Shen, K. Sun, and X. Wang, “Identification of cachexia in lung cancer patients with an ensemble learning approach”, *Front. Nutr.*, c. 11, s. 1380949, May. 2024, doi: 10.3389/fnut.2024.1380949.
- [44] A. Ferrario, B. Demiray, K. Yordanova, M. Luo, and M. Martin, “Social Reminiscence in Older Adults’ Everyday Conversations: Automated Detection Using Natural Language Processing and Machine Learning”, *J Med Internet Res*, c. 22, sy 9, s. e19133, Eyl. 2020, doi: 10.2196/19133.

- [45] B. S. Abunasser, M. R. J. AL-Hiealy, A. M. Barhoom, A. R. Almasri, and S. S. Abu-Naser, “Prediction of Instructor Performance using Machine and Deep Learning Techniques”, *IJACSA*, c. 13, sy 7, 2022, doi: 10.14569/IJACSA.2022.0130711.
- [46] J. Vamathevan, D. Clark, P. Czodrowski, A. Dunham, and A. Fisher, “Applications of machine learning in drug discovery and development”, *Nat Rev Drug Discov*, c. 18, sy 6, ss. 463-477, Haz. 2019, doi: 10.1038/s41573-019-0024-5.
- [47] Z. Abbas and S. Myeong, “Enhancing Industrial Cyber Security, Focusing on Formulating a Practical Strategy for Making Predictions through Machine Learning Tools in Cloud Computing Environment”, *Electronics*, c. 12, sy 12, s. 2650, Haz. 2023, doi: 10.3390/electronics12122650.

