

**T. C.  
İSTANBUL GELİŞİM ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

Mekatronik Mühendisliği Anabilim Dalı

**MAKİNE ÖĞRENMESİ YÖNTEMİYLE YENİLEBİLİR  
GÜNEŞ ENERJİSİ ÜRETİMİNİN METEOROLOJİK  
VERİLER KULLANILARAK TAHMİN ANALİZİ**

Yüksek Lisans Tezi

**Naciye MACİT SEZİKLİ**

Danışman

Dr. Öğr. Üyesi Ümit ALKAN

Eş Danışman

Prof. Dr. Metin ZONTUL

**İstanbul – 2023**



## TEZ TANITIM FORMU

**Yazar Adı Soyadı** : Naciye MACİT SEZİKLİ

**Tezin Dili** : Türkçe

**Tezin Adı** : Makine Öğrenmesi Yöntemiyle Yenilenebilir Güneş Enerji Üretiminin Meteorolojik Veriler Kullanılarak Tahmin Analizi

**Enstitü** : İstanbul Gelişim Üniversitesi Lisansüstü Eğitim Enstitüsü

**Anabilim Dalı** : Mekatronik Mühendisliği

**Tezin Türü** : Yüksek Lisans

**Tezin Tarihi** : 31.10.2023

**Sayfa Sayısı** : 91

**Tez** : 1. Dr. Öğr. Üyesi Ümit ALKAN

**Danışmanları** : 2. Prof. Dr. Metin ZONTUL

**Dizin Terimleri** : Güneş Enerjisi, Güneş Enerjisi Santrali, Meteorolojik Parametreler, Makine Öğrenmesi, Makine Öğrenmesi Yöntemleri, Doğrusal Regresyon, Ridge Regresyon, Çoklu Doğrusal Regresyon, Rastgele Ormanlar, Regresyon Ağaçları.

**Türkçe Özet** : Enerji sektöründe verimliliği artırmak ve enerji üretimini optimize etmek için güneş enerji panellerinden elde edilen enerji ile meteorolojik verileri kullanarak makine öğrenmesi yöntemleriyle tahminler yapılmıştır. Ayrıca, kullanılan makine öğrenmesi yöntemlerinin sonuçlarını değerlendirmek için hata performans analizleri kullanılmıştır.

**Dağıtım Listesi** : 1. İstanbul Gelişim Üniversitesi Lisansüstü Eğitim Enstitüsüne  
2. YÖK Ulusal Tez Merkezine

*Naciye*

*MACİT SEZİKLİ*

**T. C.  
İSTANBUL GELİŞİM ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

Mekatronik Mühendisliği Anabilim Dalı

**MAKİNE ÖĞRENMESİ YÖNTEMİYLE YENİLEBİLİR  
GÜNEŞ ENERJİSİ ÜRETİMİNİN METEOROLOJİK  
VERİLER KULLANILARAK TAHMİN ANALİZİ**

Yüksek Lisans Tezi

**Naciye MACİT SEZİKLİ**

Danışman

Dr. Öğr. Üyesi Ümit ALKAN

Eş Danışman

Prof. Dr. Metin ZONTUL

**İstanbul – 2023**

## BEYAN

Bu tezin hazırlanmasında bilimsel ahlak kurallarına uyulduđu, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduđu, kullanılan verilerde herhangi tahrifat yapılmadığını, tezin herhangi bir kısmının bu üniversite veya başka bir üniversitedeki başka bir tez olarak sunulmadığını beyan ederim.

Naciye MACİT SEZİKLİ

.../.../2023



**T.C.**  
**İSTANBUL GELİŞİM ÜNİVERSİTESİ**  
**LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ MÜDÜRLÜĞÜNE**

Naciye Macit Sezikli'nin "**Makine Öğrenmesi Yöntemiyle Güneş Enerji Üretiminin Meteorolojik Veriler Kullanılarak Tahmin Analizi**" adlı tez çalışması, jürimiz tarafından Mekatronik Mühendisliği anabilim dalı, Mekatronik Mühendisliği bilim dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

Başkan

*Prof. Dr. Murat BEKEN*

Üye

*Dr. Öğr. Üyesi Ümit ALKAN*  
(Danışman)

Üye

*Dr. Öğr. Üyesi Nihal ALTUNDAŞ*

Üye

*Prof. Dr. Metin ZONTUL*  
(İkinci Danışman)

**ONAY**

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylıyorum.

.../.../ 2023

*Prof. Dr. İzzet GÜMÜŞ*

Enstitü Müdürü

## ÖZET

Güneş enerji santralleri, temiz ve sürdürülebilir enerji taleplerini karşılama konusunda önemli bir rol oynar. Ancak değişken hava koşulları, mevsimsel etkiler ve benzeri faktörler, fazla üretilen enerjinin depolanması veya eksik enerjiden kaynaklanan maliyetlerle sonuçlanabilir. Bu durumlar, güneş enerjisi üretiminde verimsizliklere yol açabilir. Güneş enerjisi verilerinin meteoroloji verileri ile birleştirilmesi, makine öğrenmesi teknikleriyle güneş enerjisi üretimini tahmin etmek, sistem verimliliğini artırmak ve daha sürdürülebilir enerji stratejileri geliştirmek için önemli bir kaynak sunar.

Bu çalışmanın temel amacı, enerji sektöründe verimliliği artırmak ve enerji üretimini optimize etmek için güneş enerji panellerinden elde edilen enerji ile meteorolojik verileri kullanarak makine öğrenmesi yöntemleriyle, Rastgele Orman, Karar Ağaçları, K-En Yakın Komşu, Çoklu Doğrusal Regresyon, Ridge Regresyon, Lasso Regresyon, Doğrusal Regresyon, tahminler yapmaktır. Ayrıca, kullanılan makine öğrenmesi yöntemlerinin sonuçlarını değerlendirmek için  $R^2$ , MSE, RMSE, MAE hata performans analizleri kullanılmıştır. Bu çalışma, mevcut modellerin elde ettiği sonuçları değerlendirerek başarılarını karşılaştırmayı amaçlamaktadır. Rastgele Orman Algoritması, en yüksek  $R^2$  puanı olan 0.968420, en düşük MSE değeri olan 0.000764 ve en düşük MAE değeri olan 0.012567 ulaşarak en başarılı algoritma olarak öne çıkmaktadır. İkinci sırada ise Karar Ağaçları algoritması gelmektedir. Diğer algoritmalar başarısı sırasına göre şu şekildedir; Çoklu Doğrusal Regresyon, K-En Yakın Komşu, Ridge Regresyon, Lasso Regresyon ve Doğrusal Regresyon. Bu sonuçlar, çalışmanın temel bulgularını yansıtmaktadır.

**Anahtar Kelimeler:** Güneş Enerjisi, Güneş Enerjisi Santrali, Meteorolojik Parametreler, Makine Öğrenmesi, Makine Öğrenmesi Yöntemleri, Doğrusal Regresyon, Ridge Regresyon, Çoklu Doğrusal Regresyon, Rastgele Ormanlar, Regresyon Ağaçları, Lasso Regresyon, Karar Ağaçları, K- En Yakın Komşu.

## SUMMARY

Solar energy power plants play a significant role in meeting the demand for clean and sustainable energy. However, variable weather conditions, seasonal effects, and similar factors can result in the need for energy overproduction to be stored or lead to costs associated with energy deficiency. These situations can result in inefficiencies in solar energy production. The integration of solar energy data with meteorological data, using machine learning techniques, provides a valuable resource for predicting solar energy production, enhancing system efficiency, and developing more sustainable energy strategies.

The primary objective of this study is to enhance efficiency in the energy sector and optimize energy production by using machine learning techniques, including Random Forest, Decision Trees, K-Nearest Neighbors, Multiple Linear Regression, Ridge Regression, Lasso Regression, and Linear Regression, with energy generated from solar panels and meteorological data. Additionally, error performance analyses, such as R-squared, Mean Squared Error, Root Mean Squared Error, and Mean Absolute Error, have been employed to assess the outcomes of the applied machine learning methods.

This study aims to assess the results achieved by existing models and compare their successes. According to the research findings, the Random Forest Algorithm stands out as the most successful algorithm, achieving the highest R-squared score of 0.968420, the lowest Mean Squared Error value of 0.000764, and the lowest Mean Absolute Error value of 0.012567. The Decision Trees algorithm ranks second in performance. The other algorithms, in order of success, are Multiple Linear Regression, K-Nearest Neighbors, Ridge Regression, Lasso Regression, and Linear Regression. These results accurately reflect the fundamental findings of the study.

**Keywords:** Solar Energy, Solar Energy Plant, Meteorological Parameters, Machine Learning, Machine Learning Methods, Linear Regression, Ridge Regression, Multiple Linear Regression, Random Forests, Regression Trees, Lasso Regression, Decision Trees, K-N Nearest Neighbour



## İÇİNDEKİLER

ÖZET.....	i
SUMMARY .....	ii
İÇİNDEKİLER .....	iii
KISALTMALAR .....	v
TABLolar LİSTESİ.....	vi
GRAFİKLER LİSTESİ.....	vii
ŞEKİLLER LİSTESİ.....	viii
DENKLEMLER LİSTESİ .....	ix
RESİMLER LİSTESİ.....	x
ÖNSÖZ.....	xi
GİRİŞ .....	1

### BİRİNCİ BÖLÜM

1.1. Tezin Amacı.....	6
1.2. Tezin Kapsam ve İçeriği.....	7
1.3. Literatür Taraması .....	8

### İKİNCİ BÖLÜM MATERYAL VE YÖNTEMLER

2.1. Yazılım Araçları .....	20
2.1.1. Python.....	21
2.1.2. Anaconda Navigator .....	21
2.1.3. NumPy Kütüphanesi.....	22
2.1.4. Matplotlib Kütüphanesi .....	22
2.1.5. Pandas Kütüphanesi.....	23
2.1.6. Seaborn Kütüphanesi .....	23
2.1.7. Scikit Learn Kütüphanesi .....	24
2.1.8. SQL (Structured Query Language) .....	24
2.2. Veri Ön İşleme ve Veri Analizi .....	25
2.2.1. Verilerin Toplanması.....	25
2.2.2. Verilerin Temizlenmesi .....	27
2.2.3. Veri Setindeki Eksik/Kayıp Veriler.....	31
2.2.4. Verilerin Dönüştürülmesi .....	35
2.2.5. Verilerin Normalizasyonu .....	37
2.3. Makine Öğrenmesi Algoritmaları.....	38
2.3.1. Rastgele Orman (Random Forest ) .....	39
2.3.2. Karar Ağaçları (Decision Trees).....	41
2.3.3. Çoklu Doğrusal Regresyon (Polynomial Regression).....	41

2.3.4. K-En Yakın Komşu (K-Nearest Neighbour) .....	42
2.3.5. Ridge Regresyon (Ridge Regression) .....	43
2.3.6. Lasso Regresyon (Least Absolute Shrinkage and Selection Operator Regression) .....	44
2.3.7. Doğrusal Regresyon (Linear Regression) .....	45
2.4. Performans Değerlendirme Metrikleri.....	46
2.4.1. R-Kare ( $R^2$ , R Squared).....	46
2.4.2. Ortalama Mutlak Hata (MAE Mean Absolute Error).....	46
2.4.3. Hata Kareleri Ortalaması (MSE Mean Squared Error ) .....	47
2.4.4. Hata Karaleri Ortalaması Karekökü (RMSE Root Mean Squared Error )	48

## ÜÇÜNCÜ BÖLÜM

### ARAŞTIRMA SONUÇLARI VE TARTIŞMA

3.1. Özniteliklerin Seçimi .....	49
3.2. Veri Setinin Eğitim ve Test Olarak Bölünmesi .....	51
3.3. Makine Öğrenmesi Algoritmalarının Performansları .....	54
<b>SONUÇLAR VE ÖNERİLER .....</b>	<b>63</b>
<b>KAYNAKÇA .....</b>	<b>68</b>

## KISALTMALAR

<b>AB</b>	:	Avrupa Birliđi
<b>GES</b>	:	Güneş Enerji Santrali
<b>MAE</b>	:	Ortalama Mutlak Hata (Mean Absolute Error)
<b>MAPE</b>	:	Mutlak Hata Oranları Ortalaması (Mean Absolute Percentage Error)
<b>MSE</b>	:	Ortalama Kare Hata (Mean Squared Error)
<b>MW</b>	:	Megawatt
<b>R<sup>2</sup></b>	:	R-Kare (R Square)
<b>RMSE</b>	:	Ortalama Kare Hata Karekökü (Root Mean Squared Error)
<b>TEİAŞ</b>	:	Türkiye Elektrik İletim A. Ş.
<b>TUBİTAK MAM:</b>	:	Türkiye Bilimsel ve Teknik Araştırma Kurumu Marmara Araştırma Merkezi
<b>YSA</b>	:	Yapay Sinir Ağları

## TABLolar LİSTESİ

<b>Tablo 1.</b> Güneş enerji santralleri konumu ve verilerin miktarları.....	26
<b>Tablo 2.</b> Güneş enerji santrallerine ait parametreler ve özellikleri, verilerin miktarları.....	26
<b>Tablo 3.</b> Bilecik meteoroloji istasyonu parametreler, değişken adları, verilerin miktarları...	29
<b>Tablo 4.</b> Özniteliklere ait eksik veriler ve miktarları.....	34
<b>Tablo 5.</b> Bağımsız değişkenler arasındaki ilişki-1 .....	50
<b>Tablo 6.</b> Bağımsız değişkenler arasındaki ilişki-2.....	50
<b>Tablo 7.</b> Bağımsız değişkenler arasındaki ilişki-3.....	50
<b>Tablo 8.</b> Rastgele orman 10 katmanlı çapraz doğrulama puanlamaları.....	55
<b>Tablo 9.</b> Karar ağaçları 10 katmanlı çapraz doğrulama puanlamaları .....	57
<b>Tablo 10.</b> Çoklu Doğrusal Regresyon 5 katmanlı çapraz doğrulama puanlamaları .....	57
<b>Tablo 11.</b> K-En Yakın Komşu 3 katmanlı çapraz doğrulama puanlamaları .....	58
<b>Tablo 12.</b> Ridge regresyon 10 katmanlı çapraz doğrulama puanlamaları .....	59
<b>Tablo 13.</b> Lasso regresyon 10 katmanlı çapraz doğrulama puanlamaları.....	59
<b>Tablo 14.</b> Doğrusal regresyon 5 katmanlı çapraz doğrulama puanlamaları .....	60
<b>Tablo 15.</b> Makine öğrenmesi algoritmalarının performan karşılaştırılması .....	60

## GRAFİKLER LİSTESİ

<b>Grafik 1.</b> Özniteliklerde null miktarları.....	34
<b>Grafik 2.</b> Rastgele orman algortiması çapraz doğrulama katman sayısı karşılaştırması .....	52
<b>Grafik 3.</b> Çapraz Doğrulama katman sayısı karşılaştırmaları .....	53
<b>Grafik 4.</b> Makine öğrenmesi algortimaları gerçek-tahmin edilen değer grafikleri .....	56
<b>Grafik 5.</b> Makine öğrenmesi modelleri $r^2$ puan grafiği.....	61
<b>Grafik 6.</b> Makine öğrenmesi modelleri MAE, MSE, RMSE grafiği .....	62
<b>Grafik 7.</b> Makine öğrenmesi modelleri eğitim süreleri grafiği.....	62



## ŞEKİLLER LİSTESİ

Şekil 1. Türkiye global güneş dağılım haritası .....	2
Şekil 2. Projenin metodoloji aşamaları.....	20
Şekil 3. Kayıp/eksik veri çözümlene stratejileri.....	32
Şekil 4. Rastgele orman algoritmasına ait ağaç yapısı.....	40
Şekil 5. Korelasyon ısı haritası önce.....	51
Şekil 6. Korelasyon ısı haritası sonra.....	51



## DENKLEMLER LİSTESİ

<b>Denklem 1.</b> Min-Max normalizasyon denklemi .....	37
<b>Denklem 2.</b> GINI indeksi formülü.....	41
<b>Denklem 3.</b> Çoklu doğrusal regresyon formülü .....	42
<b>Denklem 4.</b> Öklidyen uzaklık formülü.....	43
<b>Denklem 5.</b> Ridge regresyonu .....	44
<b>Denklem 6.</b> Lasso Regresyon .....	45
<b>Denklem 7.</b> Doğrusal regresyon formülü.....	45
<b>Denklem 8.</b> R Kare formülü .....	46
<b>Denklem 9.</b> Ortalama mutlak hata formülü .....	47
<b>Denklem 10.</b> Hata kareleri ortalaması formülü .....	47
<b>Denklem 11.</b> Hata kareleri ortalaması karekökü formülü.....	48



## RESİMLER LİSTESİ

<b>Resim 1.</b> Güneş enerji santrallerinin uydu görüntüsü.....	26
<b>Resim 2.</b> GES ile meteoroloji istasyonu arasındaki mesafe.....	27
<b>Resim 3.</b> Güneş enerji santrallerinden temin edilen verinin kısmii ekran görüntüsü.....	27
<b>Resim 4.</b> Bilecik meteoroloji istasyonu toprak sıcaklığı verileri kısmi görüntüsü .....	28
<b>Resim 5.</b> Tüm verilerin toplandığı tablo.....	30
<b>Resim 6.</b> Tüm veriler .....	30
<b>Resim 7.</b> Öznitelikler ve özellikleri.....	33
<b>Resim 8.</b> Öznitelikler null miktarları .....	33
<b>Resim 9.</b> Kategorik veriler.....	36
<b>Resim 10 .</b> Kategorik verilerin dönüştürülmüş hali .....	36
<b>Resim 11.</b> Normalizasyon işleminden sonra veriler.....	38



## ÖNSÖZ

Güneş enerjisi verilerinin meteoroloji verileriyle birleştirilmesi, makine öğrenimi teknikleri kullanılarak güneş enerjisi üretimini tahmin etme, enerji sistemlerinin verimliliğini artırma ve daha sürdürülebilir enerji stratejileri geliştirme açısından önemli bir kaynak sağlar. Bu çalışma, enerji sektöründe verimliliği artırmak ve enerji üretimini optimize etmek amacıyla güneş panellerinden elde edilen enerji ile meteorolojik verileri kullanarak tahminlerde bulunmuştur.

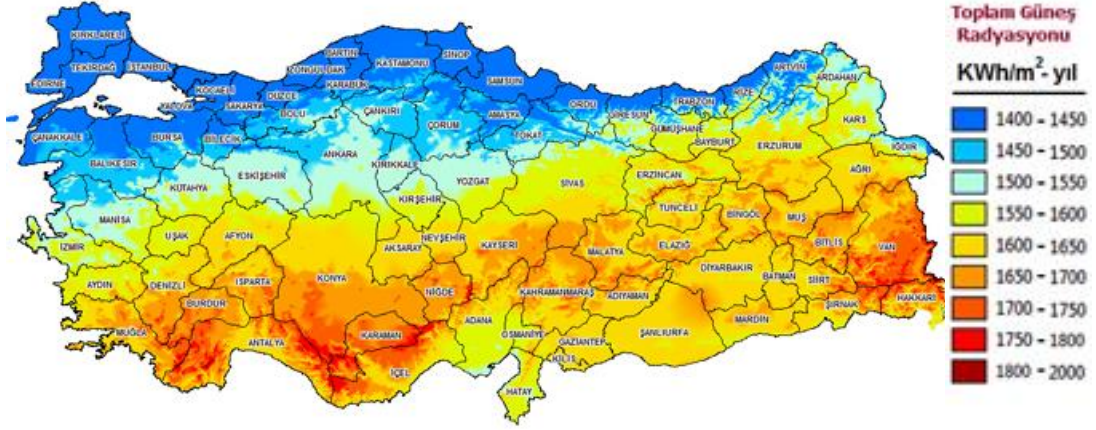
Çalışmamda kullanmam için Bilecik'te bulunan güneş enerji santrallerinin verilerini bana açan, türlü teknik destek ve bilgiyi kullanmama izin veren Kıraç Şirketi ve ETM Solar Enerji Şirketi Genel Müdürü Serkan MALÇOK'a; tez sürecinde sabır, ilgi, destek ve yardımlarından dolayı saygıdeğer tez danışmanlarım Dr. Öğr. Üyesi Ümit ALKAN, Prof. Dr. Metin ZONTUL' a ve çok kıymetli arkadaşım Zeynep ELABİAD' a; son olarakta, çalışmam boyunca beni destekleyen değerli aileme teşekkürlerimi sunarım.

## GİRİŞ

Nüfus artışı, sanayileşme ve insanların refah seviyelerinin yükselmesi gibi faktörler, enerji ihtiyacını artırmaktadır. Bu artan talep, birincil enerji kaynaklarının tükenmesine veya dışa bağımlılığımızın artmasına neden olmaktadır (Yılmaz, 2012, s. 33). Petrol, kömür, doğal gaz, nükleer enerji, hidrolik enerji, biyokütle enerjisi, dalga-gelgit enerjisi, güneş enerjisi ve rüzgâr enerjisi gibi birincil enerji kaynakları, enerji ihtiyacımızı karşılamada önde gelen kaynaklar olarak kendini göstermektedir. Ancak, bu kaynaklar sonrasında yer alan yenilenebilir enerji kaynakları, henüz yeterince kullanılmamış ve geliştirilmeyi bekleyen kaynaklarla sınırlıdır (Şenel, 2012).

Son yıllarda çevre dostu ve sınırsız bir kaynak olan güneş enerjisine olan ilgi artmıştır. Bu, alternatif enerji kaynaklarından biri olarak öne çıkmaktadır. Güneş enerjisi, çevresel sorunlara yol açmayan ve sınırsız bir rezerve sahip olan bir enerji kaynağıdır. Bu nedenle, güneş enerjisi, enerji ihtiyacımızı karşılamak için önemli bir potansiyele sahiptir.

Güneş enerjisi, dünyaya ulaşan sadece küçük bir bölümü, insanlığın enerji ihtiyacını fazlasıyla karşılayabilecek bir potansiyele sahiptir. Bu kaynağın en dikkat çekici özelliği, sınırsız ve tükenmez oluşudur. Güneş ışınları, yeryüzüne ulaştığında, bu büyük enerji potansiyelini kullanmak amacıyla çeşitli teknolojiler geliştirilmiştir. Bu teknolojiler arasında özellikle güneş panelleri, güneş kolektörleri ve güneş enerji santralleri öne çıkmaktadır. Bu teknolojiler, güneş ışınlarını hem doğrudan ısı enerjisine dönüştürebilirler hem de dolaylı olarak elektrik enerjisine çevirip kullanabilirler. Güneş enerjisi, bu yenilikçi teknolojilerin katkılarıyla temiz, sürdürülebilir ve çevre dostu bir enerji kaynağı olarak insanlığın hizmetine sunulmaktadır (Kaya ve Koç, 2015). Güneş enerjisi en fazla kullanılan ülkeler, dünya haritasında 45 derece kuzey ile 45 derece güney enlemleri arasında bulunan bölgelerde yer alan ülkelerdir (Behçet, Oral ve Gül, 2013). Adını "Güneş Kuşağı" olarak alan bu bölge, yıllık olarak 2000 ile 3500 saat arasında güneş ışığı almakta ve güneş enerjisi potansiyeli olarak 3.5 ile 7 kWh/m<sup>2</sup>/gün aralığında değerlendirilmektedir (Philibert, 2005).



Şekil 1. Türkiye global güneş dağılım haritası

Kaynak: EİE, 2012

Türkiye, Güneş Kuşağı bölgesinde konumlandığı için yüksek güneş enerjisi potansiyeline sahip olan ülkeler arasında yer almaktadır (Şekil 1). Ayrıca, Cezayir, Fas, İran, Pakistan, Suriye, Mısır, Suudi Arabistan, Libya, İspanya, İtalya, Yunanistan, İsrail, Çin, Japonya, Amerika, Meksika, Güney Afrika ve Avustralya da Türkiye gibi Güneş Kuşağı bölgesinde bulunan ülkeler arasındadır. Bu coğrafyalar, güneş enerjisi kaynaklarını en verimli şekilde kullanma potansiyeli taşıyan bölgeler olarak öne çıkmaktadır (Altuntop ve Erdemir, 2013). Türkiye'nin coğrafi konumu, güneş enerjisi açısından oldukça yüksek bir potansiyele sahip olduğunu göstermektedir. Türkiye, güneşlenme süreleri yıl içinde değişse de yıllık olarak yaklaşık 2,738 saat güneş ışığına maruz kalmaktadır. Ortalama olarak günlük 7.5 saat güneşlenme süresine sahip olan Türkiye, bu bakımdan Almanya'ya kıyasla güneş ışınlarından yüzde 60 daha fazla faydalanma avantajına sahiptir (Taktak ve İli, 2018).

Türkiye'nin güneş enerjisi potansiyeli, yapılan hesaplamalara göre en az 500 bin MW (megavat) olarak tahmin edilmektedir. 2016 yılına gelindiğinde, Türkiye'de toplamda 34 güneş enerjisi santrali bulunmakta olup bu santrallerin toplam kurulu gücü 402 MW'dır. Ayrıca, 2 güneş enerjisi santrali lisanslı olarak faaliyet göstermektedir ve toplamda 12,9 MW kurulu güce sahiptir. Lisanssız elektrik üretim santrallerinin kurulumuyla birlikte, 2016 yılı sonunda Türkiye'de 1.043 güneş enerjili santral faaliyet göstermekte ve bu santrallerin toplam kurulu gücü 819,6 MW'a ulaşmıştır. Bu verilere ek olarak, 2 adet lisanslı güneş enerjisi santrali bulunmakta ve Türkiye'nin toplam güneş enerjisi kurulu gücü 832,5 MW seviyesine çıkmıştır (Altuntop ve Erdemir, 2013).

Güneş, rüzgâr gibi yenilenebilir farklı tür enerji kaynaklarının, tükenmeyen temiz enerji kaynağı olması, bol miktarda bulunması, dışa bağımlılığı olmaması, çevre kirliliğine neden olmaması, iklim değişikliği ile mücadeleye katkı sağlaması, iş imkânlarını artırması, gelişen teknoloji ile erişebiliriliğinin artması, enerjinin sürdürülebilirliğini sağlaması gibi önemli avantajları bulunmaktadır. Bu avantajlar enerji üretimi ve kullanımında yenilenebilir kaynakların tercih edilmesini teşvik eden önemli nedenlerdir ve enerji politikalarının oluşturulmasında göz önünde bulundurulmalıdır. Bu avantajların yanı sıra, enerji kaynaklarının mevsimlere ve hava koşullarına bağlı olarak değişkenlik göstermesi, yatırım ve depolama maliyetlerinin yüksek olması, fazla enerjinin şebekeye depolanmasının gerekliliği ve eksik enerji üretiminden kaynaklanan maliyetler gibi dezavantajlar da bulunmaktadır. Bu sebeple, enerji üretiminin en etkin ve verimli şekilde gerçekleştirilmesi için çalışmaların artık bir gereklilik olduğu açıktır. Yenilenebilir enerji üretim süreçlerinde, güneş enerjisi üretimini artırmak, enerjinin mekanik sistemlere dönüşümünde verimliliği artırmak ve geleceğe yönelik kazançları tahmin etmek gibi alanlarda mekatronik, bilgisayar bilimleri ve yapay zekâ konularında çalışmalar yürütülmektedir.

Mekatronik, elektrik-elektronik, mekanik ve bilgisayar yazılımları gibi farklı disiplinleri bir araya getirerek koordinasyon sağlayıp bu disiplinlerin entegrasyonunu gerçekleştiren bir multidisipliner alandır.

Düzgün (2022), "Enerjide Yapay Zekânın Rolü Çalıştayı"nda, enerji verimliliği çalışmalarında ölçme, izleme, yönetim ve doğrulama süreçlerinin herhangi birinde dijital teknolojilerin mutlaka kullanılması gerekliliğine vurgu yapmıştır.

Makine öğrenmesi, büyük miktarda verinin işlenmesini ve bu verilerden geleceğe dönük tahminler yapılmasını sağlayarak önemli bir rol oynamaktadır. Bu büyük veri setlerinin manuel olarak işlenmesi ve analiz edilmesi imkânsızdır. Bu zorlukların üstesinden gelmek için, makine öğrenmesi yöntemleri geliştirilmiştir. Sonuç olarak, mekatronik, bilgisayar sistemleri ve makine öğrenmesi alanları, yenilenebilir enerji kaynaklarından daha verimli ve üretken enerji elde etmek için önemli bir destek sunmaktadır.

Yapay zekâ, enerji sektörünün geleceğinde önemli bir rol oynayacak ve enerji dönüşümünü hızlandırarak, bu süreçteki tüm karmaşıklıkların etkili bir şekilde yönetilmesine katkıda bulunacaktır. Yapay zekâ tabanlı uygulamalar, yenilenebilir

enerji üretimi ve tahmini, iletim ve dağıtım ağlarının işletilmesi ve optimize edilmesi, ayrıca dağıtık enerji üretimi ve enerji talebinin yönetilmesi gibi temel alanlarda öne çıkmaktadır. Dinçer, Arcaklıođlu ve Ezan (2022), bu yeni ekosistem, enerji sektöründe yapay zekâ kullanımının önemini vurgulamaktadır.

2019' da Avrupa Birliđi, Avrupa Yeşil Mutabakat strateji belgesini duyurdu. Bu mutabakat kapsamında, 2050 yılına kadar net sera gazı emisyonlarının sıfırlanması, yeşil finansmanın artırılması, temiz, ekonomik ve güvenli enerji kaynaklarının sağlanması, sürdürülebilir tarım ve akıllı ulaşım gibi yenilenebilir enerji kaynaklarının benimsenmesi ve elektrifikasyonun yaygınlaştırılması gibi uygulamalar teşvik edilmektedir (Dinçer vd., 2022, s. 18).

Güneş enerjisi üretimi için doğru hava tahminleri, güneş enerjisi santrallerinin etkili bir şekilde çalışmasını sağlamak ve enerji üretimini yönetmek açısından kritik bir öneme sahiptir. Ancak hava tahminleri, karmaşık doğal süreçlerle ilgili olduğu için beklenen sonuçlar ile gerçek sonuçlar arasında farklılıklar oluşabilir. Bu nedenle, yapay zekâ, meteoroloji alanında da önemli bir rol oynamaktadır; hava tahminlerinin doğruluđunu ve hassasiyetini artırarak, hava koşullarındaki deđişiklikleri daha hızlı ve daha iyi anlama yeteneđi sunmuştur.

Önemli bir hava durumu rapor bankasından (Watt-Sun) toplanan verileri geliştirmek için Amerika'da bir enerji bölümü IBM ile işbirliđi yaptı. Bu projenin temel amacı, güneş enerjisi üretiminin düzensizliđini, hava durumlarının deđişkenliđi ve belirsizliđi nedeniyle azaltmaktı. Bu sayede fazla güneş enerjisinin depolanması ve eksik enerji üretiminin maliyetlerinin düşürülmesi hedeflendi. Bu teknoloji, güneş ışığının deđişkenliđini kontrol edemese de tahmin edilebilirliđini artırarak tahmin doğruluđunu %30'a kadar artırma başarısını gösterdi. Güneş ışığının doğası geređi deđişken olduğu bir gerçek, ancak bu deđişkenliđi en aza indirgeme çabaları sayesinde daha öngörülebilir hale getirilebilir (PSI, 2021).

Ülkemizde enerji üretimi ve verimliliđi alanında yapılan çalışmalar desteklenmekte ve geliştirilmektedir. Ayrıca, 2021 yılında Cumhurbaşkanlıđı Dijital Dönüşüm Ofisi ve Sanayi ve Teknoloji Bakanlıđı tarafından Ulusal Yapay Zekâ Stratejisi (2021-2025) yayımlanmıştır. Bu strateji, enerji sektörüne yönelik özel hedefler içermese de yapay zekâ alanında farklı amaçlar belirlemektedir. Enerji konusunda çalışan birçok kamu ve özel sektör temsilcisi ile akademisyenlerden oluşan

geniř bir bilgi paylařım platformu, Trkiye Bilimler Akademisi (TBA) bnyesinde oluřturulan Enerji alıřma Grubu tarafından dzenlenen "TBA–Enerjide Yapay Zekânın Rol alıřtayı ve Paneli" ile kurulmuřtur (Diner vd., 2022).



# BİRİNCİ BÖLÜM

## 1.1. Tezin Amacı

Bu çalışmanın temel amacı, enerji sektöründe verimliliği artırmak ve enerji üretimini optimize etmek için güneş enerji panellerinden elde edilen enerjiler ile meteorolojik verileri kullanarak makine öğrenmesi yöntemleriyle geleceğe yönelik tahminler yapmaktır. Ayrıca, kullanılan makine öğrenmesi yöntemlerinin sonuçları üzerinde tutarlılık ve geçerlilik analizleri gerçekleştirilecektir. Mevcut modellerin elde ettiği sonuçlar değerlendirilerek başarı karşılaştırmalarına odaklanılacaktır.

Şimdiye kadar dünya genelinde ve Türkiye'de bu amaç doğrultusunda birçok çalışma, farklı coğrafi bölgelerde istatistiksel yöntemlerin yanı sıra bulanık mantık ve yapay sinir ağları gibi farklı yaklaşımlarla gerçekleştirilmiştir.

Bu çalışma, Türkiye' nin Bilecik ilinin, Bozüyük ilçesinde bulunan Cem Karakoyun Güneş Enerji Santrali (GES) , CEMKA GES ve Osman Aydın Şenol GES adlı üç güneş enerji paneli santralini ele almaktadır. Bu santraller, ETM Solar Enerji Şirketi tarafından kurulmuştur. 2018 ile 2022 yılları arasında güneş enerji panellerinde üretilen saatlik enerji miktarlarını içeren 54,000 satırlık veri, ETM Solar Enerji Şirketi tarafından çalışmamız için sağlanmıştır.

Çalışmamız için, Meteoroloji Genel Müdürlüğü, Bilecik ili için talep ettiğimiz 22 meteorolojik parametre ve her bir parametre için 30,000 satır saatlik veri de MEVBİS(Meteorolojik Veri Bilgi Satış Ve Sunum Sistemi) üzerinden sağlanmıştır.

Bu çalışma kapsamında, veri ön işleme ve veri analizi işlemleri uygulanmıştır. Ayrıca, farklı tahmin modelleri olan Rastgele Orman, Karar Ağaçları, Çoklu Doğrusal Regresyon, K-En Yakın Komşu, Ridge Regresyon, Lasso Regresyon ve doğrusal Regresyon kullanılarak enerji üretimi analizleri yapılmıştır. Elde edilen tahmin sonuçları, gerçek değerler ile karşılaştırılarak, modellerin tahmin başarısı ve gelecekteki enerji üretimini ne kadar doğru tahmin edebileceği analiz edilmiştir. Bu tez çalışması ile aşağıdaki amaçlar hedeflenmiştir: Tahmin modellerinin veri üzerindeki etkinliğini belirlemek ve enerji üretimi tahminlerinin ne kadar doğru ve güvenilir olduğunu değerlendirmek; enerji sektöründe daha etkin ve verimli enerji üretimi stratejilerinin geliştirilmesine katkı sağlamak. Bu çalışmanın özgünlüğü, daha önce kullanılmamış olan güneş enerjisi santrali (GES) verilerinin kullanılmasından,

GES verilerinin meteoroloji verileriyle birleştirilerek makine öğrenme yöntemlerinin uygulanmasından ve GES verilerinin kullanımında enerji üretiminin başlangıç ve bitiş saatlerine dayalı sınırlamaların getirilmesinden kaynaklanmaktadır. Ayrıca, farklı birçok makine öğrenme yöntemlerinin bir arada kullanılması da bu çalışmanın özgünlüğünü vurgulamaktadır.

## 1.2. Tezin Kapsam ve İçeriği

Bu çalışma altı temel aşamadan oluşmaktadır. İlk aşama, veri ön işleme aşamasını içerir. Bu aşamada, SQL kullanarak veri tabanı yönetimi ve sorgulama teknikleri ile veriler düzenlenmiş ve makine öğrenimine uygun hale getirilmiştir. Ardından null değerlerin doldurulma işlemi gerçekleştirilmiş ve bu işlemde makine öğrenmesi yöntemleri kullanılmıştır.

İkinci aşama, modelin daha güvenilir sonuçlar üretebilmesi için verilerdeki anlamsız ve hatalı değerlerin düzeltilmesini amaçlamaktadır. Bu aşamada, makine öğrenmesi yöntemleri kullanılarak bu değerlerin tamamlanması ve verilerin normalizasyon işlemi uygulanmıştır.

Veri setinin işlenmesini hızlandırmak, gereksiz ayrıntılardan arındırmak, anlamlı bilgileri öne çıkarmak, aşırı uyumu azaltmak ve böylece modelin başarısızlığını azaltmak için doğru öznitelik seçimi büyük bir öneme sahiptir. Modellerin performanslarını artırmak amacıyla otuz beş öznitelik arasından farklı makine öğrenme yöntemleri kullanılarak yirmi yedi öznitelige indirgenmiştir. Tüm bu makine öğrenme yöntemleri, Python programlama dilinde ve Anaconda dağıtımının içinde bulunan Jupyter Notebook ortamında gerçekleştirilmiştir.

Üçüncü aşamada, Rastgele Orman (Random Forest), Karar Ağaçları (Decision Trees), Çoklu Doğrusal Regresyon (Polynomial Regression), K-En Yakın Komşu (K-Nearest Neighbour), Ridge Regresyon (Ridge Regression), Lasso Regresyon (Lasso Regression) ve Doğrusal Regresyon (Linear Regression) gibi modeller kullanılarak her modelin işleyişi ve yapılandırılması incelenmiştir. Modellerimiz, eğitim ve test verileri ayrılarak eğitilmiş ve test edilmiştir. Eğitim ve test verilerinin dengeli bir şekilde dağılımını sağlamak için çapraz doğrulama yöntemi kullanılmıştır.

Dördüncü aşamada, her modelin tahmin başarısını değerlendirmek için  $R^2$  (R Square - $R^2$  Kare), MAE (Mean Absolute Error-Ortalama Mutlak Hata), MSE (Mean



Squared Error-Ortalama Kare Hata), RMSE (Root Mean Squared Error-Karekök Ortalama Kare Hata), eğitim süresi (fit time) ve test süresi (score time) gibi ölçütler kullanılmıştır. En iyi performans gösteren modelde, tahminlerin doğruluğu ve sonuçların ne kadar tutarlı olduğu gözlemlenmiştir.

Beşinci aşamada, elde edilen sonuçlar ayrıntılı bir şekilde açıklanmış ve hangi modelin en üstün performansı sergilediği belirtilmiştir. Ayrıca, bu modelin avantajları ve güvenilirliği üzerine bir analiz sunulmuştur.

Son aşamada ise, çalışmanın sonuçlarına dayalı olarak enerji üretimi tahmininde daha iyi sonuçlar elde etmek için potansiyel gelişme alanları veya alternatif modellerin kullanımı hakkında öneriler sunulmuştur. Ayrıca, gelecekte benzer çalışmalarını gerçekleştirecek olan araştırmacılara yönelik önerilerde bulunulmuştur.

### **1.3. Literatür Taraması**

Güneş enerjisi, temiz ve sürdürülebilir enerji kaynaklarının önde gelen bir bileşeni olarak kabul edilmektedir. Bilim insanları ve araştırmacılar, güneş enerjisi teknolojilerinin geliştirilmesi ve iyileştirilmesi amacıyla yoğun bir şekilde çalışmaktadırlar. Bu çalışmaların bazıları güneş enerjisi panellerinin verimliliğini artırmaya yönelikken, bazıları değişen hava koşullarının etkilerini azaltarak daha etkili ve ekonomik enerji depolama çözümleri geliştirmeyi hedeflemektedir. Diğer araştırmalar güneş panellerinin en uygun form ve konumlandırma açılarıyla ilgilenirken, bir başka grup araştırmacı ise yenilenebilir enerji kaynaklarından elde edilen verilerin daha iyi analiz edilmesi ve daha kesin tahminlerin yapılabilmesi için yapay zekâ, makine öğrenimi ve büyük veri analizi gibi teknolojileri kullanmıştır.

Ülkemizde düzenlenen "Yapay Zekânın Rolü" başlıklı panelde, yenilenebilir enerjinin enerji bağımlılığını azaltma açısından büyük bir öneme sahip olduğu vurgulanmıştır. Bu durumun yönetilmesi ve işletilmesi son derece kritik bir konu olarak belirtilmiş ve bunun çeşitli zorluklar içerdiği ifade edilmiştir. Özellikle kesintili ve değişken bir rüzgâr enerjisi kaynağının yönetimi, enerji sistem operatörleri için büyük bir zorluk oluşturmaktadır (Dinçer vd., 2022). Yeşil (2022), bu bağlamda, özellikle rüzgâr ve güneş enerjisi üretimini yapay zekâ araçlarıyla doğru bir şekilde tahmin etmenin, enerji dengeleme piyasası için büyük bir önem taşıdığı düşünülmektedir. Ülkemizde GES üretim tahmin projesi, TEİAŞ tarafından yine

TÜBİTAK-MAM işbirliğiyle yürütülmektedir. Şu an için, GES üretim tahminleri, sahadaki gerçek üretim verileri ile birlikte statik verileri kullanarak oluşturulan eğitim modelleri aracılığıyla 48 saat ve daha sonrasında güneş tahmini yapma amacını taşımaktadır (Dinçer vd., 2022).

Bu literatür taraması, GES verileri ile meteoroloji verilerinin bir araya getirilerek kullanıldığı yapay zeka ve makine öğrenimi tabanlı çalışmaları inceleyecektir. Ayrıca, bu çalışmaların enerji üretimine sağladığı katkıları değerlendireceğiz.

Shi, Lee, Liu, Yang ve Wang (2012) çalışması, Güney Çin'deki güneş enerjisi panellerinden 2010-01-13 ile 2010-10-29 tarihleri arasında elde edilen verileri kullanarak meteorolojik sistemlerden alınan maksimum sıcaklık, minimum sıcaklık ve yerel hava durumu raporlarına dayalı olarak ertesi günün ortalama sıcaklık tahminlerini içermektedir. Bu çalışma, güneş enerji panellerinden elde edilecek güç tahmini için destek vektör makinelerini kullanmış ve hava koşullarını açık, bulutlu, sisli ve yağmurlu gibi öznitelikler olarak dikkate almıştır. Araştırma sonuçlarına göre, önerilen modelin tahmin hataları, örnek güneş enerji santrallerinin kurulumu için 2.10 MW (RMSE) ve %8.64 (MRE) olarak belirlenmiştir. Bu yaklaşım, Destek Vektör Makineleri modelinin güneş enerji santrallerinin güç üretimini tahmin etmede umut verici sonuçlar sunduğunu göstermektedir.

Sivakumar, Neeraja, Jamuna Rani, Anandaram, Ramya, Padhan ve Gurusamy (2022), çalışmalarında dört küçük ölçekli güneş enerji santralinden topladıkları verileri, otuzar dakika aralıklarla kaydedilen verilerle birlikte 6 saatlik aralıklarla ölçülen bulutluluk, rüzgâr hızı, rüzgâr yönü, sıcaklık ve basınç hava durumu verileri üzerine odaklanmışlardır. Bu verileri kullanarak Lasso Regresyonu, K- En Yakın Komşu Modeli, Gradient Boosting Regression Trees (GBRT), YSA (Yapay Sinir Ağı), Persistence Model, Climatology Modelleri gibi farklı yöntemleri uygulamışlardır. Hata ölçütü olarak RMSE ve performans değerlendirme metriklerini kullanmışlardır. Elde ettikleri sonuçlara göre, YSA ve GBRT modellerinin diğer tüm yöntemlere göre en iyi performansı sergilediğini gözlemlemişlerdir.

Haba (2019) tarafından gerçekleştirilen araştırmada, üç güneş enerji santralinden elde edilen veriler, Weather Underground hava istasyonlarından alınan nem, yağış seviyesi, güneş radyasyonu ve basınç hava durumu verileri ile birleştirilmiştir. Bu çalışma, Python programlama dili kullanılarak makine öğrenme algoritmalarıyla

yürütülmüştür. Güneş enerji santrallerinin kar, kum, çamur gibi optik engelleyici malzemelerle kaplanması nedeniyle enerji üretimindeki aksaklıkları tahmin etmeye odaklanmışlardır. Bu tür aksaklıkların erken tespit edilmesinin, daha hızlı bakım faaliyetlerini mümkün kılacağına inanılmıştır.

Khalyasmaa, Eroshenko ve Tran (2020) tarafından gerçekleştirilen çalışmada, Rusya Federasyonu'ndaki bir güneş enerji santralinden her 3 dakikada bir toplanan verileri ve açık kaynak meteorolojik verilerini, Karar Ağaçları (Decision Trees) modeli ile eğitmişlerdir. Hata ölçütleri olarak RMSE ve MSE kullanmışlardır. Elde edilen sonuçlara göre, RMSE' in etkili bir ölçüt olduğu halde, bu algoritmanın çok sayıda parametreye sahip veri setlerinde düşük tahmin performansı sergilediği gözlemlenmiştir. Ayrıca, bu algoritmaların aşırı uyma eğiliminde olduğu ve bu nedenle test verilerindeki hataların önemli ölçüde artabileceği düşünülmüştür.

Özdoğan'ın (2021) çalışması, Kahramanmaraş ve Adıyaman'daki güneş enerji santrallerinden bir yıllık verileri kullanarak bölgenin sıcaklık, radyasyon ve güneşlenme değerleri ile ilişkisini incelemiştir. Bu analizi gerçekleştirmek için Eviews programında doğrusal regresyon analizi yöntemini kullanmıştır. Çalışma sonuçlarına göre, sıcaklığın güneş enerji santrallerinin üretimini doğrudan etkilediği sonucuna varılmıştır. Ayrıca, bölgenin enerji talebine göre güneş enerjisi üretiminin düşük olduğu gözlemlenmiştir. Hava durumu ve meteorolojik veriler ile enerji üretimi arasındaki ilişkiyi inceleyen çalışma, enerji üretiminin olumsuz hava koşullarından olabildiğince az etkilenmesi ve enerji tedarikinde olumsuz etkilerin minimize edilmesi için önerilerde bulunmuştur.

Bekçioğulları, Dikici, Açıkgoz ve Keçecioglu (2021) tarafından yürütülen çalışmada, güneş enerjisinin meteorolojik parametrelere bağlı olarak değiştiği ve güneş radyasyonu, sıcaklık, nem, rüzgâr hızı ve bulutluluk gibi meteorolojik faktörlerin güneş enerjisi üretimini etkilediği hipotezi üzerine odaklanılmıştır. Bu çalışma kapsamında, 2018-2019 yılları arasında Avustralya'nın Alice Springs bölgesinden elde edilen güneş enerji santrali verileri kullanılmıştır. Bu veriler, güneş enerji santralinin çıkış gücünü Ampirik Mod Ayırıştırma yöntemi ile bileşenlerine ayırıştırmak için kullanılmıştır. Çalışmanın bir diğer aşamasında, Karar Ağaçları, Destek Vektör Makinleri (Decision Vector Machine), Kernel Ridge Regresyon (Kernel Ridge Regression), Theil-Sen Regresyon ve Generalized Stepwise Regression

gibi makine öğrenme regresyon yöntemleri, tahmin performansları açısından ayrıntılı bir analize tabi tutulmuştur. Bu analizlerde  $R^2$ , RMSE ve MAE gibi performans metrikleri kullanılmıştır. Sonuçlar, tüm tahmin ufukları için Destek Vektör Makinleri modelinin en iyi tahmin performansını sergilediğini göstermiştir. Bu modeli sırasıyla Generalized Stepwise Regression, Kernel Ridge Regresyon, Theil-Sen Regresyon ve Karar Ağaçları izlemiştir. Ayrıca, gelecekteki çalışmalarda farklı ayrıştırma yöntemlerinin tahmin performansı üzerindeki etkilerinin araştırılabileceği ve daha fazla meteorolojik verinin daha kesin tahmin sonuçlarına yol açabileceği önerilmektedir.

Wolff, Kühnert, Lorenz, Kramer ve Heinemann (2016) tarafından gerçekleştirilen çalışma, 2012-2013 yılları arasında Almanya'da meteocontrol GmbH tarafından 15 dakikalık aralıklarla kaydedilen GES verileri ile Netherland Water Partnership'den alınan hava tahminleri verilerini kullanarak gerçekleştirilmiştir. Bu araştırma, makine öğrenmesi yöntemlerinden Destek Vektör Regresyonu'nu kullanarak gerçekleştirilmiştir ve hata ölçütleri olarak RMSE ve BIAS (Sapma) metriklerini kullanmışlardır. Çalışmanın ilk aşamasında, sadece güneş enerji santrali verileri kullanılarak makine öğrenmesi modeli uygulanmıştır. Ardından, hava tahmin verileri ile güneş enerji santrali verilerini birleştirerek modeli tekrar uygulamışlardır. Elde edilen sonuçlar, güneş enerji santrali ölçümlerine dayalı tahminlerin 1 saat sonrasına kadar iyi sonuçlar verdiğini, ancak hava tahmin verilerinin dâhil edildiği uygulamada tahminlerin 3 saate kadar iyi sonuçlar verdiğini göstermiştir. Ayrıca, hata oranının da %75 azaldığı gözlemlenmiştir. Gelecekteki çalışmalarda, hava tahminlerinin etkisinin daha detaylı bir şekilde araştırılması ve güneş enerji sistemlerinin tahminlerinin iyileştirilmesine yönelik çalışmaların yapılması önerilmiştir.

Mahmud, Azam, Karim, Zobaed, Shanmugam ve Mathur (2021) tarafından gerçekleştirilen araştırma, coğrafi olarak güneş enerjisi açısından zengin olan Avustralya'nın Alice Springs bölgesinde bulunan bir güneş enerjisi santralinden elde edilen verilerin tahmin edilmesi üzerine odaklanmıştır. Bu çalışmada, güneş enerjisi gücü verileri ile birlikte bağıl nem, sıcaklık, radyasyon ve günlük yağış miktarı gibi meteorolojik parametreler kullanılmıştır. Farklı makine öğrenme algoritmaları, doğrusal regresyon, çoklu doğrusal regresyon, karar ağaçları, destek vektör

makinelere, rastgele orman, uzun kısa süreli bellek ve çok katmanlı perceptron regresyon dâhil olmak üzere, güneş enerjisi tahmini için uygulanmıştır. Araştırma kapsamında, tahmin performansının değerlendirilmesi için  $R^2$ , MSE, ortalama karesel hata, medyan mutlak hata gibi çeşitli performans metrikleri detaylı bir şekilde hesaplanmıştır. Elde edilen sonuçlar, Rastgele Orman yönteminin en iyi tahmin sonuçlarını sağladığını göstermiştir. Ayrıca, bağıl nem, sıcaklık, dağınık yatay radyasyon ve küresel yatay radyasyonun güneş enerjisi üretimini önemli ölçüde etkilediği, günlük yağış miktarının ise tahmin performansında daha az önemli bir faktör olduğu bulgulanmıştır. Bulutluluk hava durumu verilerinin eğitim setine dâhil edilmesinin tahmin performansını değiştirebileceği düşünülmüştür. Ayrıca, gece meteoroloji verilerinin güneş enerjisi verilerinden çıkarılmasının hata oranını bir miktar düşürebileceği öngörülmüştür.

Yıldız ve Açıkgöz (2021) tarafından gerçekleştirilen çalışma, güneş enerjisinin birçok karmaşık parametreye bağlı olduğunu düşünerek, Kilis ilinde bulunan bir güneş enerji santrali verileri ile Kilis ilindeki Devlet Meteoroloji İşleri (TSMS) saha ölçüm istasyonundan alınan nem, radyasyon ve sıcaklık verilerini kullanarak gerçekleştirilmiştir. Bu çalışmada, güneş enerjisi tahmini için yapay sinir ağıları kullanılmıştır. Özellikle korelasyon analizi sonuçları, güneş enerji santrallerinden üretilen enerji ile nem, sıcaklık ve güneş radyasyonu arasında güçlü bir ilişki olduğunu ortaya koymuştur. Çalışmada, Learning Machine (LM), Extreme Learning Machine (ELM) ve Kernel Extreme Learning Machine (KELM) gibi farklı makine öğrenme algoritmaları kullanılmıştır. Elde edilen sonuçlara göre, KELM'in daha güçlü ve güvenilir bir tahmin performansı sağladığı gözlemlenmiştir. KELM, deneysel çalışmalar sonucunda ELM ve LM'den elde edilen R değerlerini %3,94 - %35,38 ve %2,90 - %9 oranında iyileştirmiştir. Ayrıca, üç tahmin sistemi arasında elde edilen RMSE ve MAE değerleri analiz edildiğinde, önerilen KELM'in tüm tahminlerden daha düşük hata değerlerine sahip olduğu görülmüştür. Bu nedenle, KELM ile her mevsimde daha tatmin edici tahmin sonuçlarının elde edilebileceği düşünülmüştür.

Uğuz, Oral ve Çağlayan (2019) tarafından gerçekleştirilen çalışmada, 10 farklı konumdan elde edilen güneş enerji santrali verileri ile NASA Surface Meteorology and Solar Energy'den sağlanan rakım, enlem, sıcaklık, güneşten gelen ışınım ve güneşlenme süresi gibi meteorolojik veriler kullanılmıştır. Bu verileri kullanarak,

bölgede kurulması planlanan güneş enerji santrallerinin çıkış gücünün farklı makine öğrenme metotları ile tahmin edilebilirliği araştırılmıştır. Analizler için RETScreen Clean Energy Project Analysis yazılımı kullanılmıştır. Yapılan analizler sonucunda, YSA modellerinin eğitim aşamasında en yüksek  $R^2$  değerinin 0.959 olduğu gözlemlenirken, en düşük hata değerlerinin ise MAE, MSE ve Medyan Mutlak Hata için sırasıyla 0.039, 0.002 ve 0.031 olduğu belirlenmiştir. Üç model arasında en iyi performansı YSA modelinin gösterdiği sonucuna ulaşılmıştır. Bu nedenle, YSA modelinin güneş enerjisi tahmininde daha güvenilir sonuçlar sağladığı gözlemlenmiştir.

Gök (2019) tarafından gerçekleştirilen çalışmada, Kahramanmaraş ili Elbistan ilçesinde bulunan şebekeye bağlı üç farklı Güneş Enerji Santrali'nin üretim verileri ile aynı zaman aralığındaki Küresel Tahmin Sistemi meteorolojik tahmin verilerinden elde edilen bulutluluk tahmini verileri kullanılmıştır. Bu verileri kullanarak, yapay sinir ağları ile analizler gerçekleştirilmiştir. Çalışmanın sonuçlarına göre, güneş enerjisi üretim verileri ile birlikte bulutluluk tahmini meteorolojik verilerinin kullanıldığı metodun, sadece güneş enerjisi üretim verilerinin kullanıldığı metoda göre daha başarılı olduğu tespit edilmiştir. Yani, meteorolojik verilerin kullanımı, güneş enerjisi üretimi tahminlerinin doğruluğunu artırmıştır. Bu nedenle, gelecekte yapılacak yeni çalışmalarda sıcaklık, basınç ve yağış durumu gibi diğer meteorolojik verilerin de kullanılmasının yapay zekâ metotlarının öğrenme süreçlerini geliştirebileceği düşünülmektedir.

Kıymaz (2020) tarafından gerçekleştirilen çalışmada, Konya'da bulunan dört farklı güneş enerji santralinden alınan günlük veriler detaylı bir şekilde incelenmiştir. Bu veriler üzerinde derin öğrenme, istatistiksel yöntemlerin yanı sıra, sarıma analizleri, hata performans ölçüm sistemleri, tensorflow ve keras gibi teknikler kullanılarak kapsamlı bir analiz gerçekleştirilmiştir. Tüm bu analizler, güneş enerji santrallerinin verimliliğini etkileyebilecek faktörleri detaylı bir şekilde incelemeyi amaçlamıştır. Ayrıca, İç Anadolu Bölgesi, Ege ve Akdeniz bölgelerinde kurulu olan güneş enerji santrallerinin karşılaştığı sorunlar ve bu sorunlara yönelik çıkarımlar da çalışmanın önemli bir bölümünü oluşturmuştur. Bu bağlamda, Kıymaz'ın çalışması, güneş enerjisi sektöründe verimliliği artırmak ve sorunları çözmek amacıyla değerli bilgiler sunmaktadır.

Yavuz (2018) tarafından gerçekleştirilen çalışmada, Konya bölgesindeki bir güneş enerji santralinin üretim verileri ile aynı bölge ve zaman aralığını kapsayan sıcaklık, nem, rüzgâr hızı, rüzgâr yönü, bulutluluk, radyasyon ve yağış gibi meteorolojik veriler kullanılmıştır. Veri madenciliği alanında IBM Modeler programı kullanılarak çeşitli analizler gerçekleştirilmiştir. Çalışmanın sonuçlarına göre, kurulan ve testlerde yeterli performansı gösteren modelin kısa dönemli güneş enerjisi üretim tahminlerinde kullanılabileceği tespit edilmiştir. Bu bağlamda, güneş enerjisi santrallerinin üretim tahmini yapılırken en kritik parametrelerin radyasyon, nem ve saat verileri olduğu vurgulanmıştır. Ayrıca, güneş enerji panellerinin açısı, model parametreleri gibi faktörlerin kullanılması, güneş enerji panellerinin bakım ve temizlik işlemlerinin dikkate alınması, gün ışığının çıktığı saatlerdeki enerji üretiminin özel olarak incelenmesi gibi yöntemlerin daha başarılı tahmin modellerinin geliştirilmesine katkı sağlayabileceği öne sürülmüştür.

Gabralı (2019), İstanbul ilinin Büyükçekmece ilçesinde bulunan Tarım ve Orman Bakanlığı'na ait Tarım Bilgi Sistemi istasyonuna ait 2016 yılında kaydedilen meteorolojik verileri kullanarak, bölgedeki rüzgâr ve güneş enerjisi potansiyelini ve bu potansiyeli etkileyen faktörleri mevsimsel bazda incelemiştir. Mevsimsel değişimleri gözlemlemek amacıyla her mevsimin ilk aylarını referans almıştır. Rüzgâr hızı ve güneş radyasyonu verilerini zamansal değişim grafikleriyle görselleştirmiş ve bu verilerin zaman içindeki değişimini analiz etmiştir. Çalışmada, SPSS ve MATLAB gibi istatistiksel analiz araçlarını kullanarak, dalgacık analizi ve çoklu doğrusal regresyon gibi YSA metotlarını uygulamıştır. Ayrıca, kullanılan tahmin modellerinin başarısını değerlendirmek için R, R<sup>2</sup>, RMSE, normalize hata karelerinin karekökü (nRMSE) ve ortalama mutlak hata oranları (MAPE) gibi istatistiksel yöntemleri kullanmıştır. Bu çalışmanın elde ettiği bulgular, bölgede kurulacak veya mevcut rüzgâr ve güneş enerjisi santrallerinin verimliliğini artırmak ve işletmek için önemli bilgiler sunmaktadır.

Sun, Wang, Zhang ve Zheng (2018) tarafından gerçekleştirilen çalışmada, Pekin'deki güneş radyasyonu verileri üzerinde yapay sinir ağı yöntemlerinden olan Ensemble Empirical Mode Decomposition (EEMD) kullanılmıştır. Bu veriler, gravitational search algorithm (GSA) algoritması ile optimize edilmiş ve destek vektör regresyonu (LSSVR) ile tahminler gerçekleştirilmiştir. Tahmin sonuçlarını

kümelemek için K-means yöntemi benimsenmiştir. Çalışmanın sonuçlarına göre, Differential Cumulative Evolution (DCE) öğrenme yaklaşımının, normalize kök ortalama kare hatalarına (NRMSE) göre daha düşük tahmin gücüne sahip olduğu tespit edilmiştir. Ayrıca, MAPE ve diğer karşılaştırma modellerine göre daha iyi yönlü tahminler sunarak, bir gün öncesine kadar %2,96, %2,83 ve %88,24 doğruluk oranlarına ulaşıldığı sonucuna varılmıştır. Bu nedenle, DCE öğrenme yaklaşımının güneş radyasyonu tahmini için umut verici bir yaklaşım olduğu belirlenmiştir.

Erdil ve Arçaklıoğlu (2013) çalışmasında, Türkiye Meteoroloji Müdürlüğü' den temin ettikleri verileri kullanarak günlük güneş radyasyonunu tahmin etmek için yapay sinir ağı modeli uygulamışlardır. Bu veriler, boylam, rakım, ay, günlük minimum ve maksimum nem, ortalama güneşlenme süresi, günlük minimum ve maksimum sıcaklık, günlük minimum ve maksimum basınç gibi parametreleri içermektedir. Test sonuçları  $R^2$  değerini 0.997 ve MAE ise 5.775 olarak ortaya koymuştur. Bu sonuçlar, tahmin edilen ve ölçülen değerler arasındaki farkın oldukça düşük olduğunu ve modelin günlük güneş radyasyonunu başarılı bir şekilde tahmin ettiğini göstermektedir.

Eyvazov (2019) çalışmasında, Ankara meteoroloji istasyonunda kaydedilen günlük güneş radyasyon verilerini kullanarak Ankara bölgesinin güneş enerji potansiyelini tahmin etmeyi ve güneş panellerinin verimliliğini incelemeyi amaçlamıştır. Topladığı veriler üzerinde Levenberg Marquardt, Bayesian Regularization ve Ölçekli Eşlenik Eğim algoritmalarını kullanarak Matlab uygulamasının yapay sinir ağları araçları ile analizler gerçekleştirmiştir. Modelleme aşamasında kullanılan bu üç algoritmanın, enerji potansiyeli tahminlerinde güvenilir bir şekilde kullanılabileceği sonucuna ulaşmıştır. Bu çalışma, güneş enerjisi potansiyeli ve güneş panellerinin performansının daha iyi anlaşılmasına ve gelecekteki enerji üretimi planlamalarına katkı sağlamıştır.

Oğuz (2018) çalışmasında, İstanbul Büyükşehir Belediyesi Hidayet Türkoğlu Spor Kompleksi Çatı Güneş Enerji Santrali'nden elde edilen bir yıllık verileri kullanarak, aynı şebekeye bağlı ışınım, sıcaklık ve rüzgâr hızı gibi faktörlerin güneş santrallerine olan etkisini saatlik, günlük, haftalık, aylık, mevsimlik ve yıllık bazda incelemiştir. Bu çalışma, güneş enerjisi üretimini etkileyen faktörlerin farklı zaman



dilimlerindeki deęişimini analiz ederek, güneş enerjisi sistemlerinin daha verimli bir şekilde tasarlanmasına ve işletilmesine katkı sağlamıştır.

Okur (2016) çalışmasında, Meteoroloji Genel Müdürlüğü'nden elde ettiği Akdeniz Bölgesi'nde bulunan 14 farklı veri istasyonundan toplanan çeşitli meteorolojik verileri kullanarak bir YSA modeli kullanmıştır. Bu veriler, yerleşim yeri, ay, enlem, boylam, yükseklik, ortalama sıcaklık, toprak altı ortalama sıcaklık (5 cm), bağıl nem, bulutluluk, aylık ortalama buhar basıncı, aylık toplam güneşlenme süresi, aylık ortalama toprak üstü minimum sıcaklık ve aylık ortalama günlük toplam küresel güneşlenme şiddeti gibi 13 farklı parametreyi içermektedir. MATLAB programını kullanarak bu verileri eğitmiş ve test etmiştir. Geliştirdiği YSA modeli ile yatay yüzeye düşen güneş enerjisi miktarlarını tahmin etmiş ve bu tahminleri gerçek verilerle karşılaştırmıştır. Çalışma sonuçlarına göre, tahmin edilen ve ölçülen toplam güneş ışınımı değerleri arasında önemli bir uyum sağlandığını gözlemlemiştir. Bu nedenle, YSA analizlerinin yatay düzleme düşen güneş enerjisinin belirlenmesinde etkili bir yöntem olduğu sonucuna varmıştır. Ayrıca, geliştirdiği YSA modelinin farklı yerleşim yerleri ve deęişik koşullar için de kullanılabilceğini önermiştir.

Akman (2018) çalışmasında, Orta Anadolu'daki bir elektrik iletim şebekesinden dokuz ayrı istasyona ait saatlik yük verilerini kullanmıştır. Bu çalışmada, yapay sinir aęları ve genetik algoritma tabanlı adaptif bir hibrid sistem modeli geliştirmiştir. Geliştirdiği modeli test verilerine uygulamış ve sonuçları klasik yapay sinir aęı modeli ile karşılaştırmıştır. Modelin yük tahmini performansını MAPE ölçütü ile değerlendirmiştir. Çalışmanın sonuçlarına göre, geliştirilen adaptif hibrid modelin klasik yapay sinir aęı modeline göre daha iyi performans gösterdiği görülmüştür. Bu nedenle, önerilen modelin kısa dönem yük tahminleri için daha gerçekçi sonuçlar verdiği ve doğru eğilimleri yansıttığı sonucuna varılmıştır. Akman ayrıca, daha fazla veri ve farklı tahmin yöntemleri kullanarak tahminlerin doğruluğunu daha da artırmanın mümkün olabileceğini önermiştir.

Demolli, Dokuz, Gokcek ve Ecemiş (2019), çalışmalarında Nięde bölgesindeki güneş panellerinden alınacak güç değerlerini tahmin etmek için sıcaklık, güneşlenme şiddeti ve sıcaklık verilerini kullanmışlardır. Bu çalışmanın temel amacı, gelecek yılın güneş enerjisi potansiyelini başarılı bir şekilde tahmin etmektir. Veri eğitimi sürecinde, çeşitli makine öğrenimi algoritmalarını kullanmışlardır, bunlar K-En Yakın

Komşu, Destek Vektör Makinesi ve Lasso Regresyon algoritmalarını içermektedir. Araştırmacılar, güneşlenme şiddeti ve sıcaklık verilerini kullanarak gelecek yılın güneş enerjisi potansiyelini başarılı bir şekilde tahmin ettiklerini bulmuşlardır. En iyi performansı, Destek Vektör Makinesi algoritması göstermiştir, bu algoritma en yüksek  $R^2$  skoru olan 0.997 ve en düşük hata oranı olan 0.04586 MAE değerini elde etmiştir. Ayrıca, gelecekteki çalışmalarda, diğer makine öğrenimi algoritmalarının kullanılması ve güneş enerjisi üzerinde etkili olan diğer meteorolojik parametrelerin algoritmaların girdisi olarak kullanılmasıyla elde edilen sonuçların performansa nasıl etki edeceğini araştırmayı düşünmektedirler.

Wei (2017) tarafından gerçekleştirilen bu çalışma, üç farklı veri setinin birleştirilmesiyle oluşturulmuştur. İlk veri seti, Tayvan'ın Tainan bölgesinde 2010-2016 yılları arasında 7 yıllık aralıkla toplanan saatlik atmosferik basınç, rüzgâr hızı, yağış, sıcaklık, bağıl nem ve radyasyon verilerini içermektedir. İkinci veri seti, Terra uydusu tarafından 1999 yılında yörüngeye fırlatılan MODIS (Moderate Resolution Imaging Spectroradiometer) tarafından toplanan verileri içermektedir. Bu veri seti, aerosol optik derinliği, su buharı, sirüs yansımaları ve bulut fraksiyonu gibi parametreleri içermektedir ve Dünya'nın atmosferini, okyanuslarını ve kara yüzeylerini gözlemlemek için kullanılmıştır. Son olarak, Yungkuang Hava İstasyonu için 2010-2016 döneminde toplanan saatlik güneş konumu veri seti eklenmiştir, bu veri seti sapma açısı, saat açısı, zirve açısı, yükseklik açısı ve azimut açısı gibi parametreleri içermektedir. Çalışma, dört farklı güneş radyasyonu tahmin modeli oluşturmak için çok katmanlı algılayıcı, rastgele ormanlar, k-en yakın komşu ve doğrusal regresyon (LR) gibi makine öğrenimi algoritmalarını kullanmıştır. Bu modeller, güneş panellerinin farklı eğim açılarında aldığı güneş ışınımını tahmin etmek için kullanılmıştır. Ayrıca, cihazın optimum güneş ışığı alabilmesi için en uygun eğim açılarını belirlemek için denklemler türetilmiştir. Çalışmanın bulgularına dayanarak, güneş radyasyonu tahmin modelleri arasında en düşük MAE değerini MLP'nin elde ettiği gözlenmiştir. Rastgele ormanlar ve k- en yakın komşu modelleri de iyi sonuçlar vermiş, ancak tahmin süresi arttıkça hata oranının önemli ölçüde arttığı gözlenmiştir. Doğrusal regresyon ise diğer modellere göre daha düşük performans sergilemiştir. Son olarak, güneş panelleri için en uygun eğim açısının 20-22 derece olduğu sonucuna varılmıştır. Bu, güneş panellerinin maksimum güneş ışığına en iyi şekilde maruz kalabileceği eğim açısı olarak belirlenmiştir. Bu çalışma, güneş enerjisi

üretimi için güneş radyasyonu tahmininin geliştirilmesi ve güneş panellerinin optimum konumlandırılması için önemli bilgiler sunmaktadır.

Kara (2019) tarafından gerçekleştirilen çalışma, Çorum ilinin 1983-2018 yılları arasındaki günlük güneş ışıınım verilerini incelemiştir. Bu çalışmada, derin öğrenme yöntemlerinden biri olan Long Short-Term Memory (LSTM) modeli kullanılmış ve bu modelin tahmin performansı, geleneksel makine öğrenimi modelleri olan Rastgele Orman, K-En Yakın Komşu, Gradyan Güçlendirme ve Karar Ağaçları modelleri ile karşılaştırılmıştır. Çalışmanın sonuçlarına göre, LSTM modelinin tahmin performansının geleneksel makine öğrenimi modellerine kıyasla daha etkili olduğu bulunmuştur. LSTM modeli, hata performansı değerlerinin tamamında daha iyi sonuçlar vermiştir. Diğer yandan, karar ağaçları modeli en düşük performansı göstermiş ve LSTM modelinin daha doğru tahminler sağladığı gözlemlenmiştir. Ayrıca, gelecekteki çalışmalarda sıcaklık, rüzgâr hızı, basınç gibi diğer hava durumu parametrelerinin de dâhil edilmesinin çalışmanın daha da geliştirilmesine katkı sağlayabileceği öngörülmüştür. Bu parametrelerin eklenmesi, güneş ışıınımını tahmin etme yeteneğini artırabilir ve daha kesin sonuçlar elde edilmesine yardımcı olabilir.

Tektaş ve Korkmaz (2023) çalışmalarında, 2015-2019 yılları arasında Çanakkale iline ait günlük ve saatlik güneş verilerini Meteoroloji Genel Müdürlüğü'nden elde etmişlerdir. Bu veriler, toplam yağış, nispi nem, buharlaşma saatlik, güneşlenme süresi ve aktüel basınç gibi parametreleri içermektedir. Bu verileri analiz etmek amacıyla K-En Yakın Komşular, Naive Bayes, Standartlaştırılmış Değişken Mesafeler (SVD) ve Karar Ağaçları gibi makine öğrenmesi algoritmalarını kullanarak Matlab programını kullanmışlardır. Elde ettikleri sonuçlara göre, Çanakkale güneş ışıınım verilerini işlemek için en etkili sınıflandırma yönteminin Karar Ağaçları algoritması olduğunu gözlemlenmiştir. Bu sonuçlar, bu algoritmanın veri analizi ve sınıflandırma görevlerinde diğer yöntemlere kıyasla daha iyi bir performans sergilediğini göstermektedir. Bu çalışma, Çanakkale ilindeki güneş ışıınım verilerinin analizi ve sınıflandırılması için makine öğrenmesi tekniklerinin etkili bir şekilde kullanılabileceğini göstermektedir.

## İKİNCİ BÖLÜM

### MATERYAL VE YÖNTEMLER

Güneş enerjisi, düşük çevresel etkilere sahip sürdürülebilir ve temiz bir enerji kaynağı olarak kabul edilmektedir. Bu enerji kaynağı, gelecekteki enerji taleplerini karşılama konusunda kritik bir rol oynamaktadır. Güneş enerjisi üretiminde verimliliği artırmaya yönelik yapılan çalışmalar ve güneş enerjisi tahmin çalışmaları, yenilenebilir enerji kaynaklarının daha etkili bir şekilde kullanılmasına yardımcı olmaktadır.

Güneş enerjisi üretimini etkileyen bir dizi faktör bulunmaktadır. Bunlar arasında güneş ışınlarının yoğunluğu, hava koşulları (örneğin, bulutluluk ve yağış), coğrafi konum, mevsimsel değişiklikler, günün saati, yüzey eğimi ve temizlik durumu gibi etkenler yer almaktadır. Güneş enerjisi verileri ile meteoroloji verileri bir araya getirilerek, güneş enerjisi üretiminin daha doğru bir şekilde tahmin edilmesine yardımcı olunabilir. Makine öğrenimi algoritmaları, bu verileri kullanarak hava koşullarının güneş enerjisi üretimine olan etkisini analiz edebilir ve gelecekteki enerji üretim miktarını tahmin edebilir.

Sonuç olarak, güneş enerjisi verileri ile meteoroloji verilerinin birleştirilmesi, güneş enerjisi üretimini tahmin etmek, sistemin verimliliğini artırmak ve daha sürdürülebilir enerji stratejileri geliştirmek için önemli bir kaynak sunmaktadır. Bu verilerin doğru bir şekilde analiz edilmesi ve makine öğrenimi yöntemlerinin kullanılması, güneş enerjisi sektörünün gelecekte daha verimli ve etkili bir şekilde çalışmasına katkı sağlayabilir.



**Şekil 2.** Projenin metodoloji aşamaları

Bu çalışma, güneş enerji santrallerinden elde edilen verilerin, meteoroloji verileri ile birleştirilerek makine öğrenimi yöntemleri ile analiz edileceği bir proje içermektedir. Şekil 2' deki metodoloji adımları takip edilecektir.

## 2.1. Yazılım Araçları

Makine öğrenimi ve istatistiksel analizler için kullanılan yazılım uygulamaları ve kütüphaneler, tez çalışmasının ayrılmaz bir parçasını oluşturur. Bu materyaller, analizlerin gerçekleştirilmesini, sonuçların elde edilmesini ve sonuçların açıklanmasını büyük ölçüde kolaylaştırabilir.

Bu bölümde, tez çalışmasında kullanılan uygulamalar ve yazılım kütüphaneleri, makine öğrenimi ve istatistiksel analiz için temel araçları içermektedir. Bu araçlar, tahmin çalışmalarında yaygın olarak kullanılmakta olup, zengin örnek uygulamaları ve kolay erişim imkanı sunmaları nedeniyle tercih edilmiştir.

### **2.1.1. Python**

Python programlama dili, geniş bir yelpazede veri analizi, makine öğrenimi ve derin öğrenme uygulamalarına olanak tanıdığı için sıkça tercih edilir. Ayrıca, istatistiksel analiz için numpy, pandas, scipy gibi kütüphaneler oldukça kullanışlıdır. Derin öğrenme alanında ise TensorFlow ve PyTorch gibi kütüphaneler öne çıkar. Python'un modüler yapısı, farklı veri alanlarına girişi destekler ve hemen her türlü platformda sorunsuz çalışabilir (Donat, 2015).

### **2.1.2. Anaconda Navigator**

Anaconda Navigator, veri bilimi ve makine öğrenmesi gibi alanlarda yaygın olarak kullanılan bir uygulamadır. İçerisinde bir dizi önemli kütüphane (NumPy, Pandas, Matplotlib, scikit-learn, TensorFlow, PyTorch) bulundurarak veri bilimi analizlerini destekler. Aynı zamanda Python dilinin bir dağıtımıdır ve kod yazma, arayüz oluşturma, kütüphanelerin kurulumu ve güncellenmesi gibi birçok işlemi kolaylaştırır.

Anaconda Navigator içerisinde Jupyter Notebook ve Spyder gibi araçlar da sunar. Jupyter Notebook, interaktif bir araç olarak kullanılarak hem kodların yazılmasını hem de açıklamaların belgelendirilmesini sağlar. Bu özellik, analizlerin ve sonuçların etkili bir şekilde belgelenmesi için büyük bir kolaylık sağlar. Spyder ise veri bilimi, istatistiksel analiz ve hesaplamalı bilimler için özel olarak tasarlanmış açık kaynaklı bir Python entegre geliştirme ortamıdır. Bu platform, bilimsel hesaplamalar yapmanızı, veri analizi gerçekleştirmenizi, görselleştirmeler oluşturmanızı ve makine öğrenimi projeleri geliştirmenizi kolaylaştırmak için geliştirilmiştir.

Anaconda Navigator, genellikle bilim insanları, veri bilimcileri, istatistikçiler ve mühendisler tarafından tercih edilirken, öğrenciler ve bağımsız Python geliştiricileri tarafından da kullanılabilir (Kaplan, 2019). Bu araçlar, veri analizi ve makine

öğrenimi çalışmalarını daha verimli ve etkili bir şekilde yapmak isteyen herkes için önemli kaynaklar sunar.

### **2.1.3. NumPy Kütüphanesi**

NumPy (Numerical Python), büyük veri kümeleri üzerinde hızlı ve verimli hesaplamalar yapmak için özellikle tercih edilen bir kütüphanedir. Bu kütüphane, diziler üzerinde hızlıca vektörel, matematiksel ve matris işlemleri gerçekleştirebilme yeteneği sayesinde öne çıkar. Ayrıca, farklı boyutlardaki diziler üzerinde çalışabilme, indeksleme ve dilimleme ile verilere kolay erişim sağlama kabiliyeti sunar. Lineer cebirdeki birçok matematiksel işlemi de gerçekleştirebilir ve veri okuma ve yazma gibi işlemleri de son derece kullanıcı dostu bir şekilde yapabilme özelliği taşır.

Bu nedenlerle, NumPy kütüphanesi, veri analizi, bilimsel hesaplamalar, mühendislik ve makine öğrenimi gibi geniş bir yelpazedeki alanlarda öğrenciler, analistler ve veri mühendisleri tarafından sıkça kullanılan bir araçtır. NumPy, veri odaklı işlemleri hızlandırır ve büyük veri kümeleri üzerinde etkili analizler yapmayı kolaylaştırır, bu da araştırma ve analiz projelerinin verimliliğini artırır.

### **2.1.4. Matplotlib Kütüphanesi**

Matplotlib, Python programlama dilinde 2 boyutlu sütun, dağılım, histogram, kutu, yüzey grafikleri gibi çeşitli çizimler oluşturmanıza ve bu çizimleri istediğiniz renk, etiket, gölgelendirme gibi birçok özelleştirme seçeneğiyle kişiselleştirmenize imkân tanıyan önemli bir kütüphanedir. Matplotlib, özellikle Jupyter Notebook veya Jupyter Lab gibi interaktif ortamlarda kullanım için özel olarak tasarlanmıştır. Bu kütüphane, verilerin daha anlaşılır, etkili ve yorumlanabilir hale getirilmesine yardımcı olan veri bilimcileri tarafından sıklıkla tercih edilen bir araçtır.

Matplotlib, verilerin grafiklerle görselleştirilmesi ve sunulması konusunda geniş bir esneklik sunar. Bu nedenle, tez çalışmalarında verilerin daha açıklayıcı ve etkili bir şekilde iletilmesi için kullanılan önemli bir araçtır. Ayrıca, oluşturulan grafiklerin farklı formatlarda kaydedilmesine olanak tanır, bu da sonuçların daha kolay paylaşılmasını ve yayınlanmasını sağlar.

### **2.1.5. Pandas Kütüphanesi**

Pandas kütüphanesi, veri analizi alanında çok boyutlu verilerin etkili bir şekilde yönetilmesini sağlayan önemli bir araçtır. Bu kütüphane, programlama süreçlerinde verilerin saklanması, işlenmesi ve analiz edilmesi için kullanılan DataFrame ve Series gibi veri yapıları üzerinde çeşitli işlemler yapma imkânı sunar. Pandas, veri setlerini filtreleme, grublama, birleştirme, dönüştürme ve sıralama gibi işlemleri kolayca gerçekleştirme yeteneğiyle öne çıkar. Ayrıca, eksik veya hatalı verilerle başa çıkmak için etkili bir araç sunar.

Pandas, verilere erişim için indeksleme ve dilimleme işlemlerini destekler ve istatistiksel analizler için ortalama, medyan, standart sapma gibi hazır fonksiyonlar sunar. Aynı zamanda zaman serisi verilerini işlemek için özel fonksiyonlar da içerir. Pandas, verileri çeşitli kaynaklardan okuma ve yazma işlemlerini de destekler, bu da CSV, Excel, SQL veritabanları gibi farklı veri kaynaklarıyla etkileşimli çalışabilme yeteneği sağlar.

Bu nedenlerle, Pandas kütüphanesi, tez çalışmaları dâhil olmak üzere veri analizi projelerinde sıkça tercih edilen bir araçtır ve verilerin etkili bir şekilde işlenmesini ve anlaşılmasını kolaylaştırır.

### **2.1.6. Seaborn Kütüphanesi**

Seaborn kütüphanesi, birçok verinin bulunduğu platformlarda veri analizi ve görselleştirmenin daha etkili ve çekici hale getirilmesi amacıyla geliştirilmiş güçlü bir araçtır. Bu kütüphane, çeşitli avantajlar sunar: görsel temalar ve stil seçenekleri sunma, istatistiksel analizler için özel fonksiyonlar içermesi, pandas kütüphanesi ile sorunsuz entegrasyon sağlama, ilişki grafikleri, kategorik grafikler ve zaman serileri için özel çizimler sunma, matrislerdeki ilişkili veriler için ısı haritaları oluşturma gibi özelliklere sahiptir.

Seaborn kütüphanesi, veri bilimcilerin veri analizi ve görselleştirmelerini daha etkili bir şekilde yapmalarına olanak tanır. Bu nedenle, birçok veri analizi projesinde tercih edilen bir araçtır ve verilerin daha iyi anlaşılmasına ve yorumlanmasına katkı sağlar.



### 2.1.7. Scikit Learn Kütüphanesi

Scikit-Learn, makine öğrenimi projelerini geliştirmek ve uygulamak için kullanılan güçlü ve esnek, açık kaynaklı bir Python kütüphanesidir. Bu kütüphane, veri setlerinin yüklenmesi, ön işleme yapılması, model seçimi, modellerin eğitimi, değerlendirme metrikleri ve sonuçların analizi gibi bir dizi işlemi içerir ve makine öğrenimi, veri madenciliği ve veri analizi uygulamalarında yaygın olarak kullanılır. Scikit-Learn, kullanıcıların veri setlerini yüklemeleri, ön işlem yapmaları, uygun modelleri seçmeleri, bu modelleri eğitmeleri, sonuçları değerlendirmeleri ve sınıflandırma, regresyon, kümeleme, boyut indirgeme, doğrulama gibi birçok makine öğrenimi görevini gerçekleştirmelerine olanak tanır. Ayrıca temel istatistiksel işlemler ve veri görselleştirme için de imkânlar sunar ve Numpy, SciPy ve Matplotlib gibi farklı kütüphanelerle uyumlu çalışır.

Scikit-Learn, makine öğrenimi projelerinin vazgeçilmez bir parçasıdır ve veri analizi, model geliştirme ve sonuçların yorumlanması gibi işlemleri kolaylaştırır. Bu nedenle, birçok tez çalışması ve veri odaklı araştırma projesinde kullanılan önemli bir araçtır.

### 2.1.8. SQL (Structured Query Language)

SQL programı, büyük verilere erişerek verileri almak, filtrelemek, gruplamak, sıralamak, toplamak, yeni veriler eklemek, mevcut verileri değiştirmek, güncellemek, verileri silmek, diğer veri setleri ile ilişki kurarak veri bütünlüğünü sağlamak, verileri yedeklemek ve saklamak gibi bir dizi temel veri işlemi için kullanılan önemli bir araçtır. Bu program, veritabanı yöneticileri, veri analistleri ve yazılım geliştiricileri tarafından sıklıkla kullanılır ve veri yönetimi konusunda kritik bir rol oynar.

Öte yandan, makine öğrenimi verilerden öğrenme yapmayı sağlayan bir alandır. Makine öğrenimi projeleri genellikle veri hazırlığına büyük önem verir ve veri ön işleme ve temizlik aşaması, projenin başarısı için kritik bir adımdır. Bu aşamada farklı veri setlerini birleştirme, gereksiz sütunları silme, eksik verileri doldurma veya verileri birleştirme gibi işlemler gerçekleştirilir. SQL sorguları, özellikle veri birleştirme ve temizlik işlemlerinde kullanılarak verilerin hazırlanmasına yardımcı olabilir.

Bu nedenle, SQL ve makine öğrenimi, farklı amaçlara hizmet etse de, veri yönetimi ve veri analizi süreçlerinde bir arada kullanılabilirler.

## 2.2. Veri Ön İşleme ve Veri Analizi

Makine öğrenimi, bilgisayarın sunulan verileri otomatik olarak öğrenmesini ve bu öğrenmelere dayanarak tahminlerde bulunmasını sağlayan bir yapay zeka dalıdır. Ancak makine öğrenimi algoritmalarının doğru ve verimli sonuçlar üretebilmesi için öncelikle programa sunulan çeşitli kaynaklardan temin edilen verilerin veri ön işleme aşamasından geçirilmesi gerekmektedir. Bu, kirli (dirty) verilerin, yanlış veri analizi sonuçlarına yol açmaması için hayati bir adımdır.

Veri ön işleme işlemleri, verilerin toplanması, birleştirilmesi, işlenmesi, eksik verilerin doldurulması, anormal değerlerin tespit edilip uygun formatlara çevrilmesi gibi adımları içerir. Makine öğrenimi genellikle sayısal verilerle çalıştığından, metinsel, kategorik ve zaman serisi gibi farklı türdeki verilerin uygun formatlara dönüştürülmesi de bu aşamanın bir parçasıdır. Bu dönüşüm işlemleri, metin verilerinin one-hot encoding veya label encoding gibi yöntemlerle sayısal formata çevrilmesi anlamına gelir.

Veri ön işleme aşamasının bir diğer önemli kısmı, özniteliklerin seçilmesi ve çıkarılmasıdır. Toplanan tüm verilerin makine öğrenmesine sunulması, modelin karmaşıklığını artırabilir ve bazı öznitelikler gereksiz veya düşük bilgi taşıyabilir. Bu nedenle, özniteliklerin bazıları çıkarılarak verinin sadeleştirilmesi önemlidir.

Son olarak, veri normalizasyonu veya standartlaştırma adımları, özellikler arasındaki farklılıkları dengelemek ve verileri aynı ölçekte birleştirmek için kullanılır. Bu, makine öğrenimi modellerinin daha iyi çalışmasını sağlar ve sonuçların daha güvenilir olmasını sağlar.

Tüm bu veri ön işleme adımları, makine öğrenimi projelerinin başarılı bir şekilde uygulanabilmesi için kritik bir öneme sahiptir. Bu adımların doğru bir şekilde gerçekleştirilmesi, modelin daha iyi performans göstermesine ve veri analizi sonuçlarının daha güvenilir olmasına katkı sağlar.

### 2.2.1. Verilerin Toplanması

Çalışmada, Bilecik ilinin Bozüyük ilçesine bağlı Doruk köyünde bulunan CEM GES, CEMKA GES ve Yeniçepni köyünde bulunan Osman Aydın Şenol GES güneş

enerji santralleri, ETM SOLAR ENERJİ Şirketi tarafından kurulmuştur. GES' lere ait konum bilgileri ve temin edilen verilerin miktarları Tablo 1' de verilmiştir.

**Tablo 1.** Güneş enerji santralleri konumu ve verilerin miktarları

Adı	Enlem	Boylam	Satır
CEM GES	39,9143725°	30,172552°	22506
CEMKA GES	39,9142072°	30,1709703°	22425
OAS GES	39,9141966°	30,1708696°	22348
<b>Toplam</b>			<b>67209</b>



**Resim 1.** Güneş enerji santrallerinin uydu görüntüsü

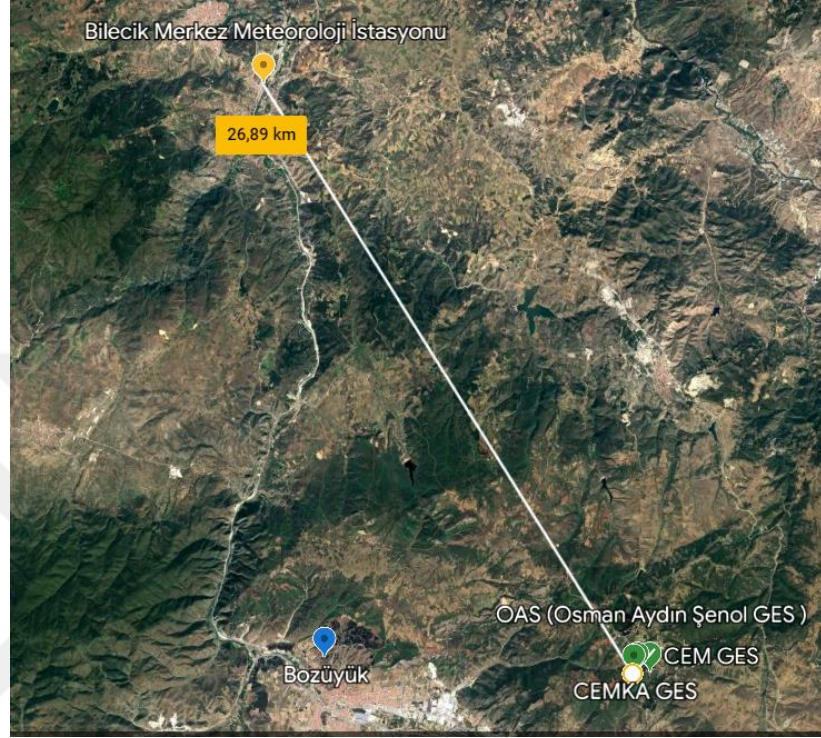
Tablo'2 de görüldüğü üzere, 09.06.2018 ile 28.02.2022 tarihleri arasında, bu üç güneş enerji santralinde saatlik periyodlarla ölçülen; Inverter Enerji Üretimi Kümülatif (MW), Inverter Enerji Üretimi (MW), Işınım Enerji (W/m<sup>2</sup>) ve Hücre Sıcaklığı (°C) parametrelerini içeren toplamda 67279 satırlık veri ve bunların çalışmadaki değişken adları paylaşılmıştır. Bu veriler senelik olacak şekilde ve her senin her ayı ayrı bir excel sayfaları olarak dosyalandırılmıştır.

**Tablo 2.** Güneş enerji santrallerine ait parametreler ve özellikleri, verilerin miktarları

Değişken Adı	Parametre Adı	Ölçüm Cinsi	Satır
TARİH(GUN, AY, YIL)	Tarih	Tarih	67279
BS_SAAT	Enerji Üretimi Başlangıç Saati	Saat	67279
BT_SAAT	Enerji Üretimi Bitiş Saati	Saat	67279
ENERJİ_KAMULATIF	Inverter Enerji Üretimi Kümülatif	MW	67279
ENERJİ	Enerji Üretimi	MW	67279
ISINIM	Işınım Enerji	W/m <sup>2</sup>	67279
HUCRE_SICAKLIK	Hücre Sıcaklığı	°C	67279

Meteoroloji Genel Müdürlüğü'nden, 40.1414° enlem, 29.9772° boylam konumunda bulunan Bilecik Merkez Meteoroloji İstasyonu'ndan, 09.06.2018 ile 28.02.2022 tarihleri arasındaki aşağıdaki 24 parametreye ait saatlik veriler, çalışmamız

için talep edilmiştir. Temin edilen veriler, her bir parametre için ayrı bir Excel belgesi olarak sunulmuştur. Toplamda, işleme alınacak 24 adet dosya bulunmaktadır. Bu istasyonun, güneş enerji santrallerine uzaklığı, Google Earth üzerinde kuşbakışı olarak Resim 2' de görüldüğü üzere 26,89 kilometredir.



**Resim 2.** GES ile meteoroloji istasyonu arasındaki mesafe

### 2.2.2. Verilerin Temizlenmesi

Dasu ve Johnson (2003) çalışmalarında, veri analiz sürecinin %80' inin veri temizleme ve hazırlama aşamasına ayrıldığını vurgulamışlardır. Resim 3' te gösterildiği gibi, güneş enerji santrallerinden elde edilen veriler saatlik periyotlarla kaydedilmiştir.

Saat	Inverter Enerji Üretimi Kümülatif	Inverter Enerji Üretimi	Işınım Enerji	Hücre Sıcaklığı º
9/6/2018--13:00 - 14:00	4.8774 MWh	0.9656 MWh	1009.2 Wh/m <sup>2</sup>	62°C
9/6/2018--14:00 - 15:00	5.8345 MWh	0.9571 MWh	946.9 Wh/m <sup>2</sup>	60°C
9/6/2018--15:00 - 16:00	6.5179 MWh	0.6834 MWh	835.9 Wh/m <sup>2</sup>	54°C
9/6/2018--16:00 - 17:00	7.1746 MWh	0.6567 MWh	664.5 Wh/m <sup>2</sup>	47°C
9/6/2018--17:00 - 18:00	7.6618 MWh	0.4872 MWh	433.4 Wh/m <sup>2</sup>	40°C
9/6/2018--18:00 - 19:00	7.8579 MWh	0.1961 MWh	194.6 Wh/m <sup>2</sup>	31°C
9/6/2018--19:00 - 20:00	7.921 MWh	0.0631 MWh	57 Wh/m <sup>2</sup>	25°C

**Resim 3.** Güneş enerji santrallerinden temin edilen verinin kısmii ekran görüntüsü

Güneş enerjisi üretiminin başlangıç saatleri her gün farklılık göstermektedir. Santraldeki veri kayıt cihazları, güneş ışınlarını tespit ederek elektrik üretimini başlattığı anda kayıtları tutmaya başlamış ve güneş ışınlarının sona erdiği ve elektrik üretiminin kesildiği anda kayıt tutmayı sonlandırmıştır. Bu durum her gün için farklı başlangıç ve bitiş saatlerinin oluşmasına neden olmuştur. Ayrıca, verilerin üretime başladığı ve sona erdiği saatler arasındaki farkların önemli ölçüde değiştiği gözlemlenmiştir.

Ayrıca, veriler bir bütün halinde değil, her ay ayrı bir sayfa üzerine kaydedilmiştir. Bu durum verilerin dağınık bir şekilde bulunmasına neden olmuş ve bunların toplanması zorunlu hale gelmiştir.

SQL programı kullanılarak, dağınık halde bulunan bu veriler tek bir tablo altında düzenlenmiştir. Tablo içerisinde CEM, CEMKA ve OAS Santralleri (YERLESKE), Tarih (TARİH), enerji üretimi başlangıç saati (BS\_SAAT), enerji üretimi bitiş saati (BT\_SAAT), Inverter Enerji Üretimi Kamulatif (ENERJI\_KAMULATIF), Inverter Enerji Üretimi (ENERJI), Işınım Enerjisi (ISINIM), Hücre Sıcaklığı (HUCRE\_SICAKLIGI) sütunları bulunmaktadır.

Meteoroloji müdürlüğünden alınan veriler, her bir parametre için ayrı dosyalarda bulunuyordu. Resim 4' te, sadece Toprak Sıcaklığı verilerinin bir görüntüsü gösterilmektedir. Meteoroloji verileri her gün 0:00 - 23:00 saatleri arasında tutulmasına rağmen her parametre ve her gün için farklılık gösterdiği gözlemlenmiştir. Aynı zamanda güneş enerji santralinden elde edilen veriler de güneş enerjisinin üretilmeye başladığı ve bittiği saatler ile sınırlılık göstermektedir.

Istasyon_No	Istasyon_Adi	YIL	AY	GUN	SAAT	TOPRAK_SICA KLIGI_5_°C
17120	BİLECİK	2018	6	1	0	18.7
17120	BİLECİK	2018	6	1	1	18.3
17120	BİLECİK	2018	6	1	2	17.9
17120	BİLECİK	2018	6	1	3	17.6
17120	BİLECİK	2018	6	1	4	17.5
17120	BİLECİK	2018	6	1	5	17.9
17120	BİLECİK	2018	6	1	6	19.2
17120	BİLECİK	2018	6	1	7	20.7
17120	BİLECİK	2018	6	1	8	22.1
17120	BİLECİK	2018	6	1	9	24.3
17120	BİLECİK	2018	6	1	10	26.8
17120	BİLECİK	2018	6	1	11	28.7
17120	BİLECİK	2018	6	1	12	29.1

**Resim 4.** Bilecik meteoroloji istasyonu toprak sıcaklığı verileri kısmi görüntüsü

**Tablo 3.** Bilecik meteoroloji istasyonu parametreler, deęişken adları, verilerin miktarları

Deęişken Adı- Parametre Id	Parametre Adı	Ölçüm Cinsi	Satır
PARAMETRE_ID_1	5 cm Toprak Sıcaklığı (Soil Temperature)	°C	30737
PARAMETRE_ID_2	Aktüel Basınç (actual pressure)	hPa	31206
PARAMETRE_ID_3	Buhar Basıncı (Vapor pressure )	hPa	30431
PARAMETRE_ID_4	Buharlařma Evapotranspirasyon (Evaporation Evapotranspiration)	mm	18699
PARAMETRE_ID_5	Bulut Taban Yükseklięi (Cloud Base Height)	m	15374
PARAMETRE_ID_6	Bulutluluk Miktarı (Cloud cover)	8 Okta	21357
PARAMETRE_ID_7	Güneşlenme Şiddeti (insolation intensity)	cal÷cm <sup>2</sup>	13559
PARAMETRE_ID_8	Güneşlenme Süresi (sunbathing duration)	saat	11810
PARAMETRE_ID_9	Islak Termometre Sıcaklığı (Wet Globe Temperature)	°C	21348
PARAMETRE_ID_10	Küresel Güneş Radyasyonu (Global Solar Radiation)	wattsaat÷m <sup>2</sup>	13559
PARAMETRE_ID_11	Maksimum Nispi Nem ( Humidity)	%	27325
PARAMETRE_ID_12	Maksimum Sıcaklık (Maximum temperature)	°C	30971
PARAMETRE_ID_13	Mevcut Kar Kalınlığı (snow thickness)	cm	2574
PARAMETRE_ID_14	Minimum Nispi Nem (Minimum Relative Humidity)	%	27324
PARAMETRE_ID_15	Minimum Sıcaklık (Minimum temperature)	°C	30971
PARAMETRE_ID_16	Nispi Nem (Relative Humidity)	%	31019
PARAMETRE_ID_17	Sıcaklık (temperature)	°C	31263
PARAMETRE_ID_18	Toplam Küresel Güneş Radyasyonu (Total Global Solar Radiation)	watt÷m <sup>2</sup>	13559
PARAMETRE_ID_19	Toplam Yaęış (Total Precipitation)	mm=kg÷m <sup>2</sup>	31064
PARAMETRE_ID_20	Toprak Üstü Minimum Sıcaklık (Above Ground Minimum Temperature)	°C	29589
PARAMETRE_ID_21	Rüzgâr Yönü (Wind Direction )	Derece	30005
PARAMETRE_ID_22	Rüzgâr Yönü Hızı (Wind Speed)	m÷sn	30005
PARAMETRE_ID_23	Maksimum Rüzgârın Yönü(Maximum Wind Direction )	Derece	27997
PARAMETRE_ID_24	Maksimum Rüzgârın Hızı (Maximum Wind Speed)	m÷sn	27997

SQL programı kullanılarak, ayrı dosyalarda bulunan meteoroloji verileri tek bir tablo altında birleştirilmiştir. Tablo 3' te görüldüğü üzere her bir meteorolojik parametre için karşısına bir kimlik (ID) atanmıştır.

Yapılan incelemeler sonucunda, meteoroloji verilerinin gün boyunca sürekli kayıt altına alındığı, ancak güneş enerji santralinde sadece enerji üretiminin olduğu saatlerde verilerin kayıt altına alındığı görülmüştür. Meteoroloji verilerindeki saat parametresindeki tutarsızlıklar nedeniyle saatlik verilerin kullanılması uygun görülmemiştir. Ayrıca, temin edilen verilerin çok fazla satır içermesi vesilesiyle, veriler SQL programı kullanılarak günlük veriler olarak düzenlenmiştir.



Mahmud ve diğerkleri (2021), gece boyunca kaydedilen meteoroloji verilerini güneş enerjisi verilerinden ayrı tutmanın, hata oranını azaltabileceğini öne sürmüşlerdir. Bu nedenle, meteoroloji verileri, güneş enerjisi üretiminin başladığı ve sona erdiği saatler arasında referans alınarak düzenlenmiştir. Örneğin, güneş enerji panellerinde gece boyunca elektrik üretimi olmadığı için gece saatlerindeki hava sıcaklığı verileri, güneş enerjisi üretimini incelemek amacıyla kullanıldığında analiz sonuçlarına yanıltıcı etkiler yaratabilir. Bu nedenle, gece boyunca kaydedilen hava sıcaklığı verilerini, aynı günün sıcaklık ortalamasına eklemek, analiz sonuçlarındaki hata miktarlarını artırabilir. Bu durum, sıcaklık ile enerji üretimi arasındaki ilişkiyi ve sonuçların doğruluğunu olumsuz etkileyebilir.

Bu nedenle, güneş enerjisi üretiminin başladığı ve bittiği saatlerdeki enerji üretimi toplamı günlük olarak kaydedilirken, aynı gün için meteoroloji verilerinde güneş enerjisi üretimi başlangıç ve bitiş saatleri referans alınarak ortalama değerleri SQL programı kullanılarak hesaplanmış ve düzenlenmiştir. Sonuç olarak, verilerin son hali Resim 5' te gösterilmiştir.

YERLESKE	GUN	AY	YIL	BS_SAAAT_MIN	BT_SAAAT_MAX	ISINIM_ORT	HUCRE_SICAKLIĞI_ORT	PARAMETRE_ID_1	PARAMETRE_ID_2	PARAMETRE_ID_3	PARAMETRE_ID_4	PARAMETRE_ID_17
QAS	3	2	2020	9	22	229.646153	11.538461	6.221428	950.678571	7.664285	0.114285	1750
QAS	4	2	2020	9	22	209.238461	12.692307	7.65	945.821428	10.607142	0	1535
QAS	5	2	2020	9	22	231.130769	12.692307	8.192857	933.442857	8.592857	0.35	942
QAS	7	2	2020	12	22	57.94	-2.1	3.60909	950.636363	4.9	0.027272	927
QAS	8	2	2020	9	22	182.676923	-3.461538	2.607142	961.135714	3.757142	0.028571	942
QAS	9	2	2020	9	22	346.923076	3.153846	2.164285	966.107142	3.5	0.028571	1857
QAS	10	2	2020	9	22	481.684615	10.615384	1.564285	959.307142	3.671428	0	NULL
QAS	11	2	2020	9	22	229.284615	6.76923	1.45	949.321428	5.464285	0	1185
QAS	12	2	2020	9	22	119.615384	4.153846	3.028571	953.964285	7.271428	0	971
QAS	13	2	2020	9	22	446.661538	15.76923	4.635714	961.078571	5	0.014285	1333
QAS	14	2	2020	9	22	281.523076	6.615384	3.6	957.692857	5.857142	0.121428	2500
QAS	15	2	2020	9	22	60.115384	5.120769	5.120769	956.940857	7.564285	0	017

Resim 5. Tüm verilerin toplandığı tablo

SQL'de tek bir tabloda bir araya getirilen veriler, makine öğrenmesi için kullanılmak üzere ".csv" uzantılı dosyalara dönüştürülmüş ve Jupyter Notebook programı kullanılarak içe aktarılmıştır.

YERLESKE	GUN	AY	YIL	BS_SAAAT	BT_SAAAT	ISINIM_ORT	HUCRE_SICAKLIK	PARAMETRE_ID_1	PARAMETRE_ID_2	...	PARAMETRE_ID_17	PA
0	CEMKA	9	6	2018	13	22	460.888888	40.111111	31.680000	947.280000	...	30.380000
1	CEMKA	10	6	2018	5	22	320.452941	31.588235	28.782352	943.258823	...	24.982352
2	CEMKA	11	6	2018	6	22	487.768750	30.000000	28.105882	945.641176	...	22.829411
3	CEMKA	12	6	2018	5	22	474.023529	29.705882	28.011111	946.022222	...	22.333333
4	CEMKA	13	6	2018	5	22	468.964705	34.764705	28.944444	942.705555	...	24.133333
...	...	...	...	...	...	...	...	...	...	...	...	...
3867	QAS	28	12	2021	8	22	232.542857	10.071428	4.753333	950.406666	...	9.333333
3868	QAS	29	12	2021	8	22	214.878571	9.071428	5.533333	947.413333	...	9.500000
3869	QAS	30	12	2021	8	22	206.985714	9.285714	6.293333	948.340000	...	10.146666
3870	QAS	31	12	2021	8	22	323.300000	15.642857	5.466666	956.006666	...	6.713333
3871	QAS	1	1	2022	8	22	64.428571	8.142857	6.360000	954.493333	...	6.046666

3872 rows x 34 columns

Resim 6. Tüm veriler

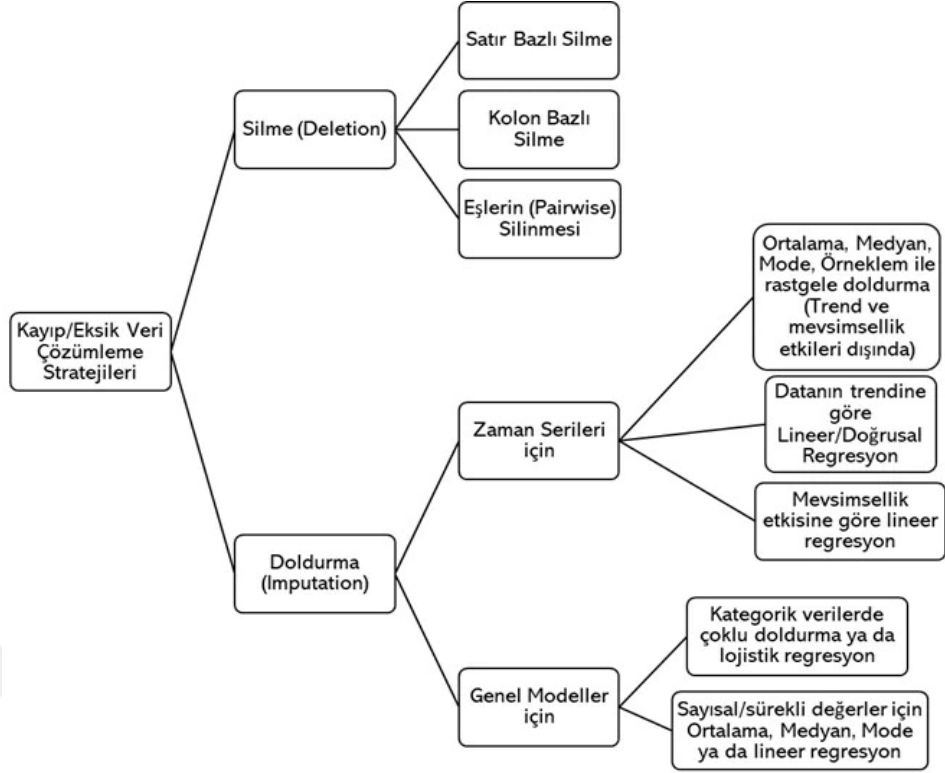
Yukarıda Resim 6' da görülen tabloda, CEM, CEMKA ve OAS "YERLESKE" sütunu; gün, ay ve yıl için ayrı ayrı "GÜN", "AY", "YIL" sütunları; o gün için elektrik üretimine başlama saati "BS\_SAAT\_MIN" sütunu; o gün için elektrik üretiminin sonlandığı saat "BT\_SAAT\_MAX" sütunu; güneş enerji santral verilerinden gelen ışınım miktarları "ISINIM" sütunu; güneş enerji santral verilerinden gelen hücre sıcaklıkları "HUCRE\_SICAKLIGI" sütunu ve meteorolojiden gelen veriler, Resim 6' da belirtilen parametre sıralamasına göre "PARAMETRE\_ID\_1" ile "PARAMETRE\_ID\_24" arasında adlandırılmış sütunlardan oluşmaktadır. Yukarıda Resim 6' da görüldüğü üzere toplam 3872 satır ve 34 sütundan oluşan verilerimiz diğer işlemler için hazır hale getirilmiştir.

### **2.2.3. Veri Setindeki Eksik/Kayıp Veriler**

Veri setlerinin toparlanma sürecinden sonra karşılaşılan ilk sorun, eksik veya kayıp verilerdir (missing-values). Eksik veriler, birçok nedenden kaynaklanabilir, örneğin gözlemlerin kaydedilmemesi, kayıt cihazlarının arızalanması veya verilerin bozulması gibi. Eksik, yanlış veya gürültülü veriler, makine öğrenmesi modellerinin performansını olumsuz etkileyebilir ve birçok makine öğrenmesi algoritması eksik verileri işlemekte zorlanır (Ye, Wang, Li, Gao ve Cheng, 2016). Bu nedenle eksik veya kayıp verilerin düzenlenmesi, makine öğrenmesi projelerinde başarılı sonuçlar elde etmek için kritik bir adımdır (Leeuw, 2009).

Eksik verilerin ele alınması için ilk yaklaşım, kayıp verilere sahip gözlemleri yok saymaktır. Ancak bu yöntem, eksik verilerin çok olduğu veya az gözlem içeren veri setlerinde yanlış sonuçlara yol açabilir (Metin, 2021). Bu tür veri setlerinde eksik değerlerin yerine yeni değerler atamak daha iyi bir yaklaşım olabilir (Sezgin, 2013). Eksik verilerin belirli bir desen oluşturmadığı durumlarda, farklı çözüm yöntemleri önerilmektedir (Carpita ve diğerleri, 2011). Bu yöntemler, verileri silme, yaklaşık değerlerle doldurma (ortalama, medyan vb.) ve model tabanlı tahmin yöntemleri olarak sınıflandırılabilir (Demir ve diğerleri, 2011).





Şekil 3. Kayıp/eksik veri çözümleme stratejileri  
Kaynak : Şener, 2020

Eksik veri tamamlama için kullanılabilecek yöntemler Şekil 3 'te gösterilmiştir.

```

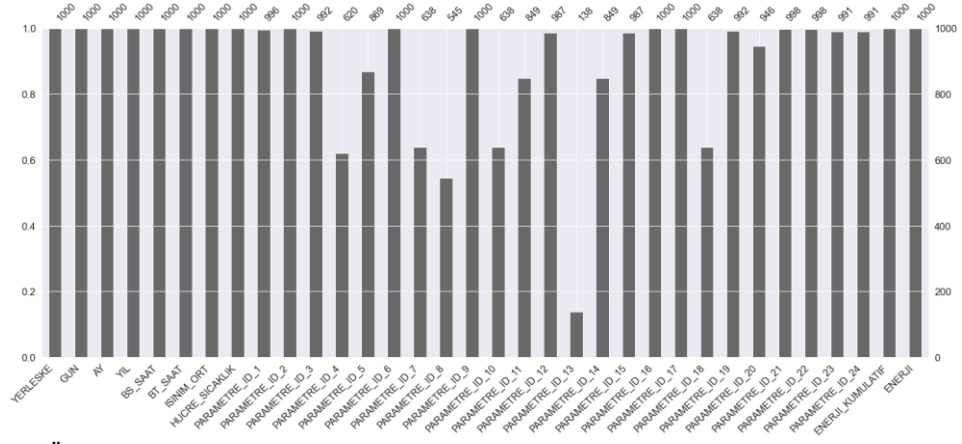
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3872 entries, 0 to 3871
Data columns (total 35 columns):
#   Column                Non-Null Count  Dtype
---  -
0   YERLESKE              3872 non-null   object
1   GUN                   3872 non-null   int64
2   AY                    3872 non-null   int64
3   YIL                   3872 non-null   int64
4   BS_SAAAT              3872 non-null   int64
5   BT_SAAAT              3872 non-null   int64
6   ISINIM_TOP            3872 non-null   float64
7   ISINIM_ORT            3872 non-null   float64
8   HUCRE_SICAKLIK        3872 non-null   float64
9   PARAMETRE_ID_1        3848 non-null   float64
10  PARAMETRE_ID_2        3872 non-null   float64
11  PARAMETRE_ID_3        3839 non-null   float64
12  PARAMETRE_ID_4        2329 non-null   float64
13  PARAMETRE_ID_5        3368 non-null   float64
14  PARAMETRE_ID_6        3872 non-null   float64
15  PARAMETRE_ID_7        2392 non-null   float64
16  PARAMETRE_ID_8        2077 non-null   float64
17  PARAMETRE_ID_9        3872 non-null   float64
18  PARAMETRE_ID_10       2392 non-null   float64
19  PARAMETRE_ID_11       3376 non-null   float64
20  PARAMETRE_ID_12       3824 non-null   float64
21  PARAMETRE_ID_13       549 non-null    float64
22  PARAMETRE_ID_14       3376 non-null   float64
23  PARAMETRE_ID_15       3824 non-null   float64
24  PARAMETRE_ID_16       3872 non-null   float64
25  PARAMETRE_ID_17       3872 non-null   float64
26  PARAMETRE_ID_18       2392 non-null   float64
27  PARAMETRE_ID_19       3839 non-null   float64
28  PARAMETRE_ID_20       3680 non-null   float64
29  PARAMETRE_ID_21       3868 non-null   float64
30  PARAMETRE_ID_22       3868 non-null   float64
31  PARAMETRE_ID_23       3833 non-null   float64
32  PARAMETRE_ID_24       3833 non-null   float64
33  ENERJİ_KUMULATIF      3872 non-null   float64
34  ENERJİ                 3872 non-null   float64
dtypes: float64(29), int64(5), object(1)
memory usage: 1.0+ MB

```

**Resim 7.** Öznitelikler ve özellikleri

**Resim 8.** Öznitelikler null miktarları

Eksik verilerle ilgili işlem yapmadan önce ön keşif yapılması önemlidir. Veri setinin kaç gözlemden ve kaç değişkenden oluştuğu, değişkenlerin veri tipleri, eksik verilerin miktarları gibi çeşitli bilgileri gözlemlemek gerekmektedir. Yukarıda Resim 7' de sunulan verilere ilişkin öznitelik isimleri, boş olmayan hücrelerin sayısı ve verilerin veri tipleri gözlemlenmiştir. Ayrıca, yukarıda Resim 8' de her bir özniteliğin içindeki eksik veri sayısı ve özniteliklerdeki eksik verilere ilişkin sütun grafiği de aşağıdaki Grafik 1' de sunulmuştur. Bu grafik, siyah alanların dolu verileri, gri alanların ise eksik verileri temsil ettiği şekilde düzenlenmiştir. Grafikte her sütunun en üstünde yer alan değerler, o değişkendeki gözlem sayısını göstermektedir. Meteorolojik verilerde 19 öznitelikte eksik veriler bulunmaktadır. Eksik verilerin bulunduğu öznitelik sütunları ve bu sütunlardaki eksik veri miktarları aşağıda Tablo 4' te sunulmuştur.



**Grafik 1.** Özniteliklerde null miktarları

**Tablo 4.** Özniteliklere ait eksik veriler ve miktarları

NO	Değişken Adı	Parametre Adı	Eksik Veri
1	PARAMETRE_ID_1	5 cm Toprak Sıcaklığı	24
2	PARAMETRE_ID_3	Buhar Basıncı	33
3	PARAMETRE_ID_4	Buharlaştırma Evapotranspirasyon	1543
4	PARAMETRE_ID_5	Bulut Taban Yüksekliği	504
5	PARAMETRE_ID_7	Güneşlenme Şiddeti	1480
6	PARAMETRE_ID_8	Güneşlenme Süresi	1795
7	PARAMETRE_ID_10	Küresel Güneş Radyasyonu	1480
8	PARAMETRE_ID_11	Maksimum Nispi Nem	496
9	PARAMETRE_ID_12	Maksimum Sıcaklık	48
10	PARAMETRE_ID_13	Mevcut Kar Kalınlığı	3323
11	PARAMETRE_ID_14	Minimum Nispi Nem	496
12	PARAMETRE_ID_15	Minimum Sıcaklık	48
13	PARAMETRE_ID_18	Toplam Küresel Güneş Radyasyonu	480
14	PARAMETRE_ID_19	Toplam Yağış	33
15	PARAMETRE_ID_20	Toprak Üstü Minimum Sıcaklık	192
16	PARAMETRE_ID_21	Rüzgâr Yönü	4
17	PARAMETRE_ID_22	Rüzgâr Hızı	4
18	PARAMETRE_ID_23	Maksimum Rüzgârın Yönü	39
19	PARAMETRE_ID_24	Maksimum Rüzgârın Hızı	39

Bu tabloda en yüksek null sayısı "Mevcut Kar Kalınlık" özniteliğinde (3323) bulunmaktadır. Bilecik, Bozüyük bölgesi tipik olarak yoğun kar yağışının görülmediği bir bölge olduğu için sadece kar yağışının olduğu günler için kayıtlar alınmış ve kar yağışının olmadığı günler için kayıtlar eksik bırakılmıştır. Bu nedenle, bu sütunun %70'ten fazlası eksik verilerle dolu olmasına rağmen, bu sütunu tamamen silmek uygun bir yaklaşım olmayabilir. Bu eksik veriler için kar kalınlık miktarı 0 olarak atanmıştır. Bu işlemle, eksik verilere yönelik bir değer ataması yapılmış ve veri setinin tutarlılığı sağlanmıştır.

Diğer sütunlardaki eksik verilerin doldurulması için kullanılan yöntemlerden ortalama, medyan ve mod değerleri atama yöntemi seçilmiştir. Ancak güneş enerji

üretiminin güneş ışığına bağlı olarak değişmesi ve farklı mevsimlerde hatta farklı aylarda güneş ışığının değişkenlik göstermesi hata miktarını artırabileceğinden, bu eksik veriler için daha özgün bir yaklaşım tercih edilmiştir. Bu nedenle, bütün sütunlar için eksik veriler, ilgili kayıp verinin bulunduğu aylardaki verilerin ortalamaları alınarak doldurulmuştur.

Örneğin, 5 cm Toprak Sıcaklığı sütunu için eksik bir veri bulunduğu anda, bu veri ilgili ayın tüm 5 cm toprak sıcaklığı verilerinin ortalamasıyla doldurulmuştur. Bu yaklaşımın, bütün sütunlar için daha doğru sonuçlar verdiği ve  $R^2$  Skorunda bir iyileşme sağladığı gözlemlenmiştir. Dolayısıyla, 5 cm Toprak Sıcaklığı, Buhar Basıncı, Buharlaşma Evapotranspirasyon, Bulut Taban Yüksekliği, Güneşlenme Şiddeti, Güneşlenme Süresi, Küresel Güneş Radyasyonu, Maksimum Nispi Nem, Maksimum Sıcaklık, Minimum Nispi Nem, Minimum Sıcaklık, Toplam Küresel Güneş Radyasyonu, Toplam Yağış, Toprak Üstü Minimum Sıcaklık, Rüzgâr Yönü, Rüzgâr Hızı, Maksimum Rüzgâr Hızı ve Maksimum Rüzgâr Yönü sütunları için yukarıda belirtilen eksik verilerin bulunduğu ayların ortalamaları kullanılarak doldurulmuştur.

#### **2.2.4. Verilerin Dönüştürülmesi**

Makine öğrenimi modelleri genellikle sayısal verilerle daha etkili bir şekilde çalışır. Bu nedenle, kategorik (sınıfsal) verilerin sayısal biçime dönüştürülmesi, modelin bu verileri anlamasına yardımcı olmak için önemlidir. Kategorik verilerin sayısal verilere dönüştürülmesi için birkaç farklı yöntem kullanılabilir. Bu yöntemler arasında ordinal encoder, label encoder ve one-hot encoder gibi popüler yöntemler bulunmaktadır.

Tek-Çok Kodlaması (One-Hot Encoding), sırasız (nominal) kategorik verilerin dönüştürülmesi için yaygın olarak kullanılan bir yöntemdir. Bu yöntem, her kategorik veriyi ayrı bir sütun olarak ele alır ve her bir sütuna ait gözlemlere 1 veya 0 gibi boolean değerler atar. Her sütun, orijinal kategorik değişkenin bir seviyesini temsil eder. Bu sayede, kategorik verilerin farklı seviyeleri sayısal bir biçimde temsil edilir (Uyanık ve Kasapbaşı, 2021). Örneğin, bir renk kategorisi düşünelim: Kırmızı, Yeşil ve Mavi. One-hot encoding kullanıldığında, bu renkler ayrı ayrı sütunlarda temsil

edilir. Her bir sütun, belirli bir rengin varlığını (1) veya yokluğunu (0) gösterir (Potdar ve diğ., 2017).

Bu dönüşüm sayesinde, makine öğrenimi modelleri kategorik verileri anlayabilir ve bu verileri kullanarak daha iyi tahminler yapabilir. One-hot encoding, kategorik değişkenlerin farklı seviyelerini birbirinden bağımsız olarak ele alır ve bu nedenle sırasız (nominal) veriler için uygundur. Ayrıca, herhangi bir sıralama veya öncelik durumu olmadan kategorik verileri temsil etmek için kullanışlıdır. Sonuç olarak, kategorik verilerin sayısal verilere dönüştürülmesi, makine öğrenimi projelerinde önemli bir adımdır ve uygun dönüşüm yöntemi seçilmelidir. One-hot encoding ve sıcak kodlama gibi yöntemler, sırasız kategorik verileri başarılı bir şekilde temsil etmek için yaygın olarak kullanılan yöntemlerdir.

	YERLESKE	GUN	AY	YIL
0	CEMKA	9	6	2018
1	CEMKA	10	6	2018
2	CEMKA	11	6	2018
3	CEMKA	12	6	2018
4	CEMKA	13	6	2018
...	...	...	...	...
3867	QAS	28	12	2021
3868	QAS	29	12	2021
3869	QAS	30	12	2021
3870	QAS	31	12	2021
3871	QAS	1	1	2022

**Resim 9.** Kategorik veriler

	CEMKA	CEM	OAS	GUN	AY	YIL
0	1.0	0.0	0.0	9	6	2018
1	1.0	0.0	0.0	10	6	2018
2	1.0	0.0	0.0	11	6	2018
3	1.0	0.0	0.0	12	6	2018
4	1.0	0.0	0.0	13	6	2018
...	...	...	...	...	...	...
3867	0.0	0.0	1.0	28	12	2021
3868	0.0	0.0	1.0	29	12	2021
3869	0.0	0.0	1.0	30	12	2021
3870	0.0	0.0	1.0	31	12	2021
3871	0.0	0.0	1.0	1	1	2022

**Resim 10 .** Kategorik verilerin dönüştürülmüş hali

Bu çalışmadaki veri setinde, kategorik olan tek sütun "YERLESKE" sütununu içermektedir (Resim 9). Bu sütunun içeriği, güneş enerji santrallerinin isimlerini içermektedir, yani "CEM", "CEMKA" ve "OAS" gibi veriler bulunmaktadır. Bu sütunu işlemek için Resim 10' da görüldüğü gibi One-Hot Encoding yöntemi uygulanmıştır. Bu işlem sonucunda her bir kategori ayrı bir sütun haline getirilmiş ve her gözlem için bu kategoriler 0 ve 1'lerle ilişkilendirilmiştir.

Örneğin, "YERLESKE" sütununda "CEM" kategorisi için yeni bir sütun oluşturulmuş ve bu sütunun değeri "CEM" olan gözlemler için 1 olarak işaretlenmiş, diğer gözlemler için ise 0 olarak işaretlenmiştir. Aynı işlem "CEMKA" ve "OAS" kategorileri için de uygulanmıştır.

Bu dönüşüm sayesinde, "YERLESKE" sütunundaki kategorik veriler, makine öğrenimi modelleri tarafından daha iyi anlaşılabilir hale gelir. Her bir kategori ayrı bir özellik olarak ele alınır ve bu özelliklerin her biri gözlemler için 0 ve 1'lerle temsil

edilir. Bu şekilde, kategorik veriler sayısal verilere dönüştürülerek modelin daha iyi performans göstermesi sağlanır.

### 2.2.5. Verilerin Normalizasyonu

Normalizasyon, genellikle makine öğrenmesi için veri hazırlığı sürecinin bir parçası olarak kullanılan bir tekniktir (Gültepe, 2019). Özellikler arasındaki farklılıkları dengelemek için veriyi normalleştirmek veya standartlaştırmak gerekebilir. Bu işlem, sütunlardaki özelliklerin değerlerini değiştirmeden, diğer sütunlarla aynı ölçekte olmalarını sağlar. Yani normalizasyon işlemi, verilerin farklı sütunlardaki değer aralıkları nedeniyle bir sütunun diğerlerine üstünlük kurmasını veya modele ağırlıklı olarak etki etmesini önlemek amacıyla kullanılır. Bu işlem, verileri belirli bir aralığa sıkıştırmak veya standart sapmaları birim olacak şekilde ölçeklendirmek gibi farklı yöntemlerle gerçekleştirilebilir. Bu şekilde normalizasyon, makine öğrenimi modellerinin daha tutarlı sonuçlar üretmesine yardımcı olur ve farklı özelliklerin modeldeki etkilerini dengelemek için önemli bir adımdır.

Makine öğrenmesi için her veri kümesini normalleştirmek gerekli olmayabilir.(Gültepe, 2019). Normalleştirme, verilerin boyutunu azaltarak sonuçların anlamlı ve kolayca yorumlanabilir olmasını ve işlemlerin uygun aralıklarda gerçekleştirilmesini sağlamak amacıyla kullanılan bir yöntemdir (Dondurmacı ve Çınar, 2014). Birçok veri normalizasyon çeşidi bulunmaktadır. Bunlar minimum-maksimum (min-max), ondalık ölçeklendirme, z-skor ve sigmoid gibi sıralanabilir (Jayalakshmi ve Santhakumaran, 2011).

Bu çalışmada, veri seti değerlerini 0 ile 1 arasında normalize etmek için min-max normalizasyon yöntemi uygulanmıştır. Min-Max normalizasyonu için Denklem(1) kullanılmıştır (Yapraklı ve Erdal, 2016). Verilerin normalize edilmiş hali aşağıda Resim 10' de gösterilmektedir.

$$x_{yeni} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

**Denklem 1.** Min-Max normalizasyon denklemi

**Kaynak:** Gültepe, 2019.

Bu denklemde;

$x_{yeni}$  : Dönüştürülmüş değerler,  
 $x$  : Gözlem değerleri ,  
 $x_{min}$  : En küçük gözlem değeri ,  
 $x_{max}$  : En büyük gözlem değeri

	CEMKA	CEM	OAS	GUN	AY	YIL	BS_SAAAT	BT_SAAAT	ISINIM	HUCRE_SICAKLIK	...	PARAMETRE_ID_8	PARAMETRE_ID_9	I
0	1.0	0.0	0.0	0.266667	0.454545	0.00	0.615385	1.0	0.620939	0.885983	...	0.514036	0.878285	
1	1.0	0.0	0.0	0.300000	0.454545	0.00	0.000000	1.0	0.431735	0.717794	...	0.514036	0.892265	
2	1.0	0.0	0.0	0.333333	0.454545	0.00	0.076923	1.0	0.657154	0.686453	...	0.514036	0.827106	
3	1.0	0.0	0.0	0.366667	0.454545	0.00	0.000000	1.0	0.638635	0.680649	...	0.514036	0.778770	
4	1.0	0.0	0.0	0.400000	0.454545	0.00	0.000000	1.0	0.631820	0.780478	...	0.514036	0.818813	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3867	0.0	0.0	1.0	0.900000	1.000000	0.75	0.230769	1.0	0.313297	0.293187	...	0.057784	0.448329	
3868	0.0	0.0	1.0	0.933333	1.000000	0.75	0.230769	1.0	0.289498	0.273453	...	0.195024	0.454963	
3869	0.0	0.0	1.0	0.966667	1.000000	0.75	0.230769	1.0	0.278865	0.277682	...	0.108346	0.459301	
3870	0.0	0.0	1.0	1.000000	1.000000	0.75	0.230769	1.0	0.435571	0.403132	...	0.050562	0.433019	
3871	0.0	0.0	1.0	0.000000	0.000000	1.00	0.230769	1.0	0.086802	0.255129	...	0.000000	0.424598	

3872 rows × 27 columns

**Resim 11.** Normalizasyon işleminden sonra veriler

### 2.3. Makine Öğrenmesi Algoritmaları

Makine Öğrenmesi algoritmaları üç kategori altında sınıflandırılmıştır. Bu sınıflandırma öğrenme stilleri temel alınarak yapılmıştır. Bunlar Gözetmeli/ Denetimli Öğrenme (Supervised Learning), Gözetmesiz/ Denetimsiz Öğrenme (Unsupervised Learning) ve Takviyeli/ Kuvvetlendirmeli Öğrenme (Reinforcement Learning) yöntemleridir (Filiz, 2017; J. H. Lee vd., 2018). Denetimli Öğrenmede girdiler ve çıktılar arasında eşleme yapan fonksiyon regresyon veya sınıflandırma yöntemleriyle oluşturulur. Eğitim verileri hem girdileri hem de çıktıları içerir. Denetimsiz öğrenme yönteminde, bilinmeyen bir yapı tahmin edilir ve giriş verilerinin sınıfı önceden belirlenmemiştir. Takviyeli öğrenme ise girdilerin çevreleriyle etkileşime girerek ödül olarak adlandırılan geri bildirimleri maksimize etmeyi ve en uygun eylem stratejilerini bulmayı amaçlar. Bu nedenle, takviyeli öğrenme diğer öğrenme türlerinden farklı bir yaklaşım benimser (Şafak, 2017).

Gökalp, denetimli öğrenme, önceden etiketlenmiş veri örnekleriyle eğitilen ve makine öğrenme alanında en yaygın ve en popüler yöntem olduğunu belirtmiştir. Bu yaklaşım, eğitim veri setindeki çeşitli sınıfları direkt olarak tanımlamaya odaklanır ve yeterli miktarda veri mevcut olduğunda, eğitim veri setindeki sınıflandırma hataları, bilinmeyen verilere dair tahminlerde de etkili olabilir (Hastie, Friedman ve Tibshirani, 2008).

Denetimli öğrenme, temel olarak sınıflandırma ve regresyon olmak üzere iki ana alt kategoriden oluşur. Sınıflandırma, girdi verilerini belirli bir sınıfa atama işlemini ifade ederken, regresyon ise girdi verilerini bir çıktı değişkenini tahmin etme işlemini içerir.

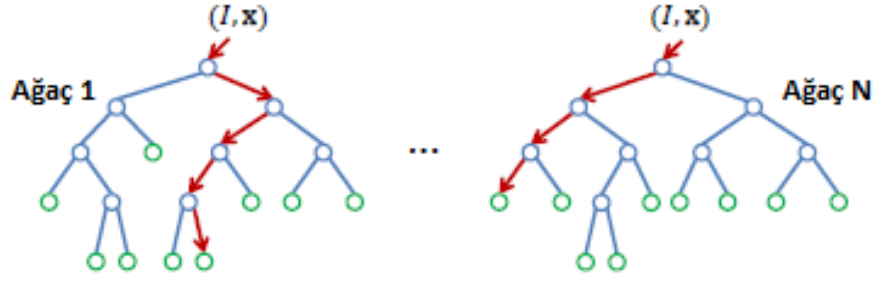
Çalışmamızda, veri setindeki tüm özniteliklerin etiketlere sahip olması, verilerin tamamen sayısallaştırılabilir olması ve hedef çıktının enerji ile ilgili olması nedeniyle, denetimli öğrenme yöntemlerine başvurulmuştur. Bu yöntemler içerisinde sınıflandırma ve regresyon algoritmaları bulunmaktadır. Bu çalışmada kullanılan özellikle şunlar vardır: Rastgele Orman (Random Forest), Karar Ağaçları (Decision Trees), Çoklu Doğrusal Regresyon (Polynomial Regression), En Yakın Komşuluk (K-Nearest Neighbors), Ridge Regresyon (Ridge Regression), Lasso Regresyon (Least Absolute Shrinkage and Selection Operator Regression) ve Doğrusal Regresyon (Linear Regression) algoritmalarıdır.

### **2.3.1. Rastgele Orman (Random Forest )**

Rastgele Orman, makine öğrenmesi alanında yaygın olarak kullanılan bir algoritmadır ve hem sınıflama hem de regresyon problemlerinde yüksek performans gösterir. Sınıflandırma görevlerinde, her sınıfın olasılığını hesaplamak veya veriyi belirli bir sınıfa atamak için kullanılabilir. Regresyon problemleri için ise, sayısal değerleri tahmin etmek amacıyla etkili bir şekilde kullanılabilir.

Topluluk (ensemble) modeli, birden fazla karar ağacının (decision tree) bir araya getirilmesiyle oluşturulur ve bu yaklaşım, daha tutarlı ve güçlü modellerin üretilmesine olanak tanır. Bu model, birden çok karar ağacının birleştirilmesiyle elde edildiği için, mevcut veri setinin daha doğru bir sınıflandırma yapmasına olanak sağlar.(Hegde, Wallace ve Gray, 2015). Her bir ağaç, veri setinin alt kümelerinde eğitilir ve tahminler sonrasında birleştirilir. Rastgele Orman algoritmasının performansını etkileyen önemli bir parametre, kullanılan ağaç sayısıdır. Genellikle daha fazla ağaç, daha iyi sonuçlar sağlayabilir, ancak bu, daha fazla hesaplama gücü gerektirebilir. Ayrıca, ağaçların derinliği de ayarlanabilir. Aşağıda Şekil 4'teki görsel, bölünmüş ve yaprak düğümlerinden oluşan T adet karar ağacının bir araya getirilerek oluşturduğu ormanı göstermektedir. Bu ormanda, kökten başlayarak her bir ağaç, t eşiği üzerinden sağa ve sola dallanarak belirli bir özelliği öğrenir ve bu dağılımın depolandığı bir yapı oluşturur. (Shotton, Sharp, Kipman, Fitzgibbon, Finocchio, Blake ve Moore, 2013).





**Şekil 4.** Rastgele orman algoritmasına ait ağaç yapısı

**Kaynak:** Shotton vd., 2013

Veri setinden rastgele örneklemeler (bootstrap örnekleme) alınır, bu da her ağacın farklı bir veri alt kümesi üzerinde eğitilmesini sağlar; ardından düğümde en iyi bölünme kriterini seçmek yerine, rasgele bir öznelik alt kümesi seçilerek rastgelelik sağlanır. Bölünme özelliği sayesinde sınıflandırmada bağımsız ve birden fazla ağaç üretilir (Akar ve Güngör, 2012). Bu, her ağacın birbirinden farklı olduğunu ve aşırı uyum (overfitting) eğilimini azalttığını garanti eder. Aykırı değerlerin varlığında çok yüksek boyutlu parametre uzaylarında diğer makine öğrenimi algoritmalarına göre daha kararlı kabul edilir (Caruana ve Niculescu-Mizil, 2006; Menze, Kelm, Masuch, Himmelreich, Bachert, Petrich ve Hamprecht, 2009).

Rastgele Orman algoritması, kullanıcı tarafından başlangıçta iki önemli parametrenin tanımlanmasıyla başlar. Bu parametrelerin yanı sıra, algoritmanın ağaç sayısını belirlemek için  $N$  ve her düğümde kullanılacak değişken sayısını ifade etmek için  $m$  belirlenir. Özellikle, eğitim verilerinin %66' sının her düğümde kullanılacak olan  $m$  değişkenlerini, mevcut tüm değişkenler arasından rastgele seçerek en iyi dalların belirlenmesine katkı sağlar (Akalin, 2020). Rastgele orman algoritması ağaç üretmek için CART (Classification and Regression Tree) algoritmasını kullanır (Beriman ve Cutler, 2005). Bu algoritma hangi özneliğin tahminlerde daha belirleyici olduğunu bulmak ve bölünme-seçme işlemi esnasında düğümlerin homojen dağılması istendiğinden CART algoritmasının kriteri olan GINI indeksi hesaplanır (Akalin, 2020; Hlasny, Ceriani ve Verme, 2022). Gini indeksi, homojenlik ilkesine dayalı olarak regresyon veya sınıflandırma özneliklerin tahmin gücünün bir ölçüsüdür (Strobl, Boulesteix, Zeileis ve Hothorn, 2007 ). GINI indeksi azaldıkça sınıfın homojenliği artmakta ve alt düğümün GINI indeksinin üst düğümün GINI indeksinden

daha az olması o dalın başarılı olarak kabul görmesini sağlamaktadır (Akar ve Güngör, 2012). GINI indeksi bir n düğümü hesaplaması Denklem (2)' de verilmiştir.

$$GINI(n) = 1 - \sum_{j=1}^2 (p_j)^2 \quad (2)$$

**Denklem 2.** GINI indeksi formülü

**Kaynak:** Okolo, 2022

Bu denklemde;  $p_j$ , n düğümündeki j sınıfının göreceli sıklığıdır.

### 2.3.2. Karar Ağaçları (Decision Trees)

Karar Ağaçları (Decision Trees), veri madenciliği ve makine öğrenme alanlarında yaygın olarak kullanılan bir sınıflandırma ve regresyon yöntemidir. Sınıflandırma ağaçları, veriyi belirli sınıflara atanması amacı güderken, regresyon ağaçları ise sayısal tahminler yaparlar. Karar ağaçları, mevcut veri setini çeşitli karar kurallarına tabi tutarak, veriyi daha küçük veri seti gruplarına bölme işlemiyle çalışan bir algoritmadır (Akalin, 2020). Karar ağaçları, kararları ve sonuçları hiyerarşik bir ağaç yapısı içinde görsel ve mantıksal olarak temsil eder. Bu ağaç yapısı, veri setini adım adım bölen ve sınıflandıran işlemleri gerçekleştirerek kararların sistematik bir şekilde alınmasını sağlar.

Her karar ağacı, bir ağaç yapısı içinde her düğümün bir özelliği (niteliği) ifade ettiği, her bağlantının (dalın) bir kararı temsil ettiği ve her bir yapıda sonuçların bulunduğu bir yapıdır. Bu ağaç, düğüm ve dallardan oluşur. Her düğüm, sınıflandırma işlemi gerçekleştirilecek olan bir grup içindeki özellikleri simgeler; buna karşılık her bir dal, bu düğümün alabileceği bir değeri yansıtır (Gültepe, 2019).

Karar ağaçları, çok yönlü bir makine öğrenme algoritmasıdır ve geniş bir uygulama yelpazesi sunar. Ancak, aşırı uyarlanma (overfitting) sorunlarına karşı hassas olabilirler. Bu nedenle, ağaçların derinliği ve diğer parametreler dikkatle ayarlanmalıdır.

### 2.3.3. Çoklu Doğrusal Regresyon (Polynomial Regression)

Çoklu doğrusal regresyon, veriler arasındaki ilişkiyi modellemek ve tahminlerde bulunmak amacıyla kullanılan bir regresyon analizi türüdür. Bu yöntem, bağımlı bir değişkenin (sonucun) bağımsız değişkenlerle (girdilerin) bir polinom fonksiyonu ile

ilişkilendirilmesi olarak düşünülebilir ve lineer regresyonun daha genişletilmiş bir versiyonunu temsil eder.

Çoklu doğrusal regresyon, veriler arasındaki bazen karmaşık ve eğrisel ilişkileri modellemek amacıyla kullanılır. Doğrusal regresyon sadece doğrusal ilişkileri ifade edebilirken, çoklu doğrusal regresyon daha karmaşık ilişkileri temsil edebilir (Arslankaya ve Toprak, 2021). Esas olarak, verileri daha yüksek dereceli bir polinom denklemine uyum sağlar. Çoklu doğrusal regresyon denklemi genellikle Denklem(3)' teki formatta ifade edilir :

$$y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad (3)$$

**Denklem 3.** Çoklu doğrusal regresyon formülü

**Kaynak:** Ostertagová, 2012

Burada:

y bağımlı değişkeni temsil eder.

x, bağımsız değişkeni temsil eder.

$a_0, a_1, a_2, \dots, a_n$  regresyon katsayılarıdır ve modelin en uygun polinomunu temsil eder.

n polinomun derecesini ifade eder.

Çoklu doğrusal regresyon, özellikle veriler arasındaki ilişkinin doğrusal olmadığı senaryolarda kullanışlı bir araçtır. Bununla birlikte, aşırı uyum (overfitting) sorunuyla karşılaşma riski bulunur; bu nedenle polinom derecesi (n), dikkatli bir şekilde seçilmelidir. Çünkü çok yüksek bir polinom derecesi, verilere mükemmel bir uyum sağlayabilir, ancak yeni verilere genelleme yapma yeteneğini azaltabilir. Bu nedenle, Çoklu doğrusal regresyon kullanırken modeli doğru bir şekilde ayarlamak büyük önem taşır.

#### **2.3.4. K-En Yakın Komşu (K-Nearest Neighbour)**

K En Yakın Komşu (k-NN) algoritması, sınıflandırma ve regresyon problemleri için kullanılan bir makine öğrenme yöntemidir. Temel prensip, veri noktalarının genellikle gözlemlenen komşularının sınıfını veya değerini tahmin etmeye dayanır (Dilki ve Başar, 2020).

K-NN, veri noktalarının birbirine yakınlığına dayalı tahminlerde bulunur ve benzer veri noktalarının genellikle aynı sınıfa veya benzer bir değere sahip olduğu düşüncesine dayanır. Modelin performansını etkileyen parametrelerden biri olan "K," tahminler yapılırken kaç komşunun göz önünde bulundurulacağını belirler. Küçük K değerleri, modele aşırı uyum eğilimi kazandırabilirken, büyük K değerleri daha düşük karmaşıklık sunarken, daha düşük ayrıntı seviyelerine neden olabilir (Sinap, 2023).

K-NN'de kullanılan uzaklık metriği, veri noktaları arasındaki benzerliği değerlendirir. Bu amaçla, en yaygın olarak kullanılan uzaklık metrikleri öklidyen uzaklık (Euclidean distance) ve Manhattan uzaklığıdır. Bununla birlikte, farklı veri analizlerinde ve problemlerde diğer uzaklık metrikleri de kullanılabilir. Öklidyen mesafe hesaplaması, verilen iki nokta arasındaki uzaklığı ölçmek için kullanılır ve Denklem (4)' teki formülle ifade edilir:

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \quad (4)$$

**Denklem 4.** Öklidyen uzaklık formülü

**Kaynak:** Cunningham ve Delany, 2021

K-NN algoritması kullanıldığında, veri normalizasyonu (örneğin, Min-Max ölçeklendirme veya z-skor normalizasyonu) büyük bir öneme sahiptir. Çünkü farklı özellikler genellikle farklı ölçeklerde olabilir ve bu durum, uzaklık hesaplamalarını etkileyebilir.

K-NN, basit ve anlaşılır bir algoritma olmasının yanı sıra bazı avantajlar sunar. Ancak büyük veri setleri ve yüksek boyutlu veri uzayları gibi durumlarda hesaplama maliyeti artabilir. Ayrıca, uygun bir K değeri seçimi ve veri normalizasyonu gibi önemli hiperparametrelerin ayarlanması gerekebilir.

### 2.3.5. Ridge Regresyon (Ridge Regression)

Ridge Regresyon (Ridge Regression), makine öğrenme ve istatistiksel regresyon analizinde kullanılan bir regresyon yöntemidir. Ridge regresyon, aynı zamanda L2 düzenlemesi olarak da bilinir ve aşırı uyum (overfitting) kontrolü için kullanılır.

Hoerl ve Kennard (1970), Ridge Regresyonunun geleneksel doğrusal regresyon ile karşılaşılan çoklu bağlantı sorununu çözmek için L2 düzenlemesini kullandığını gösterir. L2 düzenlemesi, regresyon modelinin tahmin varyanslarını azaltırken, tamamen sıfırlamaz. Bu, modelin daha istikrarlı hale gelmesini sağlar ve aşırı uyum eğilimini azaltır.

Ridge regresyonunda hedef, hata kareler toplamını en aza indiren katsayıları bulmaktır.(Aslan ve Yıldız, 2022). L2 düzenlemesi durumunda,  $\lambda$  ayar parametresini ifade eder. Bazı kaynaklarda bu, büzülme yöntemi olarak da adlandırılır.(Gokpinar, Ebegil ve Gokpinar, 2017). Ridge regresyonunda, düzenlemenin şiddeti alpha

parametresi ile kontrol edilir. Alpha'nın artması, düzenlemenin etkisini artırır ve modelin daha basit ve daha düzgün bir yapısına yol açabilir. Öte yandan, alpha'nın azalması, daha fazla özneliği etkin hale getirebilir ve modelin karmaşıklığını artırabilir.(Gupta, 2021). Gupta (2021),  $\lambda$  değerinin, modelin esnekliğinin ne kadar cezalandırılacağıyla ilgili olduğunu ifade etmiştir. Hata kareler toplamını minimize eden katsayılar, bu katsayılara uygulanan ceza ile belirlenir.(Hoerl ve Kennard, 1970). Ridge regresyon Denklem(5)' teki şekilde ifade edilmiştir:

$$SSE_{L_2} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^P \beta_j^2 \quad (5)$$

**Denklem 5.** Ridge regresyonu

**Kaynak:** Nacar ve Erdebilli, 2021

### **2.3.6. Lasso Regresyon (Least Absolute Shrinkage and Selection Operator Regression)**

Lasso Regresyonu (Least Absolute Shrinkage and Selection Operator Regression), makine öğrenme ve istatistiksel regresyon analizinde yaygın olarak kullanılan bir regresyon yöntemidir. Lasso regresyon aynı zamanda L1 düzenlemesi olarak da bilinir ve özellik seçimi (feature selection) ile aşırı uyum (overfitting) kontrolü için etkili bir şekilde kullanılır.

Tibshirani (1996), Lasso regresyonunun geleneksel doğrusal regresyondaki çoklu bağlantı sorununu ele almak için L1 düzenlemesini kullandığını belirtmiştir. L1 düzenlemesi, regresyon modelinin katsayılarını sıfıra yaklaştırır ve bazı katsayıları tamamen sıfır yapabilir, diğer katsayıları ise küçültür. Bu, aşırı uyum eğilimini azaltarak modelin daha istikrarlı hale gelmesini sağlar. Ayrıca, bu yaklaşım öznelik seçimi yapabilme yeteneği sunar, yani yalnızca önemli öznelikleri dahil edebilir. Bu nedenle, Lasso regresyonu özellikle öznelik seçimi için etkili bir yöntem olarak kabul edilir. Yüksek boyutlu veri setlerinde ve çok sayıda öznelikle çalışırken modelin karmaşıklığını azaltabilir. Özellikle gereksiz veya zayıf öznelikleri elemek için oldukça kullanışlı bir araçtır (Nacar ve Erdebilli, 2021).

Frank ve Friedman (1993), bazı senaryolarda Lasso regresyonunun daha iyi tahmin sonuçları sağladığına, diğer bazı durumlarda ise Ridge regresyonunun daha etkili olduğuna dikkat çekmişlerdir. Lasso regresyonunda amaç, hata kareler toplamını

minimize eden katsayıları tespit etmektir (Aslan ve Yıldız, 2022). Lasso regresyon Denklem (6)' daki şekilde ifade edilmiştir:

$$SSE_{L_1} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (6)$$

**Denklem 6.** Lasso Regresyon

**Kaynak:** Nacar ve Erdebilli, 2021

### 2.3.7. Doğrusal Regresyon (Linear Regression)

Doğrusal Regresyon (Linear Regression), veri analizi, istatistik ve makine öğrenme alanlarında temel bir regresyon analiz yöntemi olarak yaygın olarak kullanılır. Temel hedefi, bağımlı bir değişkenin (sonucun) bağımsız değişkenlerle (girdilerin) lineer bir ilişkisini modellemek ve bu model aracılığıyla gelecekteki değerleri tahmin etmektir.

Doğrusal regresyon, bağımlı değişken ile bağımsız değişkenler arasındaki bir lineer ilişkiyi varsayar. Bu, bir bağımlı değişkenin, bağımsız değişkenlerin ağırlıklarıyla çarpımının toplamı olarak ifade edilebileceği anlamına gelir. Her bağımsız değişkenin bir katsayısı (ağırlık) bulunmaktadır ve bu katsayılar, bağımsız değişkenlerin bağımlı değişkene olan etkisini gösterir. Doğrusal regresyon modelini oluştururken, en küçük kareler yöntemi kullanılır. Bu yöntem, gerçek gözlemler ile model tahminleri arasındaki kare hataların toplamını minimize ederek en uygun katsayıları hesaplar.(Nacar ve Erdebilli, 2021). Doğrusal regresyon denklemi Denklem (7)' de verilmiştir.

$$y = a_0 + a_1x + a_2x_2 + \dots + a_nx_n \quad (7)$$

**Denklem 7.** Doğrusal regresyon formülü

**Kaynak:** Kutner, Nachtsheim, Neter ve Li, 2005

Burada:

y bağımlı değişkeni temsil eder.

x, bağımsız değişkeni temsil eder.

$a_0, a_1, a_2, \dots, a_n$  regresyon katsayılarıdır ve modelin belirli bir doğrusal kombinasyonunu temsil eder.

Doğrusal regresyon, özellikle bir değişkenin diğer değişkenlere göre nasıl etkilendiğini anlamak veya bağımlı bir değişkenin tahminlerini yapmak amacıyla yaygın olarak kullanılan bir yöntemdir. Bununla birlikte, veri setinde bağımsız değişkenlerle doğrusal bir ilişki bulunmuyorsa veya bu ilişki çok karmaşık ise, diğer

regresyon yöntemleri veya makine öğrenme algoritmaları daha uygun bir seçenek olabilir.

## 2.4. Performans Değerlendirme Metrikleri

Bu çalışma da makine öğrenimi algoritmalarının değerlendirilmesi için R-Kare ( $R^2$ , R-Squared), Ortalama Mutlak Hata (MAE, Mean Absolute Error), Hata Kareleri Ortalaması (MSE, Mean Squared Error), Hata Kareleri Ortalamasının Karekökü (RMSE, Root Mean Squared Error) performans değerlendirme metrikleri kullanılmıştır.

### 2.4.1. R-Kare ( $R^2$ , R Squared)

$R^2$ , regresyon analizlerinde bağımsız değişkenin bağımlı değişken üzerindeki varyansın açıklanma derecesini ölçen bir metriktir. Bu ölçüt, regresyon modelinin ne kadar iyi uyarlandığını ve tahmin yeteneğini değerlendirmek için kullanılır. R-kare değeri genellikle 0 ile 1 arasında bir sayıdır veya yüzde olarak ifade edilir. 0 değeri, bağımsız değişkenlerin bağımlı değişkenin varyansını hiç açıklamadığını gösterirken, 1 değeri bağımsız değişkenlerin bağımlı değişkenin varyansını tamamen açıkladığını ifade eder. Başka bir deyişle, R-kare değeri ne kadar yüksekse, regresyon modelinin bağımlı değişken üzerindeki varyansı o kadar etkili bir şekilde açıkladığı söylenebilir. R-kare değeri, bir regresyon modelinin kalitesini yalnız başına belirlemez (Yetiz, Terzioğlu ve Kayakuş, 2021). R-kare formülü Denklem (8)'de verilmiştir.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{ort})^2} \quad (8)$$

**Denklem 8.** R Kare formülü

**Kaynak:** Yetiz vd.2021.

Burada:

$y_i$  :gerçek değer.

$\hat{y}_i$ :tahmin edilen değer.

$y_{ort}$ :ortalama değer.

### 2.4.2. Ortalama Mutlak Hata (MAE Mean Absolute Error)

Ortalama Mutlak Hata (MAE), regresyon analizlerinde kullanılan bir performans ölçüsüdür ve bir regresyon modelinin tahminlerinin gerçek değerlerden ne kadar sapma gösterdiğini ölçer. MAE, her tahminin gerçek değerden farkını mutlak değer olarak hesaplanır ve bu mutlak hataların ortalaması alınır. MAE'nin sıfıra

yakın olması veya sıfır olması, modelin iyi bir performans sergilediğini gösterir. Bu nedenle, MAE, regresyon ve zaman serisi problemlerinde sıklıkla tercih edilen bir performans ölçütüdür, çünkü sonuçları kolayca yorumlanabilir (Özen, Saraç ve Koyuncu, 2021).

MAE performans metriğinin denklemi şu şekildedir: MAE, gerçek değerler ile tahmin edilen değerler arasındaki farkların mutlak değerlerinin toplamını ifade eder. Bu metrik, iki sürekli değişken arasındaki farkın ölçüsüdür ve MAE değeri 0'dan  $\infty$ 'a kadar değişebilir.(Chai ve Draxler, 2014). MAE formülü Denklem (9)'de verilmiştir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

**Denklem 9.** Ortalama mutlak hata formülü

**Kaynak:** Chai ve Draxler, 2014.

Burada:

$y_i$  : Gerçek değer.

$\hat{y}_i$  : Modelin tarafından tahmin edilen değer.

n : Veri sayısı

### 2.4.3. Hata Kareleri Ortalaması (MSE Mean Squared Error )

Hata Kareleri Ortalaması (MSE), regresyon analizlerinde yaygın bir performans ölçütüdür ve bir regresyon modelinin tahminlerinin gerçek değerlere göre ne kadar hata yaptığını ölçer. MSE, her tahminin gerçek değerle arasındaki farkın karesinin alınarak hesaplandığı ve bu karelerin ortalamasının alındığı bir metriktir. MSE'nin değeri ne kadar sıfıra yakınsa, modelin tahminleri gerçek değerlere o kadar yakındır ve daha düşük MSE, daha iyi bir model performansını yansıtır.(Özşahin Yılmaz ve Akcan, 2022). Ancak, kareler kullanıldığı için orijinal hata ölçüsünün birimini kare alır, bu nedenle yorumlaması bazen daha zor olabilir. MSE' nin formülü Denklem(10)' da verilmiştir.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (10)$$

**Denklem 10.** Hata kareleri ortalaması formülü

**Kaynak:** Seyhan, 2021

Burada:

$y_i$  : Gerçek değer.

$\hat{y}_i$  : Modelin tarafından tahmin edilen değer.

n : Veri sayısı



#### 2.4.4. Hata Karaları Ortalaması Karekökü (RMSE Root Mean Squared Error )

Hata Kareleri Ortalaması Karekökü (RMSE), regresyon analizlerinde kullanılan bir performans ölçüsüdür ve bir regresyon modelinin tahminlerinin gerçek değerlere göre ne kadar hata yaptığını ölçer. RMSE, hata kareleri ortalamasının karekökü alınarak hesaplanır. Bu sayede, hataların orijinal birimde ifade edilen bir değeri temsil eder ve modelin tahmin yeteneğini daha anlaşılır bir şekilde değerlendirmeye yardımcı olur. RMSE'nin değeri sıfıra ne kadar yakınsa, modelin tahminleri gerçek değerlere o kadar yakındır ve daha düşük RMSE, daha iyi bir model performansını yansıtır.(Özşahin vd., 2022). RMSE, birçok matematiksel hesaplamada istenmeyen mutlak değer kullanılmamasını engeller (Chai ve Draxler, 2014). RMSE'nin formülü Denklem (11)' de verilmiştir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (11)$$

**Denklem 11.** Hata kareleri ortalaması karekökü formülü

**Kaynak:** Seyhan, 2021.

Burada:

$y_i$  : Gerçek değer.

$\hat{y}_i$  : Modelin tarafından tahmin edilen değer.

$n$  : Veri sayısı

## ÜÇÜNCÜ BÖLÜM

### ARAŞTIRMA SONUÇLARI VE TARTIŞMA

Çalışmanın bu aşamasına kadar, veriler toplandı, temizlendi, eksik veriler dolduruldu ve veriler dönüştürüldü. Bu aşamadan sonra özniteliklerin seçim işlemi yapılacak, veriler eğitim ve test veri seti olarak ayrılacak ardından makine öğrenmesi algoritmalarına tabi tutulacaktır.

#### 3.1. Özniteliklerin Seçimi

Makine öğrenmesi uygulamalarında, model kurulumuna başlamadan önceki en kritik adımlardan biri değişkenlerin seçimidir. Veri madenciliği veya makine öğrenimi projelerinde, doğru ve anlamlı özelliklerin tespit edilmesi sonuçların kalitesini artırabilir. Tüm özelliklerin kullanılması her zaman en iyi yaklaşım olmayabilir; bu, modelin performansını düşürebilir ve aşırı öğrenmeye yol açabilir. Model ne kadar basit ve anlaşılır ise, modeli gerçek dünyada uygulamak o kadar kolay olur. Bu nedenle doğru özelliklerin seçimi ve oluşturulması, aynı problemi ele alan farklı sonuçları büyük ölçüde etkileyebilir. Özniteliklerin seçimi için birçok yaygın kullanılan yöntem bulunmaktadır. Bu yöntemlerden biri olan Pearson Korelasyon Yöntemi, özelliklerin seçiminde sıklıkla kullanılan bir yöntemdir.

Pearson Korelasyon Yöntemi, sayısal iki öznitelik arasındaki ilişkiyi ifade eden bir yöntemdir. Korelasyon katsayısı, matematiksel olarak -1 ile 1 arasında değişir. Değer 1, pozitif bir korelasyonu temsil eder ve iki değişken arasındaki ilişkinin doğrusal ve en yüksek olduğunu gösterir. Değer 0, herhangi bir ilişki veya korelasyonun olmadığını belirtir. Değer -1, negatif bir korelasyonu gösterir ve ilişkinin ters orantılı olduğunu, yine en yüksek korelasyona sahip olduğunu ifade eder. Bu aralıktaki değerler, bu temel ilişkilerin ara değerlerini gösterir. Isı haritaları, verileri görsel olarak sunarak veri dağılımını, desenleri ve korelasyon ilişkilerini daha iyi anlamamıza yardımcı olur.

Bu aşamada, öncelikle Pearson korelasyon yöntemi kullanılarak değişkenler arasındaki ilişkileri incelemek amacıyla ısı haritaları ve çeşitli yöntemler kullanılmıştır. Veri setinde, 0.95 ve üzerinde veya -0.95 ve altında yüksek doğrusal ilişkilere sahip olan öznitelikler veri setinden çıkarılmıştır.

**Tablo 5.** Bağımsız değişkenler arasındaki ilişki-1

P arametre Adı	Parametre Adı	İlişki
5 cm Toprak Sıcaklığı	Islak Termometre Sıcaklığı	0.95
5 cm Toprak Sıcaklığı	Sıcaklık	0.95
5 cm Toprak Sıcaklığı	Minimum Sıcaklık	0.95
5 cm Toprak Sıcaklığı	Maksimum Sıcaklık	0.95
5 cm Toprak Sıcaklığı	Toprak Üstü Minimum Sıcaklık	0.97
Islak Termometre Sıcaklığı	Sıcaklık	0.97
Islak Termometre Sıcaklığı	Minimum Sıcaklık	0.97
Islak Termometre Sıcaklığı	Maksimum Sıcaklık	0.97
Islak Termometre Sıcaklığı	Toprak Üstü Minimum Sıcaklık	0.95
Sıcaklık	Minimum Sıcaklık	1
Sıcaklık	Maksimum Sıcaklık	1
Sıcaklık	Toprak Üstü Minimum Sıcaklık	0.98
Minimum Sıcaklık	Maksimum Sıcaklık	1
Minimum Sıcaklık	Toprak Üstü Minimum Sıcaklık	0.98

Tablo 5' te gösterildiği gibi, Sıcaklık, Minimum Sıcaklık ve Maksimum Sıcaklık öznitelikleri arasında +1 pozitif korelasyon olduğu, Sıcaklık ile Toprak Üstü Minimum Sıcaklık arasındaki ilişkinin 0.98 olduğu ve Islak Termometre Sıcaklığı ile Sıcaklık arasındaki ilişkinin 0.97 olduğu gözlemlenmiştir. Bu nedenle, sıcaklık ile ilgili olarak sadece 5 cm Toprak Sıcaklığı ve Islak Termometre Sıcaklıkları özniteliklerini kullanmış; diğer öznitelikleri veri setinden çıkartılmıştır.

**Tablo 6.** Bağımsız değişkenler arasındaki ilişki-2

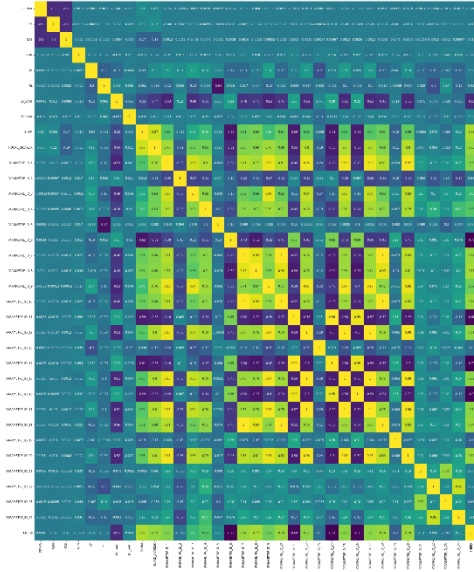
P arametre Adı	Parametre Adı	İlişki
Nispi Nem	Maksimum Nispi Nem	0.95
Nispi Nem	Minimum Nispi Nem	0.95
Minimum Nispi Nem	Maksimum Nispi Nem	0.99

Tablo 6' da görüldüğü gibi, Maksimum Nispi Nem ile Minimum Nispi Nem arasındaki ilişki katsayısı 0.99 ve bu iki öznitelik ile Nispi Nem arasındaki ilişki katsayısı 0.95 olarak hesaplanmıştır. Bu yüksek ilişki değerleri nedeniyle, Maksimum Nispi Nem ve Minimum Nispi Nem öznitelikleri veri setinden çıkarılmıştır.

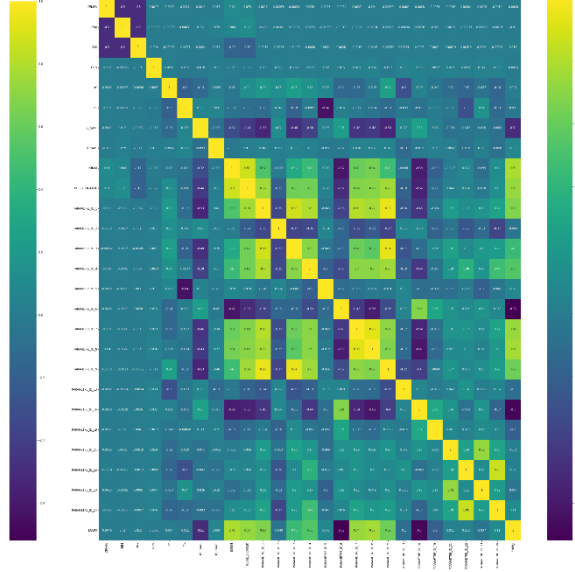
**Tablo 7.** Bağımsız değişkenler arasındaki ilişki-3

P arametre Adı	Parametre Adı	İlişki
Güneşlenme Şiddeti	Küresel Güneş Radyasyonu	1
Güneşlenme Şiddeti	Toplam Küresel Güneş Radyasyonu	1
Küresel Güneş Radyasyonu	Toplam Küresel Güneş Radyasyonu	1

Tablo 7' de görüldüğü üzere Küresel Güneş Radyasyonu ve Toplam Küresel Güneş Radyasyonu öznitelikleri arasındaki ilişki katsayısı 1 çıktığı için bu iki öznitelik veri setinden çıkarılmıştır.



Şekil 5. Korelasyon ısı haritası önce



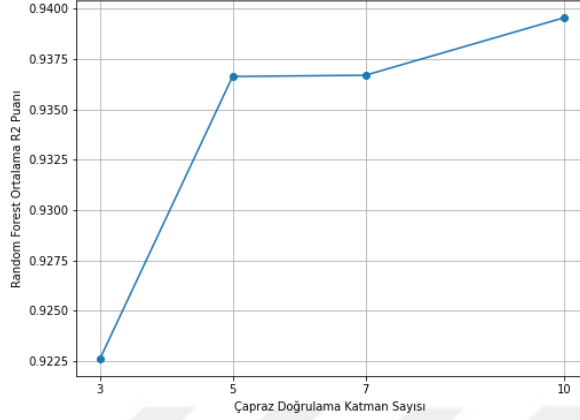
Şekil 6. Korelasyon ısı haritası sonra

Yüksek korelasyona sahip değişkenlerin veri setinden çıkarılmasının ardından, ısı haritası önce sarı renkte yoğun bir dağılıma sahipken (Şekil 5), korelasyon sonrası (Şekil 6) mavi renkteki dağılımın arttığını göstermektedir. Bu da aşırı uyarlamaya sebebiyet verebilecek özniteliklerin verilerden başarılı şekilde çıkarıldığını göstermektedir.

### 3.2. Veri Setinin Eğitim ve Test Olarak Bölünmesi

Bu aşamada, araştırma çalışmamız veri setini iki ayrı parça olarak, eğitim ve test verileri olarak böldüğü önemli bir aşamada bulunmaktadır. Temel amacımız, bir veri setini eğitim ve test verilerine bölerken, makine öğrenimi modellemesi için gerekli olan veri hazırlığını başarıyla tamamlamaktır. Bu aşama, modelin doğruluğunu ve performansını değerlendirmek için hayati bir öneme sahiptir ve makine öğrenimi projelerinin temelini oluşturur. Bu işlem, genellikle verilerin %70' ini eğitim ve %30' unu test için ayırma yaklaşımına dayanır. Ancak bu yaklaşımın dezavantajı, verinin rastgele ayrılmasının modelin eğitim ve test aşamalarında bazı sapmalara ve hatalara yol açabilmesidir. İşte çapraz doğrulama (cross-validation) bu sorunu çözmek için kullanılır. Veriyi belirtilen bir k sayısına göre eşit parçalara böler ve her bir parçayı hem eğitim hem de test için kullanır. Bu yöntem, modelin performansının veriye bağımlı olup olmadığını kontrol etmek ve daha güvenilir sonuçlar elde etmek için yaygın olarak kullanılır. Çapraz doğrulama, aşırı uydurma (overfitting) sorunlarını tespit etmek ve modelin genelleme yeteneğini değerlendirmek için önemli bir araçtır.

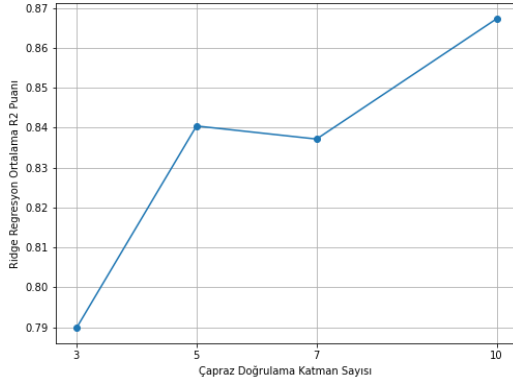
Çapraz doğrulama (cross-validation), modelin veriye duyarlı olup olmadığını ve aşırı uyarlanma (overfitting) sorunuyla karşı karşıya olup olmadığını anlamak için kullanılan bir tekniktir (Sorhun,2021).



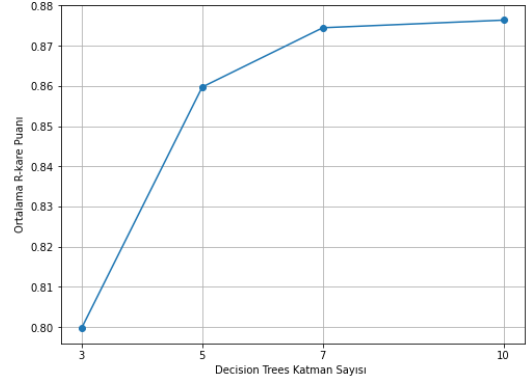
**Grafik 2.** Rastgele orman algoritması çapraz doğrulama katman sayısı karşılaştırması

Çalışmanın bu aşamasında, veri serisi eğitim ve test olarak bölünürken, çapraz doğrulama yöntemi kullanılarak en etkili katmanın belirlenmesi amaçlanmıştır. Çapraz doğrulama işlemi sırasında katmanlar 3, 5, 7 ve 10 olarak seçilmiştir. Bu katmanların her biri için elde edilen  $R^2$  puanları Grafik 2’de incelenmiş ve en yüksek  $R^2$  puanının 10 katmanlı çapraz doğrulama yöntemine ait olduğu görülmüştür. Bu nedenle, Random Forest yöntemi ile en etkili sonuçları elde etmek için 10 katmanlı çapraz doğrulama yöntemi seçilmiştir.

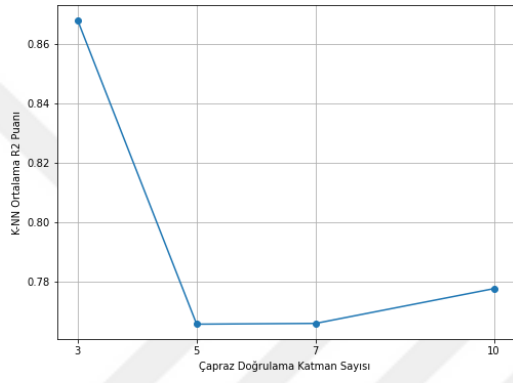
Yukarıda açıklanan işlemler, farklı makine öğrenimi algoritmalarına uygulanmış ve yedi farklı makine öğrenimi algoritması için en etkili çapraz doğrulama katmanları belirlenmiştir. Sonuçlara göre, Rastgele Orman için 10 katman, Karar Ağaçları için 10 katman, Çoklu Doğrusal Regresyon için 5 katman, K-En Yakın Komşular için 3 katman, Ridge Regresyon için 10 katman, Lasso Regresyon için 10 katman ve Doğrusal Regresyon için 5 katmanlı modellerin seçilmesi, en etkin sonuçların elde edilmesini sağlayacaktır. Bu nedenle, bu çalışmada her bir makine öğrenimi algoritması için aşağıdaki Grafik 3’te belirtilen katman sayısının tercih edilmesi önerilmektedir.



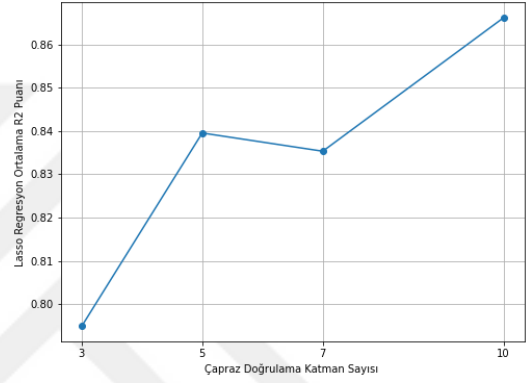
(a)



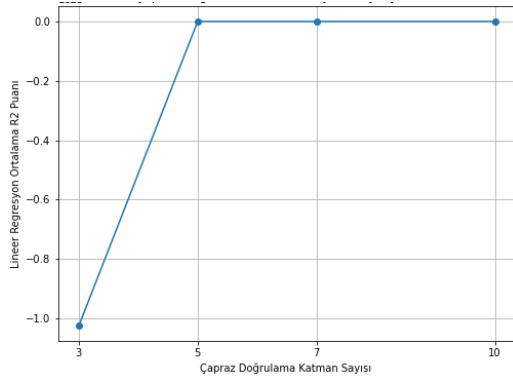
(b)



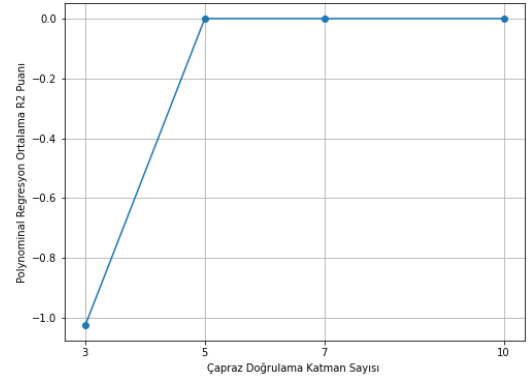
(c)



(d)



(e)



(f)

**Grafik 3.** Çapraz Doğrulama katman sayısı karşılaştırmaları

(a) Ridge Regresyon (b) Karar Ağaçları (c) K-En Yakın Komşu  
 (d) Lasso Regresyon (e) Doğrusal Regresyon (f) Çoklu Doğrusal Regresyon

### 3.3. Makine Öğrenmesi Algoritmalarının Performansları

Çalışmanın şimdiye kadar olan bölümü aşağıdaki şekilde özetlenebilir: Bilecik ilinin Bozüyük ilçesinde yer alan üç farklı güneş enerji santralinin verileri, Bilecik merkezindeki Meteoroloji Müdürlüğü istasyonundan kaynaklanan meteorolojik verilerle birleştirildi. Veri toplama süreci sırasında ve enerji üretimi başlangıç ve bitiş saatleri arasındaki zaman uyumsuzlukları göz önüne alındığında, saatlik veriler günlük periyotlara dönüştürüldü. Meteoroloji verileri, tüm günü kapsayacak şekilde kaydedilmişken, enerji üretim başlangıç ve bitiş saatleri her gün güneşin doğuşu ve batışıyla değişiyordu. Bu nedenle, meteoroloji verileri her gün için enerji üretimi başlangıç ve bitiş saatleri dikkate alınarak tüm veriler tek bir tabloda birleştirildi.

Veri seti, Jupyter Notebook programı kullanılarak içe aktarıldı ve kapsamlı bir analiz sürecinden geçirildi. Bu analizler sonucunda eksik veriler tespit edildi ve bu eksik verilerin doldurulması işlemi gerçekleştirildi. Eksik verilerin, ayın mevsimsel etkisi ve tüm yılın mevsimsel koşullarıyla uyumsuz olabileceği göz önüne alındı. Bu nedenle eksik veriler, ilgili ayın mod, medyan veya ortalama değerleri kullanılarak dikkatle dolduruldu.

Veri setinde yer alan kategorik veriler, sayısal verilere dönüştürmek için One-Hot Encoder kullanıldı ve bu veriler arasındaki sayısal farklılıkları gidermek için normalizasyon işlemi uygulandı. Sonuç olarak, 35 öznitelikten ve 3782 satırdan oluşan bir veri seti elde edildi.

Ayrıca, veri seti üzerinde yapılan analizler sonucunda Pearson Korelasyonu kullanılarak öznitelikler arasındaki ilişkiler değerlendirildi. 0.95 ve üzeri ilişkiye sahip olan 8 öznitelik, veri setinden çıkarılarak, veri setinin makine öğrenimi algoritmaları için daha uygun hale getirilmesi sağlandı. Bu sayede veri seti, modelleme süreci için hazır hale getirildi.

Modelleme işlemine başlamadan önce, veri seti eğitim ve test verileri olarak ayrıldı. Veriyi rastgele bölmek, modelin performansını etkileyebileceğinden, her makine öğrenimi algoritması için çapraz doğrulama yöntemi kullanıldı. Bu yaklaşım, modelin performansını daha güvenilir bir şekilde değerlendirmeye olanak tanır.

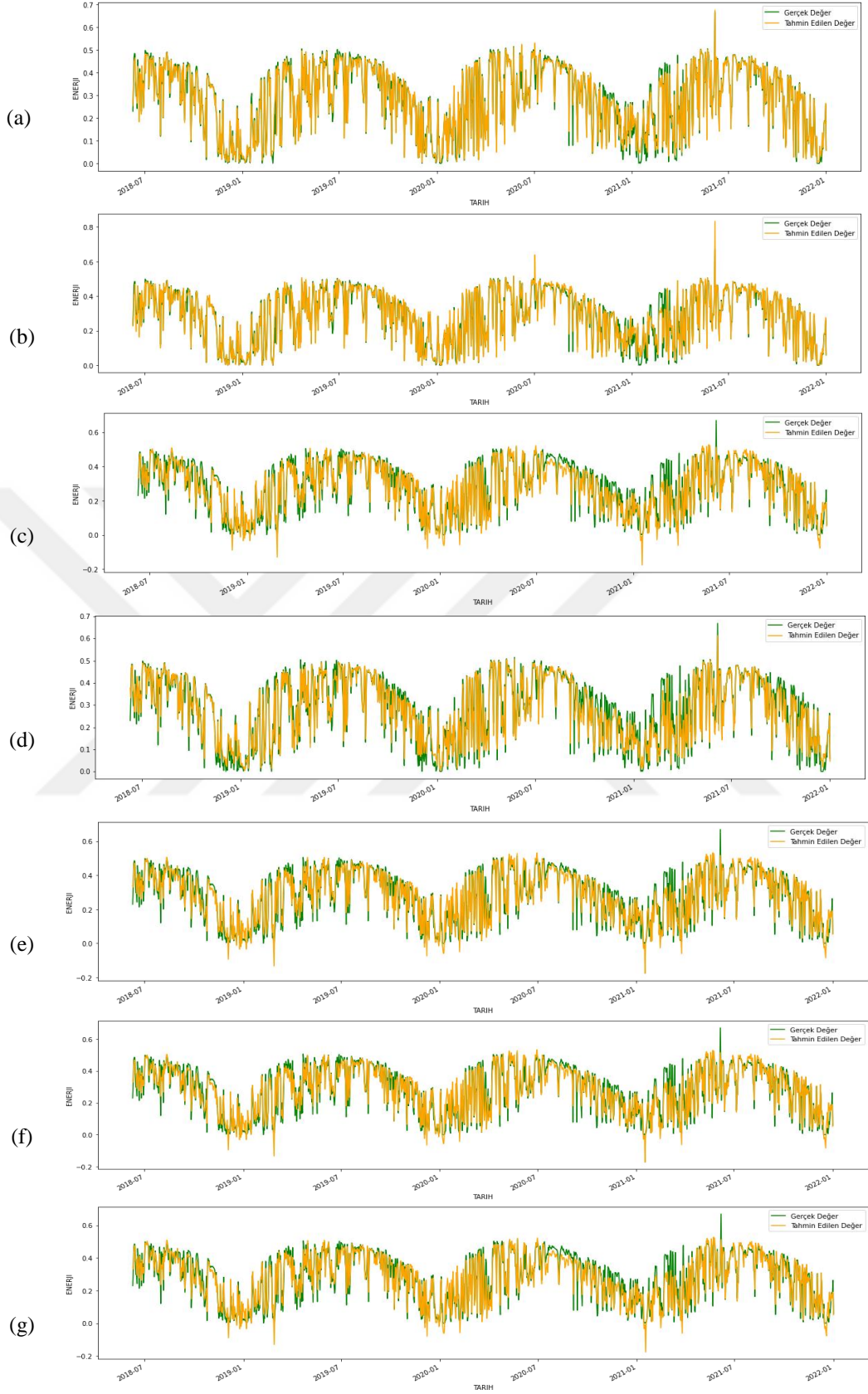
**Tablo 8.** Rastgele orman 10 katmanlı çapraz doğrulama puanlamaları

Eğitim Süresi (dakika)	Test Süresi (dakika)	MAE	MSE	R <sup>2</sup>
4.246773	0.031225	0.012567	0.000764	0.968420
4.639811	0.015613	0.015625	0.001280	0.945804
4.470950	0.031227	0.009998	0.000506	0.976169
4.531174	0.022950	0.014711	0.000968	0.959007
4.873454	0.046849	0.015791	0.001030	0.954779
5.246007	0.015594	0.012454	0.001152	0.944716
4.277527	0.031245	0.013178	0.001219	0.931728
4.559003	0.031228	0.017749	0.001334	0.942663
5.011798	0.015631	0.028700	0.002250	0.890634
4.450146	0.015615	0.030008	0.002639	0.881027

Rastgele Orman algoritması için uygulanan çapraz doğrulama katman karşılaştırma işlemi sonucunda, 10 katmanlı çapraz doğrulama ile en etkili puanlamalara ulaşıldığı gözlemlenmiştir. Veri seti üzerinde 10 katmanlı çapraz doğrulama ile Rastgele Orman algoritması uygulandığında, Tablo 8' de sunulan hata performans metrik değerleri elde edilmiştir. Bu sonuçlara göre en başarılı katmanın, R<sup>2</sup> puanı 0.968420 ile en yüksek değere ulaşmış, MSE değeri 0.000764, MAE değeri 0.012567 olarak hesaplanmıştır. Modelin eğitilmesi 4.246773 dakika sürmüştür, test edilmesi ise 0.31250 dakika almıştır.

R<sup>2</sup> puanının 1'e çok yakın olması, hata metriklerinin sıfıra yakın olması, bu modelin etkili bir şekilde çalıştığını ve tahmin edilen değerlerin gerçek değerlere çok yakın olduğunu göstermektedir. Bu nedenle, Rastgele Orman algoritması, veri setimiz üzerinde en iyi performansı gösteren algoritma olarak belirlenmiştir. Buna rağmen, modelleme süresi oldukça yüksek bir değere ulaşmıştır. Aşağıdaki Grafik 4.a' da veri setine Rastgele Orman algoritması uygulandıktan sonra gerçek değerler (actual target (yeşil)) ile modelin tahmin ettiği değerler (predicted target (sarı)) grafiği görülmektedir. Yeşil alanların azlığı modelin başarısını göstermektedir.





**Grafik 4.** Makine öğrenmesi algoritmaları gerçek-tahmin edilen değer grafikleri

(a)Rastgele Orman (b)Karar Ağaçları (c)Çoklu Doğrusal Regresyon  
(d)K- En Yakın Komşu (e) Ridge Regresyon (f)Lasso Regresyon (g) Doğrusal Regresyon

**Tablo 9.** Karar ağaçları 10 katmanlı çapraz doğrulama puanlamaları

Eğitim Süresi (dakika)	Test Süresi (dakika)	MAE	MSE	R <sup>2</sup>
0.078122	0.015624	0.016059	0.001075	0.955562
0.093740	0.000000	0.018128	0.001720	0.927177
0.125009	0.000000	0.014562	0.001395	0.934265
0.078101	0.015625	0.020338	0.002288	0.903118
0.062495	0.000000	0.017843	0.001378	0.939535
0.109368	0.000000	0.015832	0.002225	0.893177
0.093744	0.000000	0.016489	0.001495	0.916246
0.078121	0.000000	0.019785	0.001645	0.929332
0.062495	0.015626	0.037431	0.003912	0.809847
0.062494	0.000000	0.043643	0.006421	0.710576

Karar Ağaçları algoritması için 10 katmanlı çapraz doğrulama ile en etkili puanlamalara ulaşıldığı gözlemlenmiştir. Veri seti üzerinde 10 katmanlı çapraz doğrulama ile Karar Ağaçları algoritması uygulandığında, Tablo 9' da sunulan hata performans metrik değerleri elde edilmiştir. Bu sonuçlara göre en başarılı katmanın, R<sup>2</sup> puanı 0.955562 ile en yüksek ikinci değere ulaşmış, MSE değeri 0.001075, MAE değeri 0.016059 olarak hesaplanmıştır. Modelin eğitilmesi 0.078122 dakika sürmüştür, test edilmesi ise 0.015624 dakika almıştır. R<sup>2</sup> puanının 1'e çok yakın olması, hata metriklerinin sıfıra yakın ve negatif değerlerde olması, bu modelin etkili bir şekilde çalıştığını ve tahmin edilen değerlerin gerçek değerlere çok yakın olduğunu göstermektedir. Bu nedenle, Karar Ağaçları algoritması, veri setimiz üzerinde ikinci en iyi performansı gösteren algoritma olarak belirlenmiştir.

Modelleme süresi, Rastgele Orman Algoritması'na göre oldukça kısa olmasına rağmen R<sup>2</sup> puanı daha düşüktür. Yukarıda Grafik 4.b' de veri setine Karar ağaçları algoritması uygulandıktan sonra gerçek değerler (yeşil) ile modelin tahmin ettiği değerler (sarı) görülmektedir. Yeşil alanların azlığı modelin başarısını göstermektedir.

**Tablo 10.** Çoklu Doğrusal Regresyon 5 katmanlı çapraz doğrulama puanlamaları

Eğitim Süresi (dakika)	Test Süresi (dakika)	MAE	MSE	R <sup>2</sup>
0.140615	0.0	0.023647	0.001667	0.930401
0.140615	0.0	0.022007	0.001499	0.934440
0.156238	0.0	0.026565	0.001683	0.922892
0.124999	0.0	0.041460	0.003886	0.816430
0.131351	0.0	0.057116	0.009124	0.575508

Çoklu doğrusal regresyon algoritması için 5 katmanlı çapraz doğrulama ile en etkili puanlamalara ulaşıldığı gözlemlenmiştir. Veri seti üzerinde 5 katmanlı çapraz doğrulama ile çoklu doğrusal regresyon algoritması uygulandığında, yukarıdaki Tablo 10' da sunulan hata performans metrik değerleri elde edilmiştir. Bu sonuçlara göre en başarılı katmanın,  $R^2$  puanı 0.930401 ile en yüksek üçüncü değere ulaşmış, MSE değeri 0.001667, MAE değeri 0.023647 olarak hesaplanmıştır. Modelin eğitilmesi 0.140615 dakika sürerken, test edilmesi ise 0 dakika altında kalmıştır.  $R^2$  puanının 1'e çok yakın olması, hata metriklerinin sıfıra yakın ve negatif değerlerde olması, bu modelinde etkili bir şekilde çalıştığını ve tahmin edilen değerlerin gerçek değerlere çok yakın olduğunu göstermektedir. Ancak veri setine çoklu doğrusal regresyon uygulandıktan sonra diğer modellere kıyasla tahmin edilen değerlerin ile gerçek değerler arasındaki farklılık artmıştır, bu da model performansının azaldığını göstermektedir.

**Tablo 11.** K-En Yakın Komşu 3 katmanlı çapraz doğrulama puanlamaları

Eğitim Süresi (dakika)	Test Süresi (dakika)	MAE	MSE	$R^2$
0.000000	0.124990	0.021669	0.001837	0.920089
0.000000	0.124991	0.021619	0.001857	0.918472
0.015624	0.125015	0.065131	0.015403	0.280074

K- en yakın komşu algoritması için 3 katmanlı çapraz doğrulama ile en etkili puanlamalara ulaşıldığı gözlemlenmiştir. Veri seti üzerinde 3 katmanlı çapraz doğrulama ile çoklu K- en yakın komşu algoritması uygulandığında, Tablo 11' de sunulan hata performans metrik değerleri elde edilmiştir. Bu sonuçlara göre en başarılı katmanın,  $R^2$  puanı 0.920089 ile en yüksek üçüncü değere ulaşmış, MSE değeri 0.001837, MAE değeri 0.021669 olarak hesaplanmıştır. Modelin eğitilmesi 0 dakika bile sürmezken, test edilmesi ise 0.124990 dakika almıştır.  $R^2$  puanının 1'e çok yakın olması, hata metriklerinin sıfıra yakın ve negatif değerlerde olması, bu modelinde etkili bir şekilde çalıştığını ve tahmin edilen değerlerin gerçek değerlere çok yakın olduğunu göstermektedir. Ancak yukarıdaki Grafik 4.d' de veri setine K-en yakın komşu algortiması uygulandıktan sonra diğer modellere kıyasla tahmin edilen değerlerin ile gerçek değerler arasındaki farklılık artmıştır, bu da model performansının azaldığını göstermektedir.

**Tablo 12.** Ridge regresyon 10 katmanlı çapraz doğrulama puanlamaları

Eğitim Süresi (dakika)	Test Süresi (dakika)	MAE	MSE	R <sup>2</sup>
<b>0.156218</b>	0.000000	0.034323	0.002038	0.915765
<b>0.124990</b>	0.000000	0.032414	0.002561	0.891551
<b>0.124992</b>	0.015622	0.027092	0.001498	0.929428
<b>0.140620</b>	0.000000	0.035063	0.002390	0.898773
<b>0.109368</b>	0.000000	0.036170	0.002433	0.893215
<b>0.124988</b>	0.000000	0.033049	0.002144	0.897057
<b>0.109371</b>	0.000000	0.030421	0.002481	0.861025
<b>0.109367</b>	0.000000	0.036942	0.002902	0.875300
<b>0.109364</b>	0.000000	0.049912	0.004374	0.787406
<b>0.109367</b>	0.000000	0.049322	0.006135	0.723466

Ridge regresyon algoritması için 10 katmanlı çapraz doğrulama ile en etkili puanlamalara ulaşıldığı gözlemlenmiştir. Veri seti üzerinde 10 katmanlı çapraz doğrulama ile Ridge regresyon algoritması uygulandığında, Tablo 12' de sunulan hata performans metrik değerleri elde edilmiştir. Bu sonuçlara göre en başarılı katmanın, R<sup>2</sup> puanı 0.915765 ile en yüksek ikinci değere ulaşmış, MSE değeri 0.002038, MAE değeri 0.034323 olarak hesaplanmıştır. Modelin eğitilmesi 0.156218 dakika sürerken, test edilmesi ise 0 dakika bile sürmemiştir.

**Tablo 13.** Lasso regresyon 10 katmanlı çapraz doğrulama puanlamaları

Eğitim Süresi (dakika)	Test Süresi (dakika)	MAE	MSE	R <sup>2</sup>
<b>0.265605</b>	0.000000	0.034482	0.002055	0.915055
<b>0.328116</b>	0.000000	0.032605	0.002598	0.889981
<b>0.249968</b>	0.000000	0.026507	0.001471	0.930667
<b>0.249984</b>	0.000000	0.034960	0.002395	0.898565
<b>0.234357</b>	0.000000	0.036647	0.002502	0.890186
<b>0.249984</b>	0.000000	0.033124	0.002145	0.897031
<b>0.234358</b>	0.000000	0.030473	0.002506	0.859606
<b>0.234359</b>	0.000000	0.036719	0.002898	0.875467
<b>0.218736</b>	0.015623	0.050945	0.004554	0.778652
<b>0.249985</b>	0.000000	0.048838	0.006067	0.726527

Lasso regresyon algoritması için 10 katmanlı çapraz doğrulama ile en etkili puanlamalara ulaşıldığı gözlemlenmiştir. Veri seti üzerinde 10 katmanlı çapraz doğrulama ile Lasso regresyon algoritması uygulandığında, Tablo 13' te sunulan hata performans metrik değerleri elde edilmiştir. Bu sonuçlara göre en başarılı katmanın, R<sup>2</sup> puanı 0.915055, MSE değeri 0.002055, MAE değeri 0.034482 olarak hesaplanmıştır. Modelin eğitilmesi 0.265605 dakika sürmüş, test edilmesi ise 0 dakika bile sürmemiştir.

**Tablo 14.** Doğrusal regresyon 5 katmanlı çapraz doğrulama puanlamaları

Eğitim Süresi (dakika)	Test Süresi (dakika)	MAE	MSE	R <sup>2</sup>
<b>0.000000</b>	0.015625	0.034770	0.002372	0.900945
<b>0.000000</b>	0.015623	0.031955	0.002023	0.911558
<b>0.000000</b>	0.000000	0.034439	0.002290	0.895083
<b>0.015624</b>	0.000000	0.035032	0.002833	0.866153
<b>0.000000</b>	0.000000	0.058813	0.007986	0.628476

Doğrusal regresyon algoritması için uygulanan çapraz doğrulama katman karşılaştırma işlemi sonucunda, 5 katmanlı çapraz doğrulama ile en etkili puanlamalara ulaşıldığı gözlemlenmiştir. Veri seti üzerinde 5 katmanlı çapraz doğrulama ile Lasso regresyon algoritması uygulandığında, Tablo 14' de sunulan hata performans metrik değerleri elde edilmiştir. Bu sonuçlara göre en başarılı katmanın, R<sup>2</sup> puanı 0.900945 ile en yüksek ikinci değere ulaşmış, MSE değeri 0.002372, MAE değeri 0.034770 olarak hesaplanmıştır. Modelin eğitilmesi 0 dakika bile sürmezken, test edilmesi ise 0.015625 dakika almıştır.

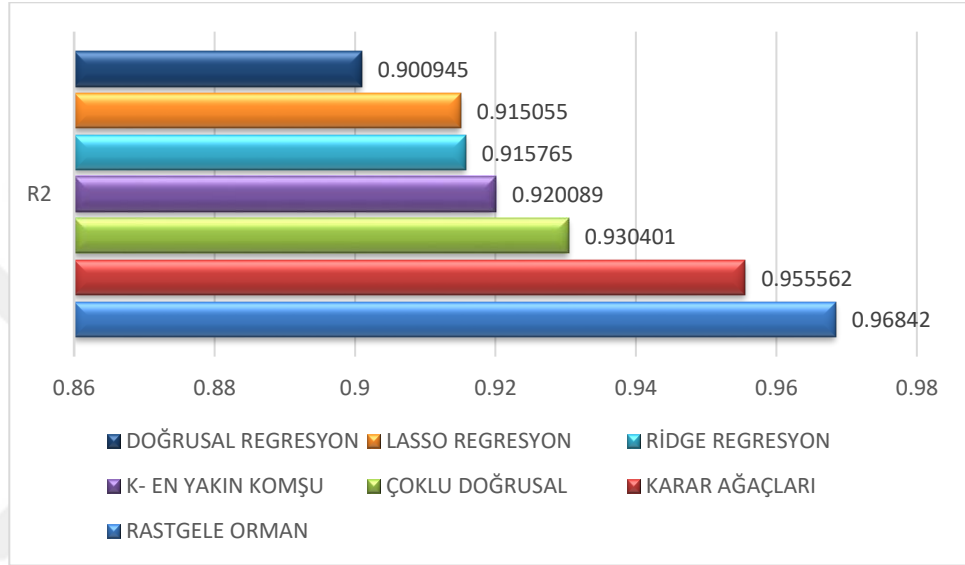
Ridge Regresyon, Lasso Regresyon ve Doğrusal Regresyon sonuçlarına göre R<sup>2</sup> puanları 1'e yakın bir değerde olmasına rağmen, bu algoritmaların diğer uygulamalara göre daha düşük performans gösterdiği gözlemlenmiştir.

Ayrıca, Sayfa 66' da bulunan Grafik 4.e, Grafik 4.f, Grafik 4.g grafiklerinde gerçek değerler ile tahmin edilen değerler arasındaki farklılık, yeşil alanların artışı, şeklinde görülmüştür. Bu durum, bu modellerin performansının diğerlerine göre düşük olduğunu işaret etmektedir.

**Tablo 15.** Makine öğrenmesi algoritmalarının performan karşılaştırılması

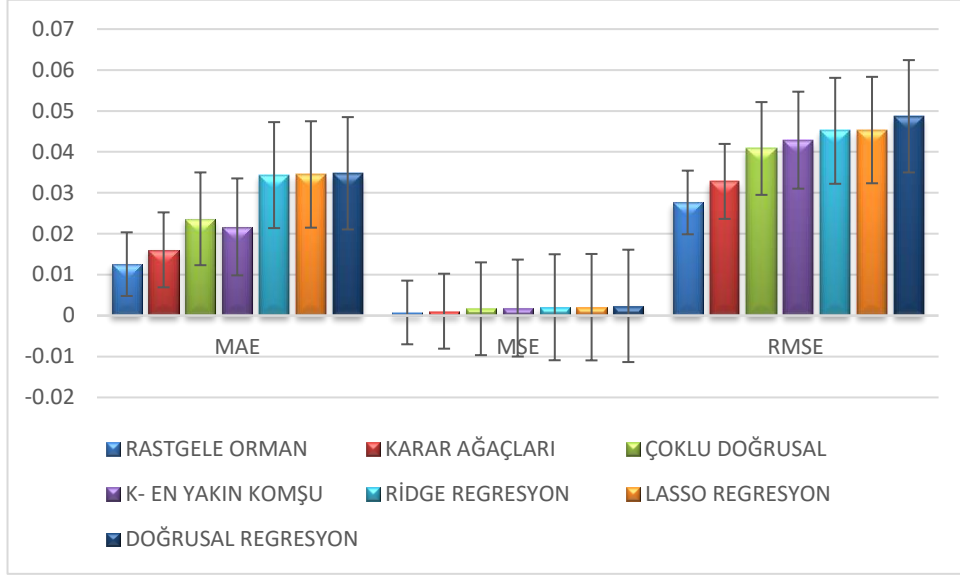
	R <sup>2</sup>	MAE	MSE	RMSE	EĞİTİM SÜRESİ (dakika)	TEST SÜRESİ (dakika)
<b>RASTGELE ORMAN</b>	0.96842	0.012567	0.000764	0.02764	4.24677	0.03122
<b>KARAR AĞAÇLARI</b>	0.955562	0.016059	0.001075	0.03278	0.07812	0.01562
<b>ÇOKLU DOĞRUSAL</b>	0.930401	0.023647	0.001667	0.04082	0.14061	0
<b>K- EN YAKIN KOMŞU</b>	0.920089	0.021669	0.001837	0.04286	0	0.12499
<b>RIDGE REGRESYON</b>	0.915765	0.034323	0.002038	0.04514	0.15621	0
<b>LASSO REGRESYON</b>	0.915055	0.034482	0.002055	0.04533	0.26560	0
<b>DOĞRUSAL REGRESYON</b>	0.900945	0.03477	0.002372	0.04870	0	0.01562

Veri seti üzerinde uygulanan makine öğrenimi algoritmalarının  $R^2$ , MSE, MAE, RMSE, Eğitim Süreleri ve Test Süreleri karşılaştırılması yukarıda Tablo 15' te yapılmıştır.  $R^2$  puanları, bir algoritmanın başarısını değerlendirmek için önemli bir performans metriği olarak kabul edilir. Bu metriğe göre en iyi modelleme sonucunu 0.96842  $R^2$  puanı ile Rastgele Orman algoritması elde etmiştir. En düşük  $R^2$  puanı ise 0.900945 puan ile Doğrusal Regresyon algoritması tarafından kaydedilmiştir.  $R^2$  skorlarının grafiği aşağıda Grafik 5' te sunulmuştur.



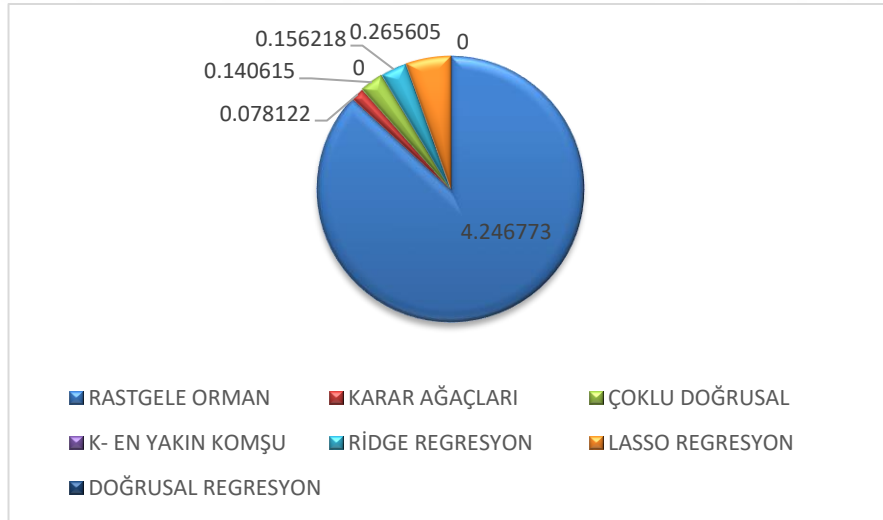
**Grafik 5.** Makine öğrenmesi modelleri  $r^2$  puan grafiği

Sadece  $R^2$  puanı, bir modelin performansını belirlemede yetersiz kalacaktır. Bu nedenle  $R^2$  puanları ile birlikte hata performans metrikleri olan MAE, MSE ve RMSE'ye de bakılması gerekmektedir. Bir modelin performansını değerlendirmek için  $R^2$  puanı 1'e ne kadar yakınsa ve hata performans metrikleri (MAE, MSE ve RMSE) ne kadar düşükse, o modelin performansı o kadar iyi kabul edilir.



**Grafik 6.** Makine öğrenmesi modelleri MAE, MSE, RMSE grafiği

Grafik 6’ da görüleceği üzere R2 puanı en yüksek olan Rastgele Orman algoritması, MSE, RMSE ve MAE değerleri açısından oldukça düşük sonuçlar elde etmiştir. Bu da Rastgele Orman algoritmasının performansının diğer kullanılan algoritmalar içinde en iyisi olduğunu göstermektedir. Bu başarı sıralamasını Karar Ağaçları ve Çoklu Doğrusal Regresyon izlemektedir.



**Grafik 7.** Makine öğrenmesi modelleri eğitim süreleri grafiği

Rastgele Orman algoritmasının tüm modellere göre en iyi performansı sergilemesine rağmen, eğitim ve test sürelerinin Grafik 7’ de görüldüğü üzere diğer modellere kıyasla belirgin şekilde daha uzun olması, bu modelin dezavantajı olarak öne çıkmaktadır.

## SONUÇLAR VE ÖNERİLER

Güneş enerji santralleri, enerji sektörünün geleceğinde hayati bir rol oynayan temiz ve sürdürülebilir enerji kaynaklarıdır. Bu kaynaklar, bol miktarda enerji üretme potansiyeline sahip olmaları nedeniyle gelecekteki enerji taleplerini karşılama konusunda kritik bir öneme sahiptir. Bu bağlamda, güneş enerjisi üretiminde verimliliği artırmak için yapılan çalışmalar son derece kritik bir rol oynamaktadır. Güneş enerjisi üretimini etkileyen bir dizi faktör bulunmaktadır. Bunlar arasında güneş ışınlarının yoğunluğu, hava koşulları, coğrafi konum, mevsimsel değişiklikler, günün saati, yüzey eğimi ve temizlik durumu gibi etkenler yer almaktadır. Bu dezavantajlar, fazla enerjinin depolanmasını zorlaştırarak ve eksik enerjinin maliyetlerini artırarak enerji üretiminde sorunlara yol açabilir. Makine öğrenimi algoritmaları, bu verileri kullanarak hava koşullarının güneş enerjisi üretimine olan etkisini analiz edebilir ve gelecekteki enerji üretim miktarını tahmin edebilir.

Sonuç olarak, güneş enerjisi verileri ile meteoroloji verilerinin birleştirilmesi, güneş enerjisi üretimini tahmin etmek, sistemin verimliliğini artırmak ve daha sürdürülebilir enerji stratejileri geliştirmek için önemli bir kaynak sunmaktadır. Bu verilerin doğru bir şekilde analiz edilmesi ve makine öğrenimi yöntemlerinin kullanılması, güneş enerjisi sektörünün gelecekte daha verimli ve etkili bir şekilde çalışmasına katkı sağlayabilir.

Bu çalışma, güneş enerji santrallerinden elde edilen verilerin, meteoroloji verileri ile birleştirilerek makine öğrenimi yöntemleri ile analiz edileceği bir proje içermektedir. Şimdiye kadar dünya genelinde ve Türkiye'de bu amaç doğrultusunda birçok çalışma, farklı coğrafi bölgelerde istatistiksel yöntemlerin yanı sıra bulanık mantık ve yapay sinir ağları gibi farklı yaklaşımlarla gerçekleştirilmiştir. Bu çalışma, kullanılan GES verilerinin ilk kez kullanılması, GES verileri ile meteoroloji verilerinin birleştirilmesi, enerji üretimine başlanan ve sonlandırılan saatlere odaklanarak verilerin sınırlanması gibi özellikleriyle özgün bir nitelik taşımaktadır. Ayrıca, farklı birçok makine öğrenme yöntemlerinin kullanılması da bu çalışmanın benzersizliğini vurgulamaktadır.

Bilecik ilinin Bozüyük ilçesinde yer alan üç farklı güneş enerji santralinin verileri, Bilecik merkezindeki Meteoroloji Müdürlüğü istasyonundan gelen



meteorolojik verilerle birleştirildi. Veri toplama sürecinde ve enerji üretimi başlangıç ve bitiş saatleri arasındaki zaman uyumsuzlukları göz önüne alındığında, saatlik veriler günlük periyotlara dönüştürüldü. Meteoroloji verileri tam günü kapsayacak şekilde kaydedilirken, parametreler arasında uyumsuzluklar gözlemleniyor, enerji üretimi başlangıç ve bitiş saatleri her gün güneşin doğuşu ve batışına göre değişiyordu. Bu nedenle, meteoroloji verileri her gün için enerji üretimi başlangıç ve bitiş saatleri dikkate alınarak tüm veriler tek bir tabloda birleştirildi.

Veri seti, Jupyter Notebook programı kullanılarak içe aktarıldı ve kapsamlı bir analiz sürecinden geçirildi. Bu analizler sonucunda eksik veriler tespit edildi ve bu eksik verilerin doldurulması gerçekleştirildi. Eksik verilerin, ayın mevsimsel etkisi ve tüm yılın mevsimsel koşullarıyla uyumsuz olabileceği göz önüne alındı. Bu nedenle eksik veriler, ilgili ayın mod, medyan veya ortalama değerleri kullanılarak dikkatle dolduruldu.

Veri setinde yer alan kategorik veriler, sayısal verilere dönüştürmek için One-Hot Encoder kullanıldı ve bu veriler arasındaki sayısal farklılıkları gidermek için normalizasyon işlemi uygulandı. Normalizasyon işlemlerinin performans değerlerini olumlu etkilediği gözlemlendi. Sonuç olarak, 35 öznitelikten ve 3782 satırdan oluşan bir veri seti elde edildi.

Ayrıca, veri seti üzerinde yapılan analizler sonucunda Pearson Korelasyonu kullanılarak öznitelikler arasındaki ilişkiler değerlendirildi. 0.95 ve üzeri ilişkiye sahip olan 8 öznitelik, veri setinden çıkarılarak, veri setinin makine öğrenimi algoritmaları için daha uygun hale getirilmesi sağlandı. Bu sayede veri seti, modelleme süreci için hazır hale getirildi.

Modelleme işlemine başlamadan önce, veri seti eğitim ve test verileri olarak ayrıldı. Veriyi rastgele bölmek, modelin performansını etkileyebileceğinden, her makine öğrenimi algoritması için çapraz doğrulama yöntemi kullanıldı. Her model için en iyi çapraz doğrulama katman sayısı ayrı ayrı belirlendi. Bu yaklaşım, modelin performansını daha güvenilir bir şekilde değerlendirmeye olanak tanır.

Rastgele Orman, Karar Ağaçları, K-En Yakın Komşu, Çoklu Doğrusal Regresyon, Ridge Regresyon, Lasso Regresyon ve Doğrusal Regresyon olmak üzere

yedi farklı model kullanılarak modelleme işlemi gerçekleştirilmiştir. Mevcut modellerin sonuçları değerlendirilmiş ve başarı karşılaştırmalarına odaklanılmıştır.

Bu sonuçlara göre, Rastgele Orman algoritması en yüksek  $R^2$  puanı olan 0.968420'yi elde etmiştir. Ayrıca, MSE değeri 0.000764 ve MAE değeri 0.012567 olarak hesaplanmıştır. Modelin eğitilme süresi 4.246773 dakika sürmüştür, test süresi ise 0.31250 dakikadır.  $R^2$  puanının 1'e çok yakın olması ve hata metriklerinin sıfıra en yakın olması, bu modelin veri seti üzerinde etkili bir şekilde çalıştığını ve tahmin edilen değerlerin gerçek değerlere çok yakın olduğunu göstermektedir. Bu nedenle, Rastgele Orman algoritması, veri setinin üzerinde en iyi performans gösteren algoritma olarak belirlenmiştir. Ancak, modelin eğitilme süresi diğerlerine oranla oldukça yüksektir.

Karar Ağaçları algoritması,  $R^2$  puanı 0.955562 ile ikinci en yüksek değeri elde etmiştir. Ayrıca, MSE değeri 0.001075 ve MAE değeri 0.016059 olarak hesaplanmıştır. Modelin eğitilme süresi 0.078122 dakika sürmüştür, test süresi ise 0.015624 dakikadır. Bu nedenle, Karar Ağaçları algoritması, veri seti üzerinde ikinci en iyi performansı gösteren algoritma olarak belirlenmiştir. Modelleme süresi, Rastgele Orman Algoritması'na göre daha kısa olmasına rağmen  $R^2$  puanı daha düşüktür.

Bu iki model, en iyi tahminleri gerçekleştiren algoritmalar olarak kabul edilirken, diğer algoritmaların başarı sıralaması şu şekildedir: Üçüncü sırada Çoklu Doğrusal Regresyon, dördüncü sırada K-En Yakın Komşu, beşinci sırada Ridge Regresyon, altıncı sırada Lasso Regresyon ve yedinci sırada Doğrusal Regresyon yer almıştır.

Her bir model için gerçek değerlerle tahmin edilen değerler arasındaki farkı gösteren grafikler çizdirildi. Bu grafiklerde gerçek değerlerin yeşil renkle temsil edildiği ve tahmin edilen değerlerin sarı renkle temsil edildiği görülmektedir. Bu grafikler incelendiğinde, Rastgele Orman algoritmasının gerçek değerlerle tahmin edilen değerler arasındaki farkın en az olduğu, Rastgele Orman algoritmasının iyi tahminlerde bulunduğu gözlemlenmiştir.

Rastgele Orman algoritmasından elde edilen tahmin sonuçları, gerçek değerlerle karşılaştırılarak modellerin tahmin başarısı değerlendirildi ve gelecekteki enerji

üretimini ne kadar doğru tahmin edebileceği gözlemlendi. Bu model, veri setini rastgele örnekler alarak böldüğü için sınıflandırmada bağımsız ağaçlar üretti ve aşırı uyum eğilimini azalttı (Akar ve Güngör, 2012). Ancak, modelin eğitim süresinin uzun olması bir dezavantaj olarak gözlemlenmiştir.

Bu çalışmanın sonuçları ve bulguları, enerji sektöründe sürdürülebilir dönüşümü teşvik etmek ve yenilenebilir enerji kaynaklarının etkin bir şekilde kullanılmasını desteklemek amacıyla bir dizi öneri sunmaktadır.

Endüstri ile yakın işbirlikleri kurularak, yenilenebilir enerji projeleri için özel sektörle ortaklıklar geliştirilmelidir. Sanayi ve enerji sektörleri arasındaki işbirlikleri, yenilenebilir enerji projelerinin daha hızlı geliştirilmesine ve uygulanmasına olanak tanır.

Meteoroloji Müdürlüğü ile sürekli bir işbirliği sürdürülmeli ve güncel iklim verileri kullanılarak enerji üretim potansiyeli analizleri düzenli olarak güncellenmelidir. Ayrıca Türkiye genelinde Güneş Enerjisi Santralleri (GES) bölgeleri belirlenmeli ve bu bölgelerin fizibilite analizleri yapılmalıdır. Bu, yenilenebilir enerji yatırımlarının en uygun bölgelerde gerçekleştirilmesini sağlar.

Bu çalışma, sadece güneş enerjisi ile sınırlı kalmamalıdır. Rüzgar enerjisi gibi diğer yenilenebilir enerji kaynaklarına odaklanan benzer çalışmalar da gerçekleştirilmelidir. Bu, enerji arzının çeşitlendirilmesini teşvik eder ve enerji güvenliğini artırır.

Karar Ağaçları algoritması, hem modelleme başarısı hem de eğitim süresindeki kısalığı ile dikkat çekmektedir. Bununla birlikte, Karar Ağaçları algoritmasında ağaç derinliklerindeki sınırlılık, aşırı uyum sorunlarına neden olabilir. Gelecekteki çalışmalarda, Karar Ağaçları algoritması kullanılırken ağaçların derinliği ve diğer parametreler üzerinde daha detaylı bir inceleme yapılabilir. Ayrıca, enerji üretimi tahmininde daha iyi sonuçlar elde etmek için Çok Katmanlı Algılayıcılar (Multi-Layer Perceptron) algoritması gibi farklı yöntemlere başvurularak sınıflandırma ve genelleme sorunlarına daha etkili çözümler üretebilir.

Güneş enerji santrallerinde bakım ve temizlik zamanlarının kayıt altına alınması ve bu verilerin kullanılması, otomatik bakım ve temizlik işlemlerini gerçekleştirebilecek otomasyonların geliştirilmesine katkı sağlayabilir. Ayrıca,

güneşin konumuna baęlı olarak deęişen otomasyonların uygulanması ve bu deęişkenlerin enerji üretimi üzerindeki etkilerinin tahmin edilmesi, verimlilięi artırmak için önemli bir adım olabilir.

Bu öneriler, enerji sektörünün sürdürülebilirlik açısından daha güçlü bir konuma gelmesine yardımcı olacaktır. Ayrıca, Türkiye'nin enerji ihtiyacını çevre dostu ve sürdürülebilir bir şekilde karşılama kapasitesini artırmak için önemli adımlar atılmasına olanak tanır.



## KAYNAKÇA

- Akalın, F. (2020). İnce bağırsak görüntüleri üzerinde sezgisel algoritma teknikleri ile polip teşhisi (Master's thesis, Sakarya Üniversitesi).
- Akar Ö. ve Güngör, O. (2012). Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması. *Jeodezi ve Jeoinformasyon Dergisi*, (106), 139-146.
- Akman, T., (2018). Yapay Zeka Kullanarak Ankara Bölgesikisa Dönem Elektrik Enerjisiyük Tahmini
- Altuntop, N. ve Erdemir, D. (2013). Dünyada ve Türkiye’de güneş enerjisi ile ilgili gelişmeler. *Mühendis ve Makina*, 69-77.
- Arslankaya, S., ve Toprak, Ş. (2021). Makine öğrenmesi ve derin öğrenme algoritmalarını kullanarak hisse senedi fiyat tahmini. *International Journal of Engineering Research and Development*, 13(1), 178-192
- Aslan, S., ve Yıldız, T. (2022). Makine Öğrenmesinde Rastgele Oran ve Sıralı Küme Örneklemesi Yöntemlerinin Doğrusal Regresyon Modellerine Etkisi. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 24(70), 29-36.
- Behçet, R., Oral, H. ve Gül, H. (2013). Adıyaman ilinin güneş enerjisi potansiyeli ve kullanımı. *Batman Üniversitesi Yaşam Bilimleri Dergisi*, 3(2), 52-67.
- Bekçioğulları, M. F., Dikici, B., Açıkgoz, H., ve Keçecioğlu, Ö. F. (2021). Güneş Enerjisinin Kısa-Dönem Tahmininde Farklı Makine Öğrenme Yöntemlerinin Karşılaştırılması. *EMO Bilimsel Dergi*, 11(22), 37-45.
- Breiman L. ve Cutler A., (2005), Random forest, Erişim Adresi: [http://www.stat.berkeley.edu/~breiman/RandomForests/cc\\_home.htm](http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm), Erişim Tarihi: 12 Temmuz 2011.
- Carpita, M. ve Manisera, M. (2011). On the imputation of missing data in surveys with Likert-type scales. *Journal of Classification*, 28, 93-112.
- Caruana, R. ve Niculescu-Mizil, A. (2006, June). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning* (pp. 161-168).
- Chai, T. ve Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE). *Geoscientific model development discussions*, 7(1), 1525-1534.
- Cunningham, P. ve Delany, S. J. (2021). K-nearest neighbour classifiers-a tutorial. *ACM Computing Surveys (CSUR)*, 54(6), 1-25.
- Dasu, T. ve Johnson, T. (2003). *Exploratory Data Mining and Data Cleaning*. Wiley-IEEE.

- Demir, E. ve Parlak, B. (2012). Türkiye’de eğitim arařtırmalarında kayıp veri sorunu. *Journal of Measurement and Evaluation in Education and Psychology*, 3(1), 230-241.
- Demolli, H., Dokuz, A., Gokcek, M. ve Ecemiř, A. (2019). Makine Öğrenmesi Algoritmalarıyla Güneř Enerjisi Tahmini: Niğde İli Örneęi. *In International Turkic World Congress on Science and Engineering*, ss (Vol. 783).
- Dilki, G. ve Bařar, Ö. D. (2020). İřletmelerin İflas Tahmininde K-En Yakın Komřu Algoritması Üzerinden Uzaklık Ölçütlerinin Karřılařtırılması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 19(38), 224-233.
- Dinçer, İ., Arçaklıoęlu, E. ve Ezan M.A. (2022). Enerjide Yapay Zekanın Rolü Raporu. TÜBA, 13-15
- Donat, W. (2015). What is Python: An Intro to a Cross-Platform Programming Language, Eriřim Adresi: <https://www.atlantic.net/vps-hosting/what-is-python-intro-crossplatform-programming-language/>: Eriřim Tarihi: 16.05.2020.
- Dondurmacı, G.A., Çınar, A. (2014). Finans Sektöründe Veir Madenciligi Uygulaması, *Akademik Sosyal Arařtırmalar Dergisi*, 2(1): 258-271.
- Düzgün, B. (2022). Enerji Verimlilięinde Dijital Dönüřüm. TÜBA - *Enerjide Yapay Zekânın Rolü Çalıřtayı ve Paneli*, TÜBA
- EİE, (2012). Eriřim Adresi: <http://www.eie.gov.tr>
- Erdil, A. ve Arçaklıoęlu, E. (2013). The prediction of meteorological variables using artificial neural network. *Neural Computing and Applications*, 22(7-8), 1677-1683. doi:10.1007/s00521-012-1210-0
- Eyvazov, A. (2019). Yapay sinir aęları ile güneř enerjisi potansiyelinin modellenmesi ve güneř pilleri verim analizi
- Frank, L. E. ve Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35(2), 109-135.
- Gabralı, D. (2019). Yapay Sinir Aęları İle Rüzgar Ve Güneř Enerjisi Potansiyelinin Modellenmesi Ve Dalgacık Dönüřümü İle Analizi
- Gök, A. O., (2019). Güneř enerjisi santralleri için bir kısa süreli üretim tahmin sistemi geliřtirilmesi
- Gökalp, Ö., Makine Öğrenmesi. Eriřim Adresi: [https://www.researchgate.net/profile/OemerGoekalp/publication/357974983\\_Makine\\_Ogrenmesi\\_\\_Machine\\_Learning/links/61e9c6fec5e3103375ac6a8f/Makine-Oegrenmesi-MachineLearning.pdf](https://www.researchgate.net/profile/OemerGoekalp/publication/357974983_Makine_Ogrenmesi__Machine_Learning/links/61e9c6fec5e3103375ac6a8f/Makine-Oegrenmesi-MachineLearning.pdf)

- Gokpinar, E., Ebegil, M. ve Gokpinar, F. (2017). A Review on Shrinkage Parameters in Ridge Regression. *GU Journal of Science*, 30(4), 565– 582. Eriřim adresi: <https://dergipark.org.tr/tr/download/articlefile/380312>
- Gültepe, Y. (2019). Makine öğrenmesi algoritmaları ile hava kirlilięi tahmini üzerine karşılařtırmalı bir deęerlendirme. *Avrupa Bilim ve Teknoloji Dergisi*, (16), 8-15.
- Gupta, P. 2017. Regularization in Machine Learning. Eriřim Adresi: <http://www.towardsdatascience.com/regularization-in-machine-learning-76441ddcf> , Eriřim Tarihi: 03.04.2021.
- Haba, C. G. (2019, May). Monitoring solar panels using machine-learning techniques. *In 2019 8th international conference on modern power systems (MPS)* (pp. 1-6). IEEE.
- Hastie, T., Friedman, J. ve Tibshirani, R., (2008), Unsupervised Learning. In: *The Elements of Statistical Learning, Springer Series in Statistics. Springer, New York*.
- Hegde, C., Wallace, S., and Gray, K. (2015). Using trees, bagging, and random forests to predict rate of penetration during drilling. *In SPE Middle East Intelligent Oil and Gas Conference and Exhibition*. OnePetro.
- Hlasny, V., Ceriani, L. ve Verme, P. (2022). Bottom incomes and the measurement of poverty and inequality. *Review of Income and Wealth*, 68(4), 970-1006.
- Hoerl, A. E. ve Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55–67. doi: <https://doi.org/10.1080/00401706.1970.10488634>
- Huang, J., Korolkiewicz, M., Agrawal, M., & Boland, J. (2013). Forecasting solar radiation on an hourly time scale using a Coupled AutoRegressive and Dynamical System (CARDS) model. *Solar Energy*, 87, 136-149. doi:10.1016/j.solener.2012.10.012
- Jayalakshmi, T. & Santhakumaran A. (2011). Statistical normalization and back propagation for classification. *International Journal of Computer Theory and Engineering*, 3(1): 89-93
- Kaplan, K., 2019, Python ve Anaconda Kurulumu, Eriřim Adresi: <https://medium.com/kodcular/python-ve-anaconda-kurulumu-b8931bd80e64>: Eriřim Tarihi: 07.03.2020.
- Kara, A. (2019). Global solar irradiance time series prediction using long short-term memory network. *Gazi Üniversitesi Fen Bilimleri Dergisi, Part C: Tasarım ve Teknoloji*, 4(7), 882-892.
- Kaya, K. ve Koç, E. (2015). Enerji Kaynakları-Yenilenebilir Enerji Durumu. *Mühendis ve Makina*, 56(668).

- Khalyasmaa, A., Eroshenko, S., ve Tran, D. C. (2020). Very-short term forecasting of photovoltaic plants generation based on meteorological data from open sources using machine learning. *In 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)* (pp. 564-568). IEEE.
- Kılıç, S. (2013). Doğrusal regresyon analizi. *Journal of Mood Disorders*, 3(2), 90-92.
- Kıymaz, Y.E. (2020). Güneş enerjisi santrallerinde derin öğrenme kullanılarak elektrik üretimi tahmininin yapılması
- Kutner, M. H., Nachtsheim, C. J., Neter, J. ve Li, W. (2005). *Applied linear statistical models*, (McGraw-Hill Irwin: New York).
- Leeuw, J. D. (2009). *Journal of statistical software. Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 128-129.
- Little, R.J.A. ve Rubin, D.B. (2002). *Statistical Analysis with Missing Data: Second Edition*. John Wiley and Sons.
- Mahmud, K., Azam, S., Karim, A., Zobaed, S., Shanmugam, B. ve Mathur, D. (2021). Machine learning based PV power generation forecasting in alice springs. *IEEE Access*, 9, 46117-46128.
- Metin, S. (2021). Benzetimli tavlama algoritması ile eksik veri tamamlama. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 33(1), 295-301.
- Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W. ve Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*, 10, 1-16.
- Nacar, E. N. ve Erdebilli, B. (2021). Makine Öğrenmesi Algoritmaları İle Satış Tahmini. *Endüstri Mühendisliği*, 32(2), 307-320.
- Oğuz, D. (2018). Güneş enerjisi sistemleri ve EMF ölçümleri analizi
- Okolo, C. T. (2022). *Diabetes Prediction Using Machine Learning Algorithm*. University of Louisiana at Lafayette.
- Okur, Y. (2016). Akdeniz bölgesine ait meteorolojik veriler kullanılarak yapay sinir ağları yardımıyla güneş enerjisinin tahmini
- Ostertagová, E. (2012). Modelling using polynomial regression. *Procedia Engineering*, 48, 500-506. <https://doi.org/10.1016/j.proeng.2012.09.545>
- Özdoğan, N., (2021). Hava Durumunun Enerji Tedariki Üzerindeki Etkisi - Kahramanmaraş Adiyaman Bölgesindeki Güneş Enerji Santrallerinin İncelenmesi Örneği



- Özen, N. S., Saraç, S. ve Koyuncu, M. (2021). COVID-19 vakalarının makine öğrenmesi algoritmaları ile tahmini: *Amerika Birleşik Devletleri örneği. Avrupa Bilim ve Teknoloji Dergisi*, (22), 134-139.
- Özşahin, M., Yılmaz, K. B. ve Akcan, S., (2022). Makine Öğrenmesi Algoritmaları İle Ev Kirasi Tahmini1.
- PSI, Pangea Strategic Intelligence Makine Öğrenmesi ve Yenilenebilir Enerji, Erişim Adresi: <https://www.pangea-si.com/machine-learning-and-renewable-energy/> Erişim Tarihi: 10 Temmuz 2021, 23:00
- Philibert, C. (2005). The Present and Use of Solar Thermal Energy as a Primary Source of Energy. *Solar Thermal Energy, IEA Yayını*, 16.
- Potdar, K., Pardawala, T. S. ve Pai, C. D. (2017). A comparative study of categorical variable encoding techniques for neural network classifiers. *International journal of computer applications*, 175(4), 7-9.
- Seyhan, Ş. (2021). Dayanma duvarlarında makine öğrenmesi ile güvenlik sayısı tahmini.
- Sezgin E, Çelik Y., (2013). Veri Madenciliğinde Kayıp Veriler İçin Kullanılan Yöntemlerin Karşılaştırılması, Akademik Bilişim Konferansı, Akdeniz Üniversitesi.
- Shi, J., Lee, W. J., Liu, Y., Yang, Y. ve Wang, P. (2012). Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Transactions on Industry Applications*, 48(3), 1064-1069.
- Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake ve Moore, R. (2013). Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1), 116-124.
- Sinap, V., (2023) Makine Öğrenmesi Teknikleri ile Counter-Strike: Global Offensive Raunt Sonuçlarının Tahminlenmesi. *Journal of Intelligent Systems: Theory and Applications*, 6(2), 119-129.
- Sivakumar, S., Neeraja, B., Jamuna Rani, M., Anandaram, H., Ramya, S., Padhan, G., ve Gurusamy, S. (2022). Machine learning approach on time series for PV-solar energy. *Advances in Materials Science and Engineering*, 2022.
- Sorhun, E. (2021). Python ile Makine Öğrenmesi. İstanbul: Abaküs Yayınları.
- Strobl, C., Boulesteix, A. L., Zeileis, A. ve Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, 8(1), 1-21.
- Sun, S., Wang, S., Zhang, G. ve Zheng, J. (2018). A decomposition-clustering-ensemble learning approach for solar radiation forecasting. *Solar Energy*, 163, 189-199.

- Şafak, H. İ. (2017). Erişim Adresi: <https://medium.com/https://medium.com/t%C3%BCrkiye/makine-%C3%B6%C4%9Frenmesi-nedir20dee450b56e>
- Şenel, M. C., (2012). Rüzgar Türbinlerinde Güç İletim Mekanizmalarının Tasarım Esasları-Dinamik Davranış. Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Samsun.
- Şener Y. Veri Biliminde Eksik/Kayıp Verilere Yaklaşım Stratejileri ve Python (Pandas) Uygulaması, 2020
- Taktak, F. ve Ilı, M. (2018). Güneş Enerji Santrali (GES) Geliştirme: Uşak Örneği. *Geomatik*, 3(1), 1-21.
- Taşkın, O. ve Korucu, T. (2014). Kahramanmaraş ili güneş enerjisi potansiyeli ve kullanım olanakları. *Tarım ve Doga Dergisi*, 17(4), 12.
- Tektaş, N. ve Korkmaz, E. (2023, May). Makine Öğrenmesi Algoritması ile Çanakkale İlinin Güneş Işınım Şiddetinin Analizi. *In International Conference on Contemporary Academic Research* (Vol. 1, pp. 48-52).
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267-288.
- Uğuz, S., Oral, O. ve Çağlayan, N. (2019). PV Güç Santrallerinden Elde Edilecek Enerjinin Makine Öğrenmesi Metotları Kullanılarak Tahmin Edilmesi. *International Journal of Engineering Research and Development*, 11(3), 769-779.
- Uyanık, F. ve Kasapbaşı, M. C. (2021). Telekomünikasyon sektörü için veri madenciliği ve makine öğrenmesi teknikleri ile ayrılan müşteri analizi. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 9(3), 172-191.
- Wei, C. C. (2017). Predictions of surface solar radiation on tilted solar panels using machinelearning models: A case study of Tainan city, Taiwan. *Energies*, 10(10), 1660.
- Wolff, B., Kühnert, J., Lorenz, E., Kramer, O. ve Heinemann, D. (2016). Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. *Solar Energy*, 135, 197-208.
- Yapraklı, T.Ş. ve Erdal, H. (2016). Firma Başarısızlığı Tahminlemesi: *Makine Öğrenmesie Dayalı Bir Uygulama*, *Bilişim Teknolojileri Dergisi*, 9(1): 21-31.
- Yavuz, C. (2018). Konya'da kısa dönemli güneş enerjisi üretim tahmini
- Ye C., Wang H., Li J., Gao, H. ve Cheng, S. (2016). Crowdsourcing-Enhanced missing values imputation based on Bayesian network, in: *International Conference on Database Systems for Advanced Applications*, Springer: 67–81.

- Yeşil, M. (2022). İletim Sisteminde Yapay Zeka Uygulamaları. TÜBA - Enerjide Yapay Zekânın Rolü Çalıştayı ve Paneli, TÜBA - Enerjide Yapay Zekânın Rolü Çalıştayı ve Paneli.
- Yetiz, F., Terzioğlu, M. ve Kayakuş, M. (2021). Makina Öğrenmesi Yöntemleri ile Türk Mevduat Bankalarının Müşteri Tahminine Yönelik Bir Uygulama. *Sosyoekonomi*, 29(50), 413-432.
- Yıldız, C. ve Açıkgöz, H. (2021). A kernel extreme learning machine-based neural network to forecast very short-term power output of an on-grid photovoltaic power plant. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 43(4), 395-412.
- Yılmaz, M. (2012). Türkiye' nin enerji potansiyeli ve yenilenebilir enerji kaynaklarının elektrik enerjisi üretimi açısından önemi. *Ankara Üniversitesi Çevre Bilimleri Dergisi*, 4(2), 33-54.
- Zhang, S., Qin, Y., Zhu, X., Zhang, J. ve Zhang, C. (2006, August). Optimized parameters for missing data imputation. *In Pacific Rim International Conference on Artificial Intelligence* (pp. 1010-1016). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Zhang, Y., Yang, S., Guo, Z., Guo, Y. ve Zhao, J. (2019). Wind speed forecasting based on wavelet decomposition and wavelet neural networks optimized by the Cuckoo search algorithm. *Atmospheric and Oceanic Science Letters*, 12(2), 107-115. doi:10.1080/16742834.2019.1569455