**T.C.**
**İSTANBUL ÜNİVERSİTESİ**
**SOSYAL BİLİMLER ENSTİTÜSÜ**
**ÜRETİM VE TEDARİK ZİNCİRİ YÖNETİMİ ANABİLİM DALI**

**YÜKSEK LİSANS TEZİ**

# MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE RÜZGÂR GÜCÜ ÜRETİM TAHMİNİ

**Emrah AKBULUT**
**2501221154**

**TEZ DANIŞMANI**
**Dr. Öğr. Üyesi Halit Alper TAYALI**

**İSTANBUL – 2024**

# ÖZ

## MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE RÜZGÂR GÜCÜ ÜRETİM TAHMİNİ

**EMRAH AKBULUT**

Bir rüzgâr türbininin üreteceği güç miktarının tahmini; türbinin tamir-bakım planlarının değerlendirilmesi, enerji üretim-tüketim dengesinin belirlenmesi gibi pek çok alanda önemlidir. Rüzgâr enerjisinin son derece değişken yapısından kaynaklı enerji sistemindeki belirsizliğin azaltılması ve rüzgâr santrallerinde daha güvenilir enerji üretim tahminleri yapılması için matematik ve istatistik temelli tahmin modelleri kullanılmaktadır. Bu tahmin modellerinin ihtiyacı olan sayısal veriler, rüzgâr türbininin faaliyetlerini gözlemleyip kaydeden sistemler SCADA ("Supervisory Control and Data Acquisition") aracılığı ile toplanmaktadır.

Bu çalışmada bir rüzgâr türbininin SCADA sisteminden aktif güç, teorik güç, rüzgâr yönü ve rüzgâr hızı değişkenlerine ilişkin 10 dakikalık zaman aralığı ile ölçülen sayısal veriler (Erisen, 2018) kullanılmıştır. Bağımlı değişken aktif güç, bağımsız değişkenler ise rüzgâr hızı, teorik güç ve rüzgâr yönü olacak şekilde türbinin ürettiği aktif gücü tahmin eden farklı modellerin hata değerleri ve performans değerleri hesaplanmıştır. Makine Öğrenmesi yöntemlerinden Karar Ağacı ("Decision Tree"), Rastgele Orman ("Random Forest"), Aşırı Gradyan Arttırma ("XGBoost") ve Çoklu Doğrusal Regresyon yöntemi ile farklı tahmin modelleri kurularak, bu modellerin performansları tahmin hata değerlerinin hesaplanması ve karşılaştırılması sonucunda en iyi performans veren tahmin yönteminin belirlenmesi amaçlanmıştır.

**Anahtar Kelimeler:** Çoklu Doğrusal Regresyon, Karar Ağacı, Rastgele Orman, XGBoost, SCADA Sistemi Verileri, Tahmin Modelleri

# ABSTRACT


**WIND POWER GENERATION FORECASTING USING MACHINE LEARNING METHODS**

**EMRAH AKBULUT**

Estimating the amount of power a wind turbine will generate is important in many areas, such as evaluating maintenance and repair schedules, determining the balance of energy production and consumption, and more. Due to the highly variable nature of wind energy, mathematical and statistical-based prediction models are used to reduce uncertainty in the energy system and provide more reliable energy production forecasts for wind farms. The numerical data required by these prediction models are collected through SCADA systems that monitor and record the operations of the wind turbine.

In this study, numerical data measured at 10-minute intervals related to the variables of active power, theoretical power, wind direction, and wind speed from the SCADA system of a wind turbine (Erisen, 2018) were used. The dependent variable was active power, while the independent variables were wind speed, theoretical power, and wind direction. The aim was to calculate the error values and performance metrics of different models predicting the active power produced by the turbine. Prediction models were established using Machine Learning methods such as Decision Tree, Random Forest, Extreme Gradient Boosting (XGBoost), and Multiple Linear Regression. The objective was to determine the best-performing prediction method by calculating and comparing the prediction error values of these models.



**Keywords:** Multiple Linear Regression, Decision Tree, Random Forest, XGBoost, SCADA System Data, Prediction Models

# ÖNSÖZ

Bu tez çalışmasının önemi ve amacı rüzgâr türbinlerinde üretilen gücün doğru tahmin edilmesi ile türbinin bakım ve işletmesinin doğru planlanmasına, üretilen elektrik enerjisinin doğru tahmin edilmesi sayesinde rüzgâr türbini işletmecilerine kılavuzluk edecek bir kaynak niteliğinde olmasıdır. Çalışmamızın kapsamı ise SCADA sisteminden alınan verilerin tahmin yöntemlerinden Makine Öğrenmesi tekniklerini kullanarak çözdüğümüz Rüzgâr Gücü Üretim Tahmini probleminin incelenmesidir.

Bu çalışmanın ortaya çıkış sürecinde destek ve katkılarını hiçbir zaman esirgemeyen, tez danışmanım sayın Dr. Öğr. Üyesi Halit Alper TAYALI' ya teşekkür ederim. Bilgi, belge ve çalışmalarını tereddütsüz paylaşması benim için çok önemliydi.

Üretim ve Tedarik Zinciri Yönetimi Ana Bilim Dalı'ndaki tüm hocalarıma ayrı ayrı teşekkür eder saygılarımı sunarım.

Hayatım boyunca her anımda yanımda oldukları ve verdikleri tüm emekler için Annem' e ve Babam' a ve sevgili Eşim' e sevgilerimle…

Emrah AKBULUT
İSTANBUL, 2024

# İÇİNDEKİLER

**ÖZ** ..... **ii**

**ABSTRACT** ..... **iii**

**ÖNSÖZ** ..... **iv**

**TABLOLAR LİSTESİ** ..... **vii**

**ŞEKİLLER LİSTESİ** ..... **viii**

**KISALTMALAR LİSTESİ** ..... **ix**

**GİRİŞ** ..... **1**

**BİRİNCİ BÖLÜM**

**RÜZGÂR ENERJİSİ**

1.1. DÜNYADA ve TÜRKİYE'DE RÜZGÂR ENERJİSİ ..... 2
1.2. RÜZGÂR TÜRBİNİ YAPISI VE ÇALIŞMA PRENSİBİ ..... 4
1.2.1. Rüzgârın Teorik Güç ve Aktif Güç Hesabı ..... 5

**İKİNCİ BÖLÜM**

**LİTERATÜR TARAMASI**

2.1. RÜZGÂR GÜCÜ ÜRETİM TAHMİNİNİN TANIMI, ÖNEMİ VE YAPILAN ÇALIŞMALAR ..... 7

**ÜÇÜNCÜ BÖLÜM**

**DOĞRUSAL REGRESYON VE MAKİNE ÖĞRENMESİ YÖNTEMLERİ**

3.1. DOĞRUSAL REGRESYON TANIMI ..... 13
3.1.1. Çoklu Doğrusal Regresyon ..... 14
3.2. MAKİNE ÖĞRENMESİ ..... 15
3.2.1. Karar Ağaçları ..... 16
3.2.2. Torbalama ..... 18
3.2.3. Arttırma ..... 19
3.2.3.1. XGBoost ..... 21
3.2.4. Rastgele Orman ..... 22

## DÖRDÜNCÜ BÖLÜM

## VERİ SETİ ANALİZİ

4.1. VERİ SETİ .....27

## BEŞİNCİ BÖLÜM

## R PROGRAMLAMA DİLİNDE UYGULAMA

5.1. MODEL PERFORMANSI VE HATA DEĞERLERİ.....30
5.1.1. Çapraz Doğrulama ile Model Performansı Değerlendirme.....31
5.2. R PROGRAMLAMA DİLİNDE ÇOKLU DOĞRUSAL REGRESYON.....32
5.2.1. Çoklu Doğrusal Regresyonda Gerçek ile Tahmin Değerleri Grafiği ve Varsayımların Kontrolü .....34
5.3. R PROGRAMLAMA DİLİNDE RASTGELE ORMAN .....38
5.3.1. Rastgele Orman Yöntemi için Gerçek ile Tahmin Değerleri Grafiği .....40
5.4. R PROGRAMLAMA DİLİNDE KARAR AĞACI .....41
5.4.1. Karar Ağacı Yöntemi için Gerçek ile Tahmin Değerleri Grafiği .....43
5.5. R PROGRAMLAMA DİLİNDE XGBOOST.....44
5.5.1. XGBoost Yöntemi için Gerçek ile Tahmin Değerleri Grafiği.....48

**SONUÇ.....50**

**KAYNAKÇA.....52**

# TABLOLAR LİSTESİ

**Tablo 1 :** Zaman ölçeğine göre sınıflandırılmış rüzgâr gücü tahmininin uygulama alanları (Roungkvist & Enevoldsen, 2020).....12
**Tablo 2 :** Çalışmada uygulanacak yöntemlerin kıyaslanması. .....26
**Tablo 3 :**Veri setinin istatistiksel ölçümleri (Singh & Rizwan, 2022).....27
**Tablo 4 :**Uygulanan modellerin hata değerleri sonuç tablosu. .....49

# ŞEKİLLER LİSTESİ

**Şekil 1 :**Türkiye'de yıllara göre rüzgâr enerjisi kurulu gücü değişim grafiği (T.C. Enerji ve Tabii Kaynaklar Bakanlığı, 2022). ..... 3
**Şekil 2 :**Türkiye'de yıllara göre rüzgâr enerjisinden sağlanan elektrik enerjisi üretim değişim grafiği (T.C. Enerji ve Tabii Kaynaklar Bakanlığı, 2022). ..... 3
**Şekil 3 :**Rüzgâr türbini bileşenleri (Ahlström, 2002). ..... 4
**Şekil 4 :**Doğrusal regresyon grafiği (Chatterjee & Hadi, 2012). .....14
**Şekil 5 :**Çoklu regresyon canlandırması (James vd., 2023).....15
**Şekil 6 :** Karar ağacı regresyon örneği (Swetapadma, 2016). .....17
**Şekil 7 :** Arttırma yönteminin aşamaları (Jankowski vd., 2011). .....20
**Şekil 8 :** Veri setindeki bağımlı ve bağımsız değişkenlerin aralarındaki ilişkiyi gösteren grafikler (Singh & Rizwan, 2023). .....28
**Şekil 9 :**Polar koordinat sisteminde rüzgâr hızı, rüzgâr yönü ve güç üretimi arasındaki ilişkiyi gösteren diyagram (Singh & Rizwan, 2023).....29

# KISALTMALAR LİSTESİ

**BI-LSTM** : Bidirectional-Long Short-Term Memory (Çift Yönlü Uzun Kısa Süreli Bellek)

**CNN** : Convolutional Neural Network (Evrişimsel Sinir Ağı)

**GBM** : Gradient Boosting Machine (Gradyan Artırma Makinesi)

**GRU** : Gated Recurrent Unit (Kapılı Tekrarlayan Birim)

**KNN** : K Nearest Neighbors (K En Yakın Komşu)

**LSTM** : Long Short Term Memory (Uzun Kısa Süreli Bellek)

**MAE** : Mean Absolute Error (Ortalama Mutlak Hata)

**MAPE** : Mean Absolute Percentage Error (Ortalama Mutlak Yüzde Hata)

**MLP** : Multi Layer Perceptron (Çok Katmanlı Algılayıcı)

**MSE** : Mean Squared Error (Ortalama Kare Hata)

**NWP** : Numerical Weather Prediction (Sayısal Hava Tahmini)

**RES** : Rüzgâr Enerji Santrali

**RMSE** : Root Mean Squared Error (Hata Kareler Ortalamasının Karekökü)

**RNN** : Recurrent Neural Network (Tekrarlayan Sinir Ağı)

**SCADA** : Supervisory Control and Data Acquisition (Denetleyici Kontrol ve Veri Toplama)

**SSE** : Sum of Squared Errors (Hata Karelerinin Toplamı)

**TCN** : Temporal Convolutional Network (Zamansal Evrişimsel Ağ)

**XA** : X-Algorithm (X algoritması)

**XGBOOST** : EXtreme Gradient Boosting (Aşırı Gradyan Arttırma)

# GİRİŞ

Ülkeler arasında enerji kaynakları ve rezerv miktarları açısından büyük farklılıklar bulunmaktadır. Enerjinin hem üretimi hem de tüketiminde dış kaynaklara olan bağımlılık politik ve stratejik problemlere yol açarken cari bütçe açığı gibi ekonomik sorunları da beraberinde getirmektedir. Bu nedenle pek çok ülke enerji kaynaklarını çeşitlendirmeye ve rezervlerini artırmaya yönelmektedir. Alternatif enerji kaynakları arasından yenilenebilir olan kaynaklar, özellikle sürdürülebilir avantajları nedeniyle ön plandadır. Elektrik enerjisi, artık fosil yakıtların yanı sıra su, güneş ve rüzgâr gibi yenilenebilir kaynaklar ile üretilebilmektedir.

Rüzgârdan üretilen gücün tahmini, enerji üretim ve tüketiminin planlanması, bakım ve onarım programlarının oluşturulması ve yatırım kararlarının alınmasında kritik bir rol oynar. Rüzgârdan elde edilen elektrik enerjisinin üretim ve tüketim miktarlarına ilişkin sayısal tahminler, elektrik piyasalarında son derecede önemlidir (Karık vd., 2017).

Tahminlerden kaynaklanan hatalar, katılımcıların ceza ödemesine neden olabilir. Sonuç olarak, rüzgâr enerjisi gibi değişken profilli kaynaklardan elektrik üretirken kontrol etmek daha zordur ve doğru tahmin yapmak çok önemlidir. Rüzgâr santrali işletmecisi tahmin hataları nedeniyle gelirlerinin yaklaşık % 10' unu yanlış teklif sonucunda veya uygulanan cezalar nedeniyle kaybetmektedir (Dukpa vd., 2010).

Elektrik piyasasında üretim tahminleri ve tüketim verileri ile üretim maliyetlerine ilişkin fiyat teklifleri, santrallerin çalışma programlarının optimize edilmesine ve elektrik maliyetlerinin düşürülmesine olanak tanır. Üretim ve tüketim dengesi sağlandığında, elektrik sistemlerindeki belirsizlik azalır, bu da elektrik fiyatlarının dengelenmesine ve sistemin daha verimli çalışmasına katkıda bulunur. Rüzgâr enerjisine dayalı tahmin modelleri, rüzgâr kaynaklarının güvenli, ekonomik ve verimli kullanımında kritik bir rol oynamaktadır. Bu tahmin modelleri, güç sistemlerinin planlanmasından rezerv yönetimine ve bakım programlarının düzenlenmesine kadar geniş bir yelpazede kullanılmaktadır. Doğru tahminler, santrallerin gün öncesi piyasalarda tahmin hatalarını azaltarak gelirlerini artırmasına ve enerji dengesizliklerini en aza indirerek maliyetleri düşürmesine imkân sağlar (Karık vd., 2017).

Bu sebeplerden dolayı rüzgâr gücü üretim tahmini seçilecek uygun modeller ile tahmin hata değerlerinin düşürülmesi sonucunda elektrik piyasasına ve türbin işletmecilerine işletim planlamasında kolaylık sağlayacaktır.

# BİRİNCİ BÖLÜM

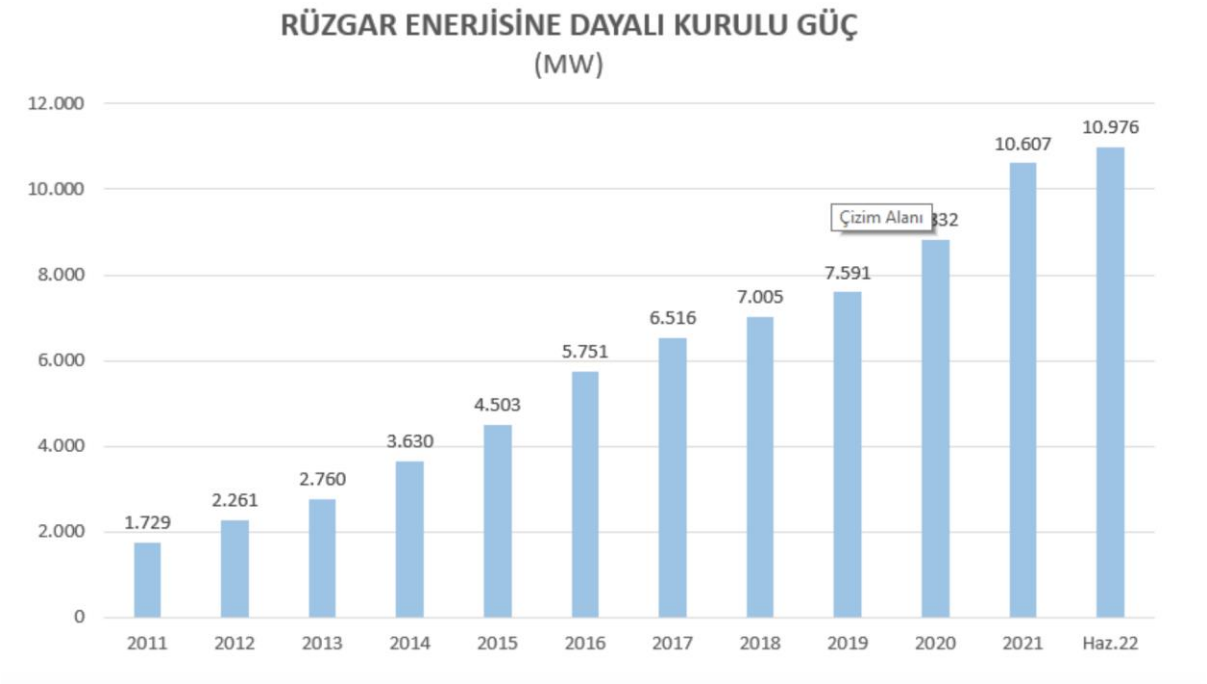
# RÜZGÂR ENERJİSİ

## 1.1. DÜNYADA ve TÜRKİYE'DE RÜZGÂR ENERJİSİ

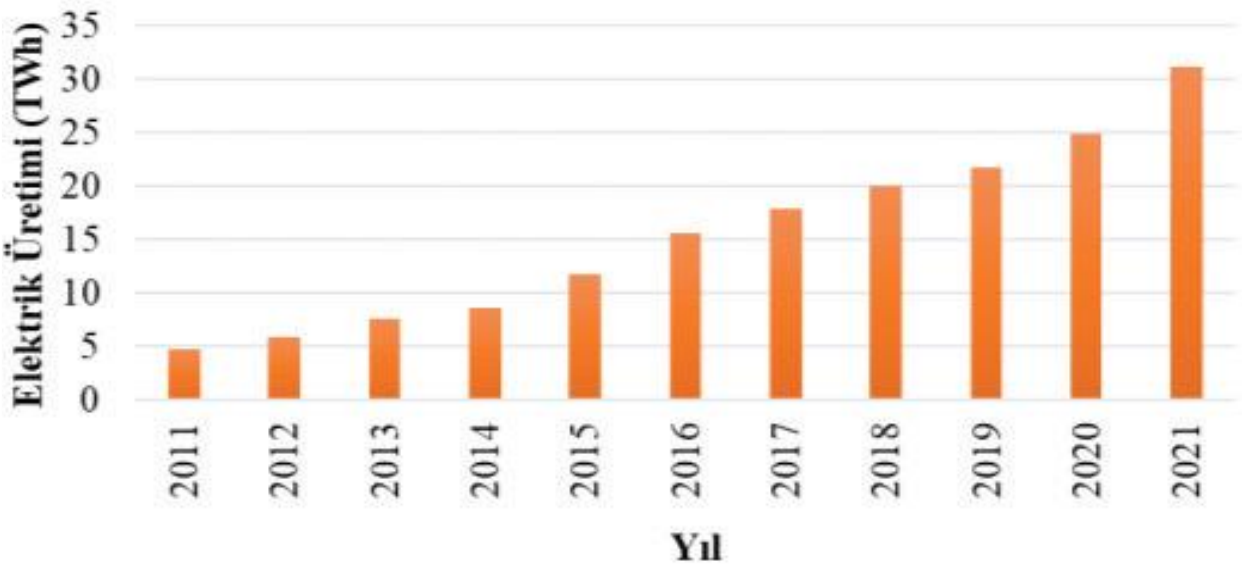
Dünya nüfusunun artması, sanayileşme ve teknolojinin yaygınlaşması, enerji tüketimini hızla artırmış, bu durum fosil yakıt rezervlerinin tükenme riskini gündeme getirmiştir. Bu nedenle, ülkeler yenilenebilir enerji kaynaklarından faydalanmayı arttırmaktadır. Rüzgâr enerjisi yenilenebilir enerji kaynakları arasında, çevre dostu olması ve sürdürülebilirliği nedeniyle giderek daha fazla tercih edilmektedir. Rüzgâr enerjisi, rüzgâr türbinleri aracılığıyla elektrik enerjisine dönüştürülerek kullanılmakta ve teknolojik gelişmeler sayesinde üretim maliyetleri düşürülmüştür. Bu durum, rüzgâr enerjisinin fosil yakıtlarla rekabet edebilir hale gelmesini sağlamış ve birçok ülkenin bu enerji kaynağını devlet politikası olarak desteklemesine yol açmıştır (Şenel & Koç, 2015).

Rüzgâr enerjisi, sürdürülebilir, yenilenebilir ve çevreye zararı olmayan bir enerji türüdür. Rüzgâr türbinleri, hava akımının oluşturduğu kinetik enerjiyi mekanik enerjiye, mekanik enerjiyi de elektrik enerjisine çevirir. Rüzgâr türbinleri karada ve denizde kurulabilmektedir. Rüzgâr enerjisinin olumlu yönleri arasında çevreye zarar vermemesi, tükenme tehlikesinin bulunmaması ve düşük işletme maliyetleri öne çıkarken, olumsuz yönleri yüksek kurulum maliyetleri, düşük enerji kapasitesi ve değişken enerji üretimi olarak sıralanabilir (T.C. Enerji ve Tabii Kaynaklar Bakanlığı, 2022).

Şenel ve Koç (2015) dünyada rüzgâr enerjisi potansiyeli en yüksek bölgeleri sırası ile Kuzey Amerika (14.000 TWh/yıl), Doğu Avrupa ve Rusya (10.600 TWh/yıl), Afrika (10.600 TWh/yıl), Güney Amerika (5.400 TWh/yıl), Batı Avrupa (4.800 TWh/yıl), Asya (4.600 TWh/yıl) ve Okyanusya (3.000 TWh/yıl) olarak belirtirken Türkiye'de faal rüzgâr enerji santrallerinin kurulu güç bakımından en yüksek olduğu bölgeler ise Marmara (1240,88 MW), Akdeniz (543 MW), İç Anadolu (222 MW), Ege (131,1 MW), Karadeniz (80 MW) ve Güney Doğu Anadolu (27,5 MW) olarak belirtilmektedir.




**Şekil 1 :**Türkiye'de yıllara göre rüzgâr enerjisi kurulu gücü değişim grafiği (T.C. Enerji ve Tabii Kaynaklar Bakanlığı, 2022).



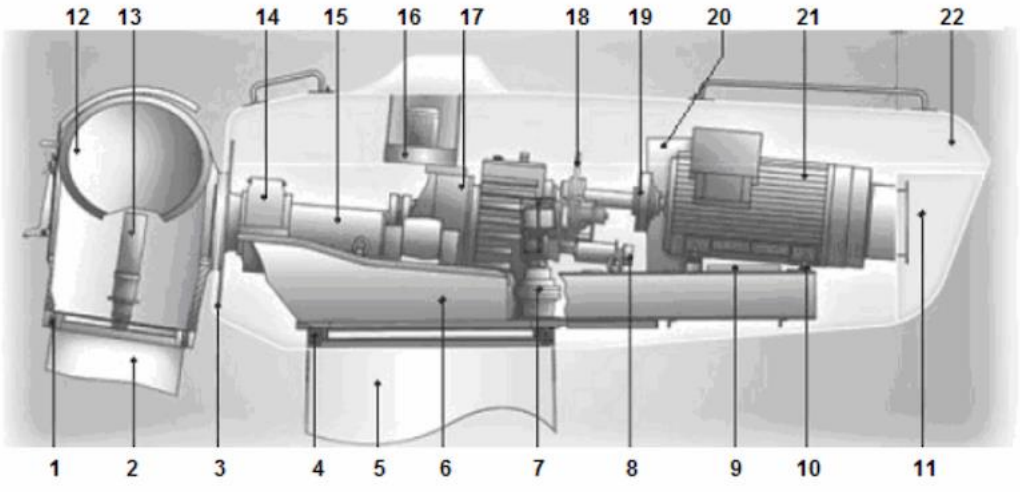

**Şekil 2 :**Türkiye'de yıllara göre rüzgâr enerjisinden sağlanan elektrik enerjisi üretim değişim grafiği (T.C. Enerji ve Tabii Kaynaklar Bakanlığı, 2022).

Şekil 1'de gösterildiği üzere, son on bir yılda kurulu rüzgâr santrallerinin gücünde önemli bir artış kaydedilmiştir. 2021 yılının ilk yarısında 1.280 MW' lık yeni kapasite artışı ile rüzgâr enerjisinin kurulu gücü toplamda 10.607 MW' a ulaşmıştır. Rüzgâr enerjisinin yıllık büyümesi, enerji sistemine etkisini ve genel enerji arzına katkısını artırmaktadır (T.C. Enerji ve Tabii Kaynaklar Bakanlığı, 2022).

### 1.2. RÜZGÂR TÜRBİNİ YAPISI VE ÇALIŞMA PRENSİBİ

Rüzgâr türbini, rüzgârdan gelen enerjiyi kullanarak elektrik enerjisi üreten bir sistemdir. İlk olarak, rüzgâr türbininin kanadı, rüzgârdan gelen kinetik enerjiyi yakalar. Bu enerji, kanadın dönmesine ve rotor miline iletilmesine neden olur. Rotor mili, döndükçe mekanik enerji oluşturur. Oluşan mekanik enerjiyi jeneratör elektrik enerjisine çevirir. Jeneratör, rotor milinin dönme hareketini elektrik enerjisine çevirir. Elde edilen elektrik enerjisi, genellikle bir aküye depolanır veya doğrudan kullanıma yönlendirilir. Bu şekilde, rüzgâr türbini rüzgârdan gelen doğal kaynak olan kinetik enerjiyi elektrik enerjisine çevirir (Elibüyük & Üçgül, 2014).

Rüzgâr türbinleri, temel olarak dikey ve yatay eksenli türbinler olarak iki ana gruba ayrılır. Dikey eksenli türbinler genellikle deneysel amaçlarla kullanılırlar, ticari kullanımları sınırlıdır. Günümüzde yaygın olarak kullanılan sistemler ise yatay eksenli türbin sistemleridir. Aşağıda yatay eksenli bir türbinin iç yapısı gösterilmiştir (Ahlström, 2002).




**Şekil 3 :**Rüzgâr türbini bileşenleri (Ahlström, 2002).

| | |
|---|---|
| 1- Kanat Yatağı | 12- Rotor Göbeği |
| 2- Kanat | 13- Kanat Açısı (pitch) Sürücüsü |
| 3- Rotor Kilidi | 14- Yatak Engelleyicisi |
| 4- Sapma Yatağı | 15- Rotor Şaftı |
| 5- Kule | 16- Yağ Soğutucusu |
| 6- Ana Gövde | 17- Dişli Sistemi |
| 7- Sapma Sürücüsü | 18- Fren |
| 8- Transmitter | 19- Kuplaj Sistemi |
| 9- Akü | 20- Kontrol Panel |
| 10- Ses Yalıtım | 21- Generator |
| 11- Havalandırma | 22- Dış Yüzey |

Şekil 3 'de belirtilen bileşenlerin görevleri; kanatlar, uçak kanatlarındaki kaldırma kuvveti prensibine dayanarak çalışır. Bu sayede kanatlar rüzgâr enerjisini rotora aktarır. Rüzgâr türbinlerinde rotor, rüzgâr enerjisini dairesel harekete dönüştüren ana bileşendir ve kanatlar, aerodinamik kontrol yüzeyleri ve göbekten oluşur. Kanat açısı sürücüsü, rüzgâr hızına göre kanatların açısını ayarlarken, sapma sürücüsü ise rotorun rüzgâr yönüne göre konumunu kontrol eder. Rotor şaftı, rotoru dişli sistemine veya doğrudan tahrikli sistemlere bağlar. Fren, acil durumlarda rotoru durdurur. Dişli sistemi, jeneratörlerin yüksek hızda dönmesini sağlar ve jeneratör, bu mekanik enerjiyi elektrik enerjisine dönüştürür. Bu bileşenler birlikte, rüzgâr türbinlerinin verimli çalışmasını ve rüzgârdan elektrik üretilmesini mümkün kılar (Ahlström, 2002).

### 1.2.1. Rüzgârın Teorik Güç ve Aktif Güç Hesabı

Aktif güç, rüzgâr hızı ve rüzgâr yönüne bağlı olarak gerçekleşen rüzgâr türbininden elde edilen gerçek güç değeridir. Teorik güç, rüzgâr türbininin kurulumda belirtilen mevcut yerdeki rüzgâr hızı ve rüzgâr yönü hesaplanarak bulunmuş olan rüzgâr türbininin üretebileceği güç kapasitesini belirten değerdir. Aktif güç ile teorik güç arasındaki oran rüzgâr türbininin işletme verimini belirtir. Rüzgâr gücü üretilme sırasında oluşan kayıplar ile teorik güç kavramı arasındaki oran da türbin verimini gösterir.

Rüzgârın teorik gücü olarak ifade edilen ($\mathrm{P_o}$) değeri (1) no'lu formülde gösterilmektedir. Aktif gücü ($P_{Aktif}$) değeri ise (2) no'lu eşitlikte gösterilmiştir (Ragheb & Ragheb, 2011).

$$\mathrm{P_o} = \frac{1}{2}\,\rho\,S\,V^3 \tag{1}$$

$$P_{Aktif} = \mathrm{P_o}.C_p \quad (2)$$

(1) no'lu formüldeki, $\rho$ (kg/m³) hava yoğunluğu, $S$ (m²) rüzgârın hareket yönündeki dik kesit alanı ve $V$ (m/s) rüzgâr hızıdır. Rüzgâr türbininin maksimum teorik verimliliği $\mathrm{C_p}$, rüzgârdan elde edilen aktif güç ile rüzgârda mevcut toplam güç arasındaki orandır (Ragheb & Ragheb, 2011).

# İKİNCİ BÖLÜM

# LİTERATÜR TARAMASI

## 2.1. RÜZGÂR GÜCÜ ÜRETİM TAHMİNİNİN TANIMI, ÖNEMİ VE YAPILAN ÇALIŞMALAR

Tahmin yapma, geçmiş verilere ve matematiksel modellere dayalı olarak geleceği öngörmeyi amaçlayan bir süreçtir. İşletme yöneticileri, belirsizlikle başa çıkmak için genellikle talep tahminlerine güvenirler. Bu tahminler, bir şirketin ürün veya hizmetlerine olan gelecekteki talebi belirlemeye çalışır ve bu sayede üretim, kapasite ve planlama süreçlerini optimize eder. Talep odaklı tahminler, finans, pazarlama ve insan kaynakları planlamasında da kritik rol oynar. Tahmin yöntemleri genel olarak iki kategoriye ayrılır: Nicel tahminler, geçmiş verilere ve matematiksel analizlere dayanır; nitel tahminler ise yöneticilerin sezgileri ve deneyimlerine dayalıdır. En iyi sonuçlar genellikle bu iki yöntemin birleştirilmesiyle elde edilir (Heizer vd., 2020).

Bir işletmenin etkili bir şekilde tahmin yapabilmesi için tahmin amacının belirlenmesi, talep planlaması ve tahmininin sisteme entegre edilmesi, talep tahminini etkileyen ana faktörlerin belirlenmesi, uygun düzeyde birleştirme ile tahmin yapılması ve tahmin için performans ve hata ölçümleri belirlenmesi önemlidir. Her tahmin, bu tahmine dayalı olarak alınan kararları destekler, bu yüzden tahmine dayalı kararların net bir şekilde tanımlanması gereklidir. İşletmeler, tahminlerin doğruluğunu ve zamanlamasını değerlendirmek için net performans ölçütleri oluşturmalıdır. Bu ölçütler, bu tahminlere dayalı iş kararlarının hedefleriyle bağlantılı olmalıdır. İşletme gerçek talebi tahmin edilen talep ile karşılaştırarak tahminin doğruluğunu değerlendirmelidir. Daha sonra, gelecekteki tahmin hatalarını azaltma veya gözlemlenen tahmin hatalarına yanıt verme planları yapabilmelidir (Chopra & Meindl, 2007).

Verimlilik, bir girdi başına elde edilen çıktıyı ölçerken, etkinlik, kaynakların ne kadar iyi kullanıldığını değerlendirir ve dijitalleşme ile otomasyon, bu iki faktörü artırmada büyük rol oynar. Şirketler, rekabet avantajı elde etmek için büyük veri analizi ve matematiksel modelleme ile operasyonlarını optimize etmelidir. Bir şirketin başarısı, müşteri taleplerini karşılamak için arz ve talebi etkili bir şekilde dengelemesine bağlıdır. Üretim planlama stratejileri veri toplama, modelleme ve süreçlerin otomasyonunu içermeli; üretim planlayıcıları, verimli operasyon yönetimi için en uygun kararları almalıdır. Üretim planlama ve kontrol, hedef belirleme, malzeme yönetimi, talep tahmini ve işgücü tahsisi gibi operasyonel faaliyetleri içerir ve bu süreçler, arz ve talebi dengelemeyi amaçlar (Tayali, 2023).

Rüzgâr gücü tahminleri, güç sisteminin planlamasını, bakım ve onarımların düzenlenmesini, koruyucu bakımların yapılmasını, elektrik piyasasında daha hassas tekliflerin verilmesini ve rüzgâr enerjisinin sistem içindeki payının artırılmasını sağlar (Karık vd., 2017).

Bu çalışmada rüzgâr türbinindeki SCADA sisteminden elde edilen veri setine dayanarak rüzgâr gücünün tahmin yöntemine geçmeden önce, tahmin tekniklerinin nasıl kullanıldığını gösteren çalışmalar incelenmelidir. Rüzgâr gücü tahmini yöntemlerinin üç kategorisi vardır: fiziksel, istatiksel ve fiziksel ile istatiksel yöntemlerin kullanıldığı hibrit yöntemler.

Fiziksel yöntemlerde, rüzgâr türbinlerini/çiftliklerini modellemek için sıcaklık, basınç vb. sayısal hava tahmini (NWP) verileri kullanılır. Bu değişkenler, rüzgâr hızını belirlemek için zaman alan karmaşık matematiksel modellerde kullanılır. Ardından, rüzgâr gücünü tahmin etmek için tahmin edilen rüzgâr hızı ilgili değerler rüzgâr türbini güç eğrisine (normalde türbin üreticisi tarafından sağlanan) alınır. Bu yöntemde geçmiş zaman serisi verilerine ihtiyaç yoktur sadece fiziksel verilerle tahmin yapılır (Hanifi vd., 2020).

İstatistiksel yöntemler ise meteorolojik koşullar dikkate alınmadan büyük miktarda geçmiş zaman serisi verilerine dayanmaktadır. Genellikle yapay zekâ yani makine öğrenmesi ve zaman serisi analiz yaklaşımlarını içerir (Wang vd., 2011).

Hibrit yöntemlerde fiziksel yöntemler ve istatistiksel yöntemler birleştirilerek tahmin modellemesi yapılır. Çalışmada ele alınan veri setine baktığımızda aktif güç, rüzgâr hızı, rüzgâr yönü, aktif teorik güç ile zaman değerleri mevcuttur. Sayısal hava tahmini (NWP) değerleri ile modelleme yapılamayacağı için yöntem olarak istatiksel yöntem seçilerek rüzgâr gücü tahmini yapılacaktır.

Literatürde istatiksel yöntemler kullanılarak yapılan tahminlerde genellikle zaman serisi analizi, makine öğrenmesi ve derin öğrenme yöntemleri kullanılmaktadır.

Singh ve Rizwan (2022) araştırmalarında Yalova rüzgâr santralindeki SCADA sisteminden alınan veri setinin hatalarından arındırılarak geçmiş veri kümelerini kullanarak rüzgâr gücü üretimi için makine öğrenimine dayalı tahmin modelleri tasarlamış ve geliştirmişlerdir. Rüzgâr hızı ile üretilen güç arasındaki korelasyonları gösteren dağılım eğrileri çizilmiştir. Ayrıca, tahmin edilen ortalama rüzgâr gücü ve bu türbinden gelen fiili ortalama güç, çizilen hata eğrileri kullanılarak karşılaştırılmıştır. Önerilen makine öğrenimi modellerinin Rastgele Orman, XGBoost ve Uzun Kısa Süreli Bellek (“LSTM”) tahmin yeteneklerini değerlendirmek ve incelemek için Ortalama Kare Hatası (“MSE”), Hata Kareleri Ortalamasının Karekökü (“RMSE”), Ortalama Mutlak Yüzde Hatası (“MAPE”), Ortalama Mutlak Hatası

(“MAE”) ve Determinasyon katsayısı (“R²”) gibi hata değerleri makine öğrenimi modellerinin performansını karşılaştırmak için kullanılmıştır (Singh & Rizwan, 2022).

Bir başka çalışmada Rüzgâr türbinleri tarafından üretilen güç tahmini için önerilen derin makine öğrenmesi yöntemlerinden Tekrarlayan Sinir Ağı (“RNN”), LSTM ve Geçitli Tekrarlanan Birimler (“GRU”) algoritmaları karşılaştırılmıştır. Çalışmada incelenen SCADA veri tabanı, Brüksel Belçika’da deniz seviyesinden 109 m yükseklikte bulunan 2.3 MW güç dereceli üç kanatlı rüzgâr türbini kayıtlı veri tabanıdır. Veri seti, rüzgâr türbininde kaydedilen zaman serisi verilerinden tarih/saatini, ortam sıcaklığını, rüzgârın akış yönünü, rüzgârın hızını ve rüzgâr türbininden üretilen rüzgâr gücünü içerir. Önerilen iş akışı, veri seti ön işleme aşaması, eğitim aşaması ve test aşaması olmak üzere üç adımdan oluşur. Veri seti ön işleme aşamasında, zaman serisi şeklinde olan veri seti derin öğrenme modeline uyacak şekilde biçimlendirilir ve hatalı değerler temizlenir. Eğitim aşamasında, veri seti doğrulama seti, test seti ve eğitim seti olarak ayrılır. Daha sonra derin öğrenme modeli geliştirilir, eğitilir ve eğitim ve doğrulama veri seti kullanılarak doğrulanır. Modelin performansı, ortalama kare hata hesaplanarak doğrulanır. Test aşamasında, eğitilen model, test veri seti ile test edilir ve tahmin edilen ve gözlenen değerler ile ilgili grafik çizilir (Chandran vd., 2021).

LSTM algoritması ile yapılan diğer bir çalışmada, SCADA sistemine dayalı rüzgâr türbini veri analizi ve tahmini sunulmaktadır. SCADA sistemi sürekli olarak zaman serisi verileri üretmektedir. Bu modelde rüzgâr ve aktif güç üretimi için kısa vadeli analizleri tahmin edebilir. Görselleştirmeye dayalı veri analizi, veri setinin karmaşıklığını azaltmak ve anlamak için faydalıdır. Sunulan analizden, rüzgâr hızının ve yönünün her ayın birkaç gününün başlangıcında sabitlendiği sonucuna varılabilir. Rüzgârın yönü rüzgârın estiği bölgeleri, hızı ise üretilecek güç miktarını belirler. Geliştirilen LSTM tabanlı tahmin modeli, rüzgâr ve elektrik üretimi hakkında kısa vadeli tahmin sağlar. Rüzgâr enerjisinin dinamik davranışını etkili bir şekilde yakalayabilir. Önerilen model ile mevcut istatistiksel ve makine öğrenimi teknikleri arasındaki performansı değerlendirmek için karşılaştırmalı bir analiz yapılır. Karşılaştırma, istatistiksel hareketli ortalama tekniği (“MA”) ve çok katmanlı bir algılayıcı (“MLP”) modeli dikkate alınarak yapılır. Bu karşılaştırmalı analizde kullanılan metrik, ortalama mutlak hatadır. MAE’ nin düşük değerde olması daha iyidir. Tahmin sonuçlarının MA tekniği ve MLP tekniği ile karşılaştırılması, modelin daha iyi performans gösterdiğini göstermektedir (Delgado & Fahim, 2021).

Singh vd. (2021), araştırmalarında, kısa vadeli rüzgâr gücü üretim tahmininin doğruluğunu artırmak amacıyla beş optimize edilmiş regresyon makine öğrenmesi yöntemini karşılaştırmışlardır. Bu yöntemler; Rastgele Orman, Gradyan Artırma Makinesi (“GBM”), K En

Yakın Komşu (“KNN”), Karar Ağacı ve Ekstra Ağaç Regresyonudur (“Extra Tree Regression”). Bu yöntemler, Türkiye’nin batısında yer alan rüzgâr çiftliklerindeki rüzgâr hızı ve rüzgâr yönü ile zaman verileri temel alınarak uygulanmıştır. Girdiler olan rüzgâr hızı ve yönünün rüzgâr enerjisi üretimi üzerindeki etkileri incelenmiş ve polar diyagramlar çizilmiştir. Rüzgâr hızı ile üretilen rüzgâr gücü arasındaki ilişkileri gösteren dağılım eğrileri tüm yöntemler için çizilmiş ve tahmin edilen ortalama rüzgâr gücü, çizilen hata eğrileri yardımıyla türbinden elde edilen gerçek ortalama güç ile karşılaştırılmıştır. Sonuç olarak, GBM regresyonunu içeren algoritmanın üstün tahmin performansı gösterdiğini ortaya koymuşlardır (Singh vd., 2021).

Lin vd. (2021), yaptıkları çalışmalarda iklimdeki değişikliklerin farklı hava koşulları altında rüzgâr gücü tahminini etkilediğini ve teorik olarak hem rüzgâr gücü üretiminin hem de hava değişikliklerinin istatistik tabanlı tahmin modelleri kullanılarak doğru bir şekilde tahmin edilmesinin zor olduğu belirtmişlerdir. Geleneksel makine öğrenmesi modelleri, %10 ile %17 arasında bir MAPE ile uzun vadeli rüzgâr gücü tahmini yapabilir. Derin öğrenme ağları meteorolojik özellikler ile güç üretimi arasındaki korelasyonları, tahmin hatalarını en aza indirmek için gradyan iniş algoritmaları ile çok katmanlı sinirsel evrişimli bir mimari kullanarak elde etmek amacıyla kullanılmaktadır. Bu, rüzgâr gücü tahmini alanında geniş bir uygulama alanına sahiptir. Bu nedenle, bu çalışmada, derin öğrenme ağlarının Zamansal Evrişim Ağı (“TCN”) algoritmasını kullanarak %10’dan daha az bir MAPE ile uzun vadeli (24-72 saat öncesinden) rüzgâr gücü tahminini amaçlamışlardır. Çalışmada, Türkiye’de yer alan bir rüzgâr santralinden alınan SCADA zaman, hava durumu verileri ve güç üretimi çıktıları kullanılarak TCN modeli kurulmuştur. Deneysel sonuçlar, 72 saatlik rüzgâr gücü tahmini için %5,13’lük bir MAPE göstermiştir. Son olarak, güç tahmini için dört Derin Öğrenme Ağları tabanlı tahmin modelinin performansını karşılaştırdık; bunlar TCN, LSTM, RNN ve kapılı yineleme birimi GRU modelleridir. TCN’ nin, veri girişi büyüklüğü, hata azaltma kararlılığı ve tahmin doğruluğu açısından diğer üç modelden daha iyi performans gösterdiğini doğrulamışlardır (Lin vd., 2021).

Fahim vd. (2022), tahmin modellemesinde, derin öğrenme yaklaşımı olan TCN ve KNN regresyonunu ele almışlardır. Tahmin modellemesinin iki bileşeni vardır. İlk olarak, rüzgâr hızını tahmin etmek için tek değişkenli zaman serisi verilerini işler. İkinci olarak, üretilen gücü bir haftadan bir aya kadar değişen sürelerde (orta vadeli tahmin) tahmin eder. Deneyleri karada bulunan rüzgâr türbinlerine ait halka açık veri setleri üzerinde gerçekleştirmişlerdir. Elde edilen sonuçlar, önerilen modelin uygulanabilirliğini doğrulamaktadır. Mevcut klasik tahmin modelleri olan Karar Ağacı, Rastgele Orman ve Destek Vektör Regresyonu ile yapılan karşılaştırmalı analiz, tasarlanan yaklaşımın daha iyi sonuçlar elde ettiğini göstermektedir. Model, yönetim

ekibinin rüzgâr çiftliklerini uzaktan izlemesine ve güç üretimini önceden tahmin etmesine yardımcı olabilir (Fahim vd., 2022).

Peng vd. (2023), Doğu-Çin veri seti, Yalova (Türkiye) veri seti ve 16 MW veri seti üzerinde çapraz dikkat ("XA") mekanizması kullanarak, kendi alanlarında en iyi uygulama modellerinden iki makine öğrenimi modelini birleştirerek, yenilikçi bir ultra kısa vadeli rüzgâr gücü üretim tahmin modelinin etkinliklerini göstermek için çalışma yapmışlardır. Bu modeller, sınıflandırma problemlerinde etkili olduğu bilinen derin Evrişimsel sinir ağı ("CNN") modeli ve zaman serisi tahmin problemlerinde etkili olduğu bilinen iki yönlü uzun kısa süreli bellek ("Bi-LSTM") modelidir. İki alternatif çapraz dikkat modeli vardır. Birinci modelde CNN özelliklerini anahtar-değer çifti olarak ve Bi-LSTM özelliklerini sorgu olarak kullanan model, XA modeli olarak adlandırılmıştır ve ikincisi Bi-LSTM özelliklerini anahtar-değer çifti olarak ve CNN özelliklerini sorgu olarak kullanan model, XA_I modeli olarak adlandırılmıştır. Bu nedenle, XA modeli Doğu-Çin veri setinde XA_I modelinden daha iyi sonuçlar elde ederken, XA_I modeli Yalova veri setinde ve 16 MW veri setinde XA modelinden daha iyi sonuçlar elde etmiştir. Yalova veri seti ve 16 MW veri seti üzerinde diğer son teknoloji modellerle karşılaştırılabilir sonuçlar mevcut olduğunda, XA_I modelimizin sonuçlarının üstün olduğunu bulmuşlardır (Peng vd., 2023).

Zaman ufuklarına göre literatürde karşılaşılan çalışmalar göz önüne alındığında Çok kısa vadeli tahminler, 30 dakikaya kadar olan tahminlerdir. Bu tahminler rüzgâr çiftliği kontrolünde ve işletilmesinde kullanılır. Kısa vadeli tahminler, saatlerden birkaç güne kadar olan tahminleri hesaba katar ve rüzgâr enerjisi santralleri ve elektrik ticareti için çok önemlidir. Rüzgâr gücü üretiminin dinamikleri önemli hale geldiğinden ve potansiyel olarak modelin daha iyi bir performans gösterebilmesinden dolayı meteorolojik veriler gibi fiziksel girdilerin ilave edilmesiyle kısa vadeli tahminler için istatistiksel bir modelin kurulmasına imkân verebilir. Daha uzun vadeli tahminler genellikle fiziksel yöntemlerden yararlanır ve birim taahhüdü veya bakım planlaması hakkında kararlar almak için kullanılır (González vd., 2021).

| Zaman Ölçeği | Zaman Ufku | Uygulamalar |
|---|---|---|
| Çok Kısa Dönem | 1 saate kadar | -Elektrik piyasası işlem bedeli<br>-Güç sistemi yük izleme<br>-Sistem frekansı yönetimi<br>-Şebeke yönetimi<br>-Türbin yönetimi |
| Kısa Dönem | 1-6 saat arası | -Elektrik dağıtım planı<br>-Talepteki artış veya azalış kararları |

| | | |
|---|---|---|
| Orta Dönem | 6-72 saat arası | -Enerji yönetim kararları<br>-Tüketim alanında operasyon güvenliği<br>-Şebeke yönetimi<br>-Elektrik enerjisi sektöründe fiyatlandırma Ve Enerji ticareti işlemleri<br>-Çevrim içi ve çevrim dışı seçimler<br>-Doğru yük yönetim kararları<br>-Bakım programlaması<br>-En uygun işletme maliyetleri<br>-Operasyon yönetimi |
| Uzun Dönem | 72 saatten 1 yıla kadar | -Tüketim alanında operasyon güvenliği<br>-Yedek güç kararı<br>-Elektrik ticaret işlemleri<br>-Üretici çevrim içi ve çevrim dışı seçimler<br>-Doğru yük yönetim kararları<br>-Bakım programlaması<br>-Rüzgâr çiftliği projesinin fizibilite analizi<br>-Şebeke bakım çalışmaları<br>-Bakım programı oluşturma<br>-Operasyon yönetimi<br>-En uygun işletme maliyetleri<br>-Güç sistemi planlama<br>-Yedek güç gereksinimleri<br>-Birim yükümlülük kararları<br>-Rüzgâr enerjisi stratejisi<br>-Rüzgâr çiftliği projesinin fizibilite analizi |

**Tablo 1 :** Zaman ölçeğine göre sınıflandırılmış rüzgâr gücü tahmininin uygulama alanları (Roungkvist & Enevoldsen, 2020).

# ÜÇÜNCÜ BÖLÜM

# DOĞRUSAL REGRESYON VE MAKİNE ÖĞRENMESİ YÖNTEMLERİ

## 3.1. DOĞRUSAL REGRESYON TANIMI

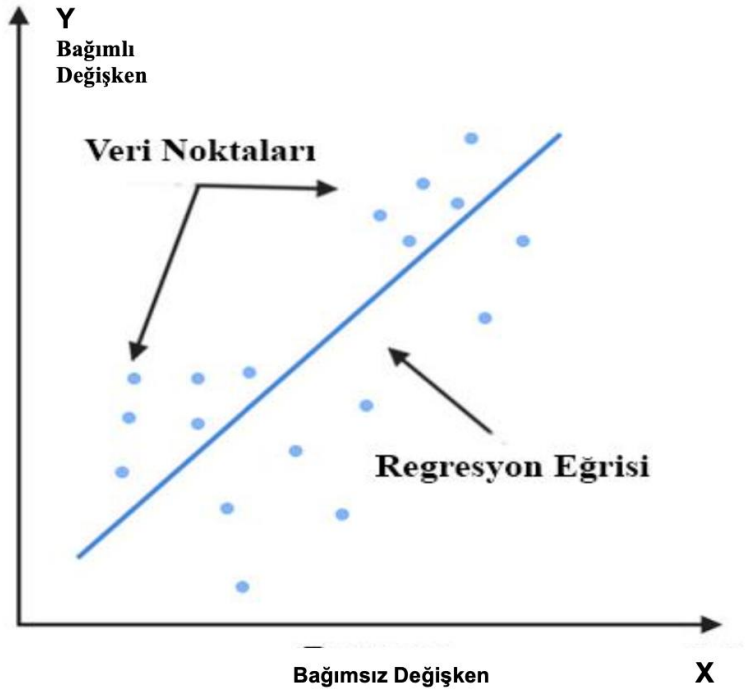
Regresyon analizi, değişkenler arası ilişkiyi modellemek için kullanılan bir istatistiksel teknik olup değişkenlerin geçmiş değerlerini kullanarak oluşturulan modellerde, ilgili değerleri tahmin etmek için girdi olarak nitel veya nicel veriler kullanılır. Regresyon modelleri değişkenlerin yapısına, değişkenler arası ilişkinin doğrusallığına ve diğer çeşitli yaklaşımlara göre sınıflandırılabilir. Bu çalışmada, örneğin, doğrusal regresyon modelleme kullanılmış olup bu modelleme tekniği literatürde basit ve çoklu doğrusal regresyon olarak incelenebilmektedir: İncelenen veride bağımlı değişkeni (tahminlenen) tahminleyen tek bir bağımsız değişken olması halinde basit regresyon, birden fazla bağımsız değişkenin olması durumunda ise çoklu regresyon yöntemi kullanılır (Chatterjee & Hadi, 2012).

Regresyon modelleri, iki ana kategoriye ayrılır: doğrusal ve doğrusal olmayan modeller. Doğrusal regresyon modelleri ise bağımsız değişken sayısına göre tek bağımsız değişkenin olduğu basit doğrusal regresyon ve birden fazla bağımsız değişkenin olduğu çoklu doğrusal regresyon olmak üzere ikiye ayrılabilir. Regresyon analizinde temel amaç özniteliklerden birinin diğerine olan bağımlılık düzeyini matematiksel olarak ortaya koyabilmektir. $\hat{y}_i$ bağımlı değişken iken $x_i$ bağımsız değişken veya açıklayıcı değişken olarak adlandırılır. Ayrıca bu süreçte, üzerinde çalışılan değişkenlere ilişkin veri seti göz önünde bulundurularak ortaya konulan ilişkinin matematiksel biçimi yani regresyon eğrisi elde edilir. Basit doğrusal regresyon formülü şu şekildedir (James vd., 2023).

$$\hat{y}_i = f(x_i, \beta) + e_i \tag{3}$$

(3) no'lu eşitlikte $\hat{y}_i$= bağımlı değişken, $f$ = fonksiyon, $x_i$= bağımsız değişken, $\beta$ = bilinmeyen parametreler, $e_i$= hata terimini göstermektedir (James vd., 2023).

Regresyon analizi, bağımlı değişkenin değerini en doğru biçimde tahmin etmek için bir veya birden çok bağımsız değişkeni içeren bir doğrusal denklemin katsayılarını belirlemeyi amaçlar. Bu süreçte, tahmin edilen ve gerçek sonuçlar arasındaki farkları en aza indiren bir düz çizgi veya düzlem oluşturulması hedeflenir (Chatterjee & Hadi, 2012).




**Şekil 4 :**Doğrusal regresyon grafiği (Chatterjee & Hadi, 2012).

Şekil 4' te gösterildiği gibi regresyon doğrusu, veri noktalarından bu doğruya olan dikey mesafeyi olabildiğince küçük tutan bir doğrudur. Bu, en uygun doğrunun, hata kareleri toplamını en aza indiren bir doğru olması gerektiği anlamına gelir (Chatterjee & Hadi, 2012).

### 3.1.1. Çoklu Doğrusal Regresyon

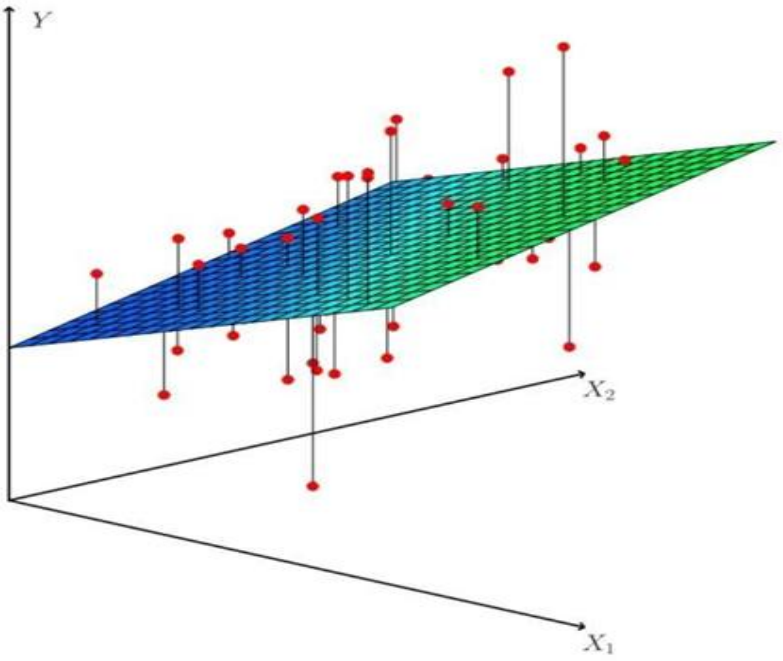
Birden fazla bağımsız değişkenin yer aldığı ve bir tahmin gerçekleştirmeye yönelik gerçekleştirilen regresyon sürecine ise çoklu regresyon denir (James vd., 2023).

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots \ldots + b_n x_n \quad (4)$$

(4) no'lu formülde $\hat{y}$= bağımlı değişken ve $x_n$ = bağımsız değişkenlerdir. Hata terimine formülde yer verilmemiştir (Chatterjee & Hadi, 2012).

Çoklu regresyon, çoklu açıklayıcı değişkenlere sahip doğrusal ve doğrusal olmayan regresyonları kapsayan daha geniş bir regresyon sınıfıdır. Doğrusal regresyonda ilişkinin eğimini etkileyen yalnızca bir bağımsız değişken varken, çoklu regresyonda birden fazla bağımsız değişken bulunmaktadır. Farklılıklarına rağmen hem basit hem de çoklu doğrusal regresyon modelleri doğrusal modellerdir. Doğrusal olmayan regresyon, iki veya daha fazla değişkenden elde edilen çıktıyı grafiksel olarak çizmeye çalışmasından dolayı doğrusal regresyona oldukça benzer. Doğrusal olmayan modellerin geliştirilmesi doğrusal modellerden daha karmaşıktır çünkü fonksiyon, deneme-yanılmadan kaynaklanabilecek tekrarlar

(iterasyon) yoluyla oluşturulur. Modelin amacı, doğrusal regresyonda olduğu gibi hata kareleri toplamını mümkün olduğunca küçük yapmaktır. Kareler toplamı, y gerçek değerlerinin $\hat{y}'$yi tahmin etmek için kullanılan doğrusal olmayan fonksiyondan ne kadar farklı olduğunu izleyen bir ölçüdür. Bu süreçte veri setindeki her $\hat{y}$ veri noktası arasındaki fark bulunarak hesaplanır. Ardından, bu farkların her birinin karesi alınır. Son olarak, tüm kareli rakamlar toplanır. Bu kareli rakamların toplamı ne kadar küçükse, fonksiyon kümedeki veri noktalarına o kadar iyi uyar (Chatterjee & Hadi, 2012).




**Şekil 5 :**Çoklu regresyon canlandırması (James vd., 2023).

### 3.2. MAKİNE ÖĞRENMESİ

Makine öğrenmesi, ele alınan veri kümesini iyi temsil eden bir model oluşturarak çeşitli problemleri kolayca çözmeyi amaçlar. Bu alan, bilgisayarların insan beynini taklit etmekten ileri bir noktaya gelmiş ve istatistik alanını, öğrenme süreçlerine ilişkin temel istatistiksel hesaplamalı teoriler üreten geniş bir disipline dönüştürmüştür. Makine öğrenmesi, tamamen bilgisayarın öğrenmesine olanak tanıyan algoritmalar geliştirmekle ilgilidir. Öğrenme, istatistiksel düzenlilikleri veya diğer veri kalıplarını bulma sürecidir. Bu algoritmalar, insan yaklaşımını taklit ederek belirli görevlerin öğrenilmesini sağlar (Nasteski, 2017).

Makine öğrenimi algoritmaları da denetimli (“supervised”), denetimsiz (“unsupervised ”), destekli veya takviyeli (“reinforcement”) vbg. çeşitli biçimlerde sınıflandırılabilir. Denetimli Öğrenme, bir dizi girdi değişkenine dayanarak çıktı değerlerini tahmin etmeye odaklanır ve regresyon ile sınıflandırma alt gruplarını içerir. Denetimsiz Öğrenme ise etiketli örnekler olmadan girdi değişkenleri arasındaki ilişkileri ve örüntüleri tanımlamaya çalışır; kümeleme algoritmaları bu tür öğrenmeye örnektir. Takviyeli Öğrenme ise sistemin girdilere karşılık çıktıları üretmesini ve ürettiği çıktının doğruluğunu değerlendiren sinyallerle öğrenme sürecini sürdürmesini içerir. Bu kategorilendirme, makine öğrenimi alanında kullanılan temel yaklaşımları anlamak için önemlidir. Her kategori, farklı problemlerin çözümü için farklı teknikler ve algoritmalar gerektirebilir (Nasteski, 2017).

### 3.2.1. Karar Ağaçları

Karar ağaçları algoritması veri kümelerinin sınıflandırılması ve regresyon işlemi için kullanılan topluluk temelli bir yöntemdir. Karar ağacı kök düğüm, dallar ve yaprak düğümlerden oluşur. Bir veri setindeki özelliğin test edilmesi her dahili düğümde yapılır, testin sonucu dal ve sınıf etiketinde, sonuç olarak da yaprak düğümde olur. Kök düğüm ağacın en üst düğümüdür (Patel & Prajapati, 2018).

Bu algoritmada amaç tahmin ediciler ve nihai karar/sonuç arasındaki ilişkiyi modellemektir. Ağaç yapısının kök düğümünden başlanarak niteliklerine göre verileri bölen eğer/değilse cevaplarına göre veriler karar düğümleri boyunca ilerler. Dallar olası seçimleri, yaprak düğümler ise nihai kararları gösterir. Algoritmada homojen bölümleri oluşturan bölmeleri belirlemek için böl-yönet taktiği uygulanır ve veri kümesini benzer değerlere sahip bölümlere ayırmaya çalışır. Bu bölümlerin her biri ağaç yapısında bir eğer/değilse kararıyla sonuçlanır. Zor ve karmaşık bir karar verme işlemi bir dizi küçük karara bölünerek daha basit hale getirilir (Lantz, 2013).

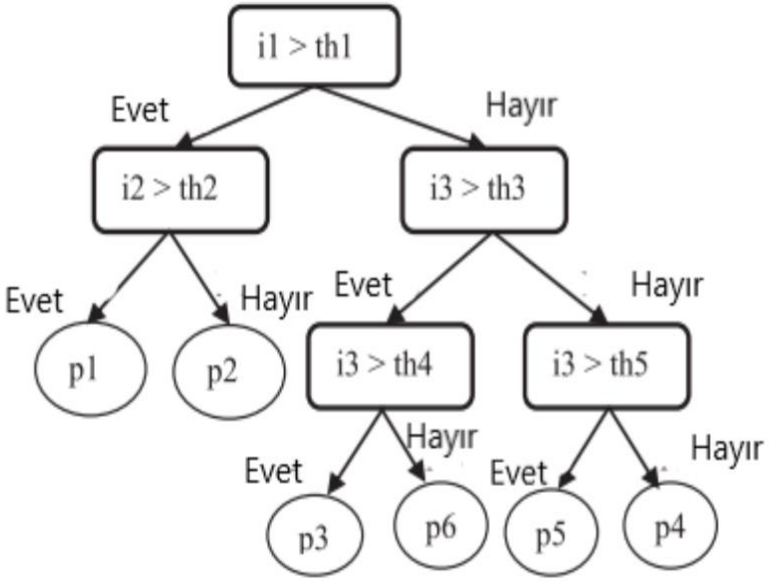
Ağaç büyütme işlemi ile daha küçük ve daha homojen bölümler oluşturulur. Karar ağacı her zaman eksen-paralel bölünmeler yaratır. Bazı durumlarda bu bölünme şekli ile çok hızlı bir şekilde çok karmaşık hale gelme eğilimi vardır. Bir ağaç yapısı aşırı büyüdüğünde ve aşırı derecede karmaşık hale geldiğinde aşırı uyum (“overfitting”) sorunu ortaya çıkabilir. Karar ağaçları eğitildikleri veri kümesine fazla uyma eğilimine sahiptir. Bu dezavantajları aşmak için budama işlemi uygulanır. Ön budama işlemi ile ağacın büyüme süreci erken durdurulur. Ağaç önceden tanımlı bir boyuta ulaştığında böl-yönet işlemini durdurur. Hata oranı, budanma yeri ve karmaşıklık arasında tanımlanan ilişkiye göre ağaç yapısı giderek karmaşık hale geldikçe modelin hata oranı düşer. Modelin beklenen performansta sonuçlar vermesi için ağaçlar çok büyük ve çok küçük olmamalıdır (Lantz, 2013).

Uygulanan karar ağacı modellerine örnek verecek olursak sınıflandırma sisteminin yasal nedenlerden dolayı şeffaf olması gereken veya karar alma sürecini kolaylaştırmak için sonuçların paylaşılması gereken uygulamalar için özellikle uygundur. Bazı potansiyel kullanım yerleri şunlardır: (Lantz, 2013).

-Başvurunun reddedilmesine neden olan kriterlerin iyi belirlendiği kredi puanlama modellerinin doğru şekilde belirlenmesi gerekmektedir.

-Yönetim veya reklam ajanslarıyla paylaşılacak müşteri kaybı veya müşteri memnuniyetine ilişkin pazarlama çalışmaları için geçerli olan verilerin paylaşılması önemlidir.

-Tıbbi durumların laboratuvar ölçümlerine dayalı olarak teşhis edilmesi ve semptomlar ile hastalığın ilerleme hızı arasındaki ilişkinin değerlendirilmesi önemlidir (Lantz, 2013).




**Şekil 6 :** Karar ağacı regresyon örneği (Swetapadma, 2016).

Şekil 6' daki $I$ = {$i1$, $i2$, $i3$} girdi parametreleri, $Th$ = {$th1$, $th2$ , $th3$ , $th4$ , $th5$ } eşik değerleri ve $P$ = {$p1$, $p2$, $p3$, $p4$, $p5$, $p6$ } tahmin değerleridir. Tahminler, girdi parametreleri için belirlenen "eğer - ise" koşullarına dayanarak yapılır (Swetapadma, 2016).

Karar ağaçlarında regresyon işleminin amacı, sürekli bir hedef değişkeni tahmin etmektir. Yapraklardaki veya bölünme kısımlarındaki hata değerleri model için önemlidir. Burada kullanılan hata fonksiyonu, hata kareleri toplamı ("SSE - Sum of Squared Errors")'dir (Williams, 2011).

$$SSE = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \quad (5)$$

(5) no'lu formülde; $y_i$: Gerçek değer, $\hat{y}_i$: Model tarafından tahmin edilen değer ve $n$: Toplam örnek sayısını ifade etmektedir (Williams, 2011).

Bu hata fonksiyonu, her yapraktaki $y$ gerçek değerleriyle $\hat{y}$ modelin tahmin ettiği değerler arasındaki farkların karesini toplar. Amaç, bu farkları en aza indiren bölünme noktalarını seçmektir. Karar ağaçlarında regresyon işleminin amacı, modelin mümkün olduğunca doğru tahminler yapabilmesi için her yapraktaki hata miktarını azaltmaya çalışmasıdır. Bölünme kararı verilirken, SSE minimize edilmeye çalışılır. Bu, bölmelerin doğruluğunu artırmak için önemli bir adımdır. Her bölünmedeki toplam SSE (6) no'lu formül ile hesaplanır. Sol ve sağ kısımlardaki toplam kare hata değeridir (Williams, 2011).

$$Toplam\ SSE = SSE_{sol} + SSE_{sağ} \quad (6)$$

Bir yaprak düğümüne ulaşıldığında, bu düğümdeki tahmin, tüm örneklerin hedef değişkenlerinin ortalamasıdır (Williams, 2011).

$$\hat{y}_{yaprak} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (7)$$

(7) no'lu eşitlikte; $\hat{y}_{yaprak}$, yaprak düğümdeki tahmini $n$ ise bu yapraktaki örneklerin sayısını ifade eder (Williams, 2011).

Karar ağaçlarında, algoritmayı durdurma prosedürü, her yaprak düğümüne atanan minimum eğitim verisi sayısını kullanmaktır. Sayı, belirli bir minimum değerden az ise, bölünme kabul edilmez ve düğüm nihai bir yaprak düğüm olarak alınır. Eğitim verilerinin sayısı, veri kümesine göre ayarlanır (Williams, 2011).

### 3.2.2. Torbalama

Önyükleme ("Bootstrap") ve toplama ("Aggregation") işlemlerinin baş harflerinin bir kısmından oluşan Torbalama işleminin amacı, eğitim setlerinden seçilen yer değiştirme ile rastgele örnekleme yapılması ve her bir örneklemenin üzerinde ayrı bir model eğitilmesidir. Sonunda, her bir modelin tahminleri bir araya getirilir ve ortalama veya oylama yöntemleri ile bir tahmin elde edilir (Breiman, 1996).

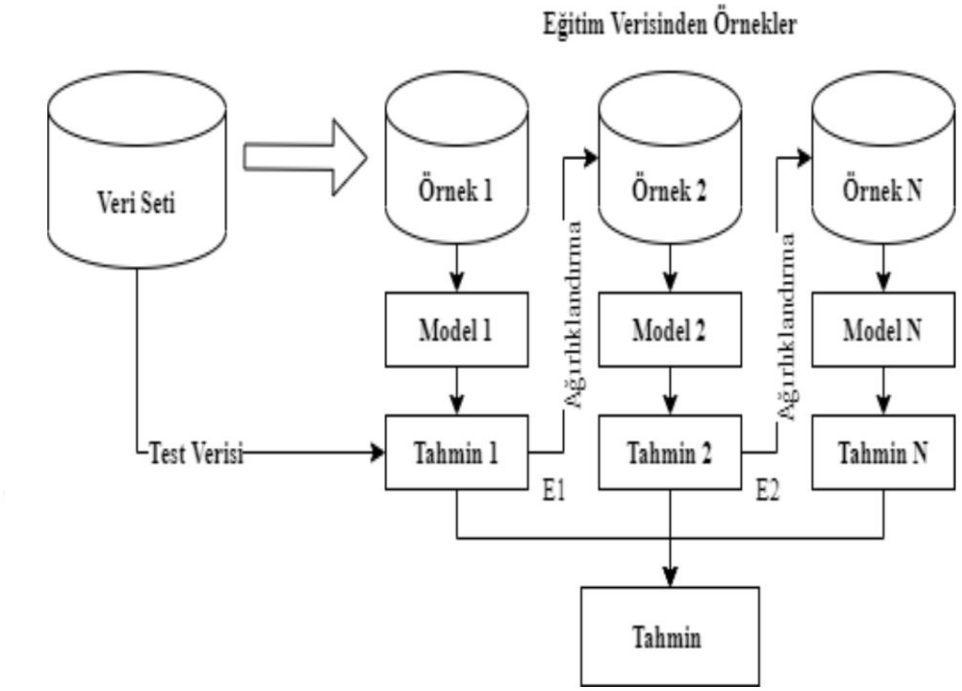
Torbalama işlemi, birden çok modelin bir araya getirilmesiyle varyansı azaltır, böylece genel modelin daha istikrarlı olmasını sağlar. Her bir örnek üzerinde farklı modellerin eğitilmesi ve bu modellerin birleştirilmesi, aşırı uyumu azaltır (Breiman, 1996).

### 3.2.3. Arttırma

Arttırma, modeldeki zayıf tahmincilerin performansını güçlü tahmin performansı olan tahmincilerin seviyesine yükseltmeyi amaçlayan topluluk öğrenimi yöntemidir. Her model bir önceki tahmincinin bilgisinden yararlanmaktadır. Bu sayede her model bir öncekinin geliştirilmiş halidir. Yinelemeli öğrenme sayesinde torbalanmış ağaçlardan daha iyi performans vermektedir. Artırmanın esas olarak üç türü vardır; Adaboost, Gradyan Artırma ("Gradient Boosting") ve XGBoost (Demir & Sahin, 2023).

Adaboost, ilk ünlü artırma algoritmasıdır. Adaboost yönteminde oluşturulan örneklerin ağırlığını değiştirerek algoritmanın yanlış sınıflandırılmış örneklere daha fazla dikkat etmesini sağlar. Daha sonra veriler üzerinde bir karar ağacı uydurur ve yanlış sınıflandırılmış örneklerin ağırlığını toplar. Daha sonra yanlış sınıflandırılmış örneklerin toplam ağırlığına göre bu tahmincinin katkısını hesaplar. Son olarak tüm örneklere katkı yaparak ağırlığı değiştirir, ilk modelin geri kalanını (artıklarını) başka bir karar ağacına sığdırır ve ağırlığı bir önceki adımdaki gibi değiştirir (Demir & Sahin, 2023).

Gradyan artırma, zayıf öğrenicilerin bir dizi halinde bir araya gelerek güçlü bir öğrenici oluşturduğu bir tekniktir. Temel fikir, her yeni modelin önceki modellerin yaptığı hataları azaltmaya odaklanmasıdır. Başlangıçta küçük bir karar ağacı gibi bir model kullanılarak başlanır ve ardışık modeller, önceki modellerin hatalarını düzeltmek için eğitilir. Her bir model, önceki modellerin hatalarını düzeltmeye odaklanarak yeni bir hedef belirler. Sonunda, tüm modellerin tahminleri birleştirilir ve toplu bir tahmin elde edilir. Adaboost yöntemine göre güçlendirilmiş ve daha verimli bir algoritmadır (Demir & Sahin, 2023).




**Şekil 7 :** Arttırma yönteminin aşamaları (Jankowski vd., 2011).

Şekil 7 incelendiğinde, arttırma yöntemi, sınıflandırma sürecinin başında veri setini örneklere bölerek Model 1' i oluşturur. Sonrasında doğruluğun az olduğu veya hiç elde edilemediği örneklerin ağırlıklarını arttırarak Model 2' yi oluşturmaktadır. Amaç oluşturulan modellerin tahminlerde başarısız olduğu örneklerin ağırlıklarının arttırılması biçimiyle, odaklanarak başka bir model oluşturmaktadır. Böylelikle en iyi sınıflandırma performansını elde edene kadar süreç devam etmektedir (Jankowski vd., 2011).

Arttırma süreci matematiksel olarak ifade edilirse; öncelikle veri setinden alınan her bir örneğe, $D_i1 = 1/n$, eşit ağırlıklar verilir. En yüksek doğruluk değeri elde edilene kadar yapılan her bir denemede dağılım (8) no'lu formül ile hesaplanmaktadır (Jankowski vd., 2011).

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} * \ e^{\pm \alpha t} \tag{8}$$

$$(e\breve{g}er\ i.\ birim\ do\breve{g}ru\ sınıflandırıldıysa\ e^{-at}$$
$$e\breve{g}er\ i.\ birim\ yanlı\breve{s}\ sınıflandırıldıysa\ e^{at})$$

(8) no'lu formülde; $t$ = (1,2 … $T$) olacak şekilde $Z_t$ değeri normalizasyon faktörünü ifade etmektedir. Başlangıçta $[\sum_{i=1}^{m} D_{t+1} = 1]$ olacak şekilde seçilmektedir. Kurulacak her yeni

modelde, sınıflandırıcıların ağırlıkları ($\alpha t$) ise (9) no'lu formül ile hesaplanmaktadır (Jankowski vd., 2011).

$$a_t = \frac{1}{2}\log\left(\frac{1-\varepsilon t}{\varepsilon t}\right) \quad (9)$$

$$H(x) = sign\left(\sum_{t=1}^{T} a_t h_t(x)\right) \quad (10)$$

Arttırma yöntemi sonrası her bir zayıf öğrenici için elde edilecek tahmin değeri (10) no'lu formül ile hesaplanmaktadır. Buradaki $h_t$ değeri sınıflandırıcıyı temsil etmektedir (Jankowski vd., 2011).

#### 3.2.3.1. XGBoost

XGBoost, bir dizi zayıf öğrenicinin tüm tahminlerini birleştirerek daha iyi tahmin performansı elde eden güçlü bir öğrenici oluşturur. Bu yöntem, modelin tahminleri için birincil hiperparametreleri belirlemek için gereklidir. Her makine öğrenimi algoritması, en iyi performansı elde etmek için en iyi hiperparametrelerle çalışır. XGBoost için bu hiperparametreler öğrenme oranı, subsample (eğitim verilerinin alt örnekleme oranı), max_depth (bir ağacın maksimum derinliği), gamma (minimum kayıp azalması), col_sample_bytree (her ağaç oluşturulurken sütunların alt örnekleme oranı), min_child_weight (minimum örnek ağırlığı toplamı) ve nrounds (boosting iterasyon sayısı) gibi hiperparametreleridir. XGBoost işlemlerini gerçekleştirmek için R'deki 'caret' kütüphanesinden 'xgboost' paketi kullanılır (Demir & Sahin, 2023).

Ağaç temelli çalışan Aşırı Gradyan Arttırma (XGBoost) yöntemi, örnek boyutu $n$ ve değişken boyutu $m$ olan bir veri setinde, $D = \{(x_i, y_i)\}(|D| = n, x_i \in R^m, y_i \in R)$, tahminlerin yapılması için (11) no'lu fonksiyonu kullanır (Chen & Guestrin, 2016).

$$\hat{Y}_i = \emptyset(x_i) = \sum_{k=1}^{K} f_k(y_i), f_k \in F \quad (11)$$

(11) no'lu fonksiyonda $F$ ağaç temelli algoritmayı ifade ederken $f_k$ zayıf sınıflandırıcıları, $K$ ise zayıf sınıflandırıcıların sayısını ifade etmektedir. Yöntem, öğrenme aşamasında (12) no'lu amaç fonksiyonunu minimize etmeyi amaçlamaktadır (Chen & Guestrin, 2016).

$$L(\emptyset) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (12)$$

(12) no'lu amaç fonksiyonunda $l$ türevlenebilen kayıp fonksiyonunu, $\Omega$ ise aşırı öğrenmeyi engelleyen düzenlileştirme terimini ifade etmektedir. Kurulacak olan model bir arttırma yöntemini içerdiğinden sınıflandırıcılar birbirlerine paralel ve bağımlı olarak çalışacaktır. Bu sebeple, $\hat{y}_i^{(t)}$, $t$. denemedeki $i$. örneğin tahmini olacak şekilde, $f_t$'nin amaç fonksiyonuna eklenmesi gerekmektedir. Böylelikle amaç fonksiyonu (13) no'lu eşitlik gibi olacaktır (Chen & Guestrin, 2016).

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (13)$$

(13) no'lu amaç fonksiyonun ise hızlı bir şekilde optimize edilmesi için Aşırı Gradyan Arttırma yöntemi, ikinci dereceden yaklaşımı kullanmaktadır. Buna göre amaç fonksiyonu düzenlenirse (14) no'lu eşitlik gibi olacaktır (Chen & Guestrin, 2016).

$$L^{(t)} \cong \sum_{i=1}^{n} [l\left(y_i, \hat{y}_i^{(t-1)} + g_i f_t(x_i) + \frac{1}{2} h_i {f_t}^2(x_i)\right) + \Omega(f_t) \quad (14)$$

(14) no'lu eşitlikte ise $g_i$ve $h_i$ değerleri, gradyan istatistiklerini temsil etmektedir. Daha basit bir amaç fonksiyonu elde etmek amacıyla $t$. adımda sabit terim eşitlikten çıkartılarak (15) no'lu eşitlik elde edilir (Chen & Guestrin, 2016).

$$L^{(t)} = \sum_{i=1}^{n} \left( g_i f_t(x_i) + \frac{1}{2} h_i {f_t}^2(x_i)\right) + \Omega(f_t) \quad (15)$$

Daha sonrasında istenirse (15) no'lu eşitlikteki $\Omega$ değeri eşitlikte içeri dağıtılabilir. Yapılan her denemede bir önceki modelin tahmin etmede başarısız olduğu örneklerin ağırlıkları arttırılarak yeni modeller oluşturulur. En iyi sınıflandırma performansını elde edilene kadar güncel model ile bir önceki model arasındaki farklar karşılaştırılır (Chen & Guestrin, 2016).

### 3.2.4. Rastgele Orman

Rastgele Orman algoritması, karar ağaçları temelli torbalanmış ağaçlar ile regresyon ve sınıflandırma işlemlerinde çözüm sağlayan bir topluluk öğrenme yöntemidir. Bu yöntemin avantajları arasında tek bir karar ağacından daha doğru sonuçlar vermesi, eksik verilerle daha iyi başa çıkması ve aşırı uyum sorununu azaltması bulunmaktadır. Rastgele Orman modelinde, düğüm seçimi süreci mevcut özellik kümesinden rastgele bir alt kümenin seçilmesi ve ardından bu alt kümeden optimize edilmiş bir özellik seçilmesiyle gerçekleştirilir. Ayrıca, artırma işleminde her yeni eğitim seti, başta oluşturulan eğitim setinden örneklenerek önyükleme işlemine dayalı olarak değiştirilir. Karar ağaçlarını oluştururken her düğümde rastgele seçilen özelliklerin kullanılması, ağaçlar arasındaki korelasyonu azaltır (Alkesaiberi vd., 2022).

Topluluk yöntemlerinden olan Rastgele Orman algoritması modelde zayıf öğrenenlerin ekip çalışmasıyla daha güçlü öğrenme kabiliyetine sahip hale gelmesi ilkesine dayanır. Bu orman yapısı verilerin karmaşıklığını yansıtan daha küçük ve daha basit ağaçların koleksiyonudur. Her ağaca rastgele bir alt küme tahsis edilir. Hem özelliklerin hem de eğitim setinden oluşturulan örneklemelerin ağaçtan ağaca farklılık gösterebileceği özel büyüme algoritmasıdır. Leo Breiman ve Adele Cutler tarafından geliştirilen bu yöntem, rastgele özellik seçimi ile torbalamanın temel ilkelerini birleştirerek karar ağacı modellerine ek çeşitlilik ekler. Orman oluşturulduğunda, model ağaçların tahminlerini birleştirmek için bir oylama veya ortalama yöntemi ile tahmin için oy kullanır. Verilerin küçük rastgele alt kümelere bölünmesi tüm veri kümesini öğrenme avantajına sahip karmaşık tek ağaçtan daha iyi performans gösterebilir (Lantz, 2013).

Rastgele Orman yönteminin algoritması şu şekildedir:

Ormanda bulunan $k$ numaralı ağaç için, geçmişteki rastgele vektörler $Q_1$,...,$Q_{k-1}$' den bağımsız olarak fakat aynı dağılıma sahip rastgele bir $Q_k$ vektörü üretilir ve eğitim seti ile $Q_k$ kullanılarak bir ağaç büyütülür ve sonuçta $x$'in bir giriş vektörü olduğu $h(x, Q_k)$ sınıflandırıcısı elde edilir. Torbalama işlemi sırasında rastgele vektör $Q$ rastgele seçilmiş $N$ cinsinden sayımlar ile oluşturulur. $N$, eğitim setindeki örneklerin toplam sayısıdır. Bölünme işleminde, $Q$, 1 ile $k$ değerleri arasından rastgele seçilir ve tam sayı değerlerinden oluşur. $Q'$değerinin boyutu, ağaç oluşturulma sırasında kullanımına bağlıdır. Bir dizi ağaç oluşturulduktan sonra, en fazla oyu alan sınıf belirlenir. Bu yöntem, rastgele ormanlar olarak adlandırılır (Breiman, 2001).

**Adım 1:**

Rastgele orman, $\{h(x, Q_k), k = 1, \dots\}$ şeklinde tanımlanan ağaç yapılı sınıflandırıcıların bir koleksiyonudur; burada $\{Q_k\}$ bağımsız ve aynı şekilde dağıtılmış rastgele vektörlerdir ve her ağaç $x$ girişindeki en popüler sınıf için bir birim oy verir (Breiman, 2001).

**Adım 2:**

$h_1(x), h_2(x), \dots\dots, h_k(x)$ bir grup sınıflandırıcı ve $Y, X$ rastgele vektörlerinin dağılımından rastgele seçilen eğitim seti ile, marj fonksiyonunu (16) no'lu eşitlik ile tanımlanır:

$$mg(X, Y) = av_k\, I(h_k(X) = Y) - max_{j \neq Y} av_k\, I(h_k(X) = j) \qquad (16)$$

(16) no'lu eşitlikte; $av_k\, I(hk(X) = Y)$ : $h_k(X)$, $k$ numaralı sınıflandırıcının $X$ gözlemi için yaptığı tahmini temsil eder. $I(h_k(X) = Y)$ ise bu tahminin, gözlemin gerçek sınıfı olan $Y$

ile eşleşip eşleşmediğini kontrol eden gösterge fonksiyonudur. $max_{j\neq Y} av_k\, I(h_k\,(X) = j)$, gözlemin $Y$ dışında herhangi bir sınıfa atanma olasılığını hesaplar. Burada $j$, $Y$'den farklı tüm olası sınıfları temsil eder. Marj, $X, Y$' deki sınıfların ortalama oy sayısının diğer sınıfların ortalama oylarını ne ölçüde aştığını ölçer. Marj ne kadar büyük olursa, sınıflandırmaya olan güven de o kadar artar (Breiman, 2001).

Genelleme hatası şu şekildedir:

$$PE^* = P_{X,Y}\,(mg(X,Y) < 0) \tag{17}$$

(17) no'lu eşitlikte; $X, Y$ alt simgeleri olasılığın $X, Y$ alanının üzerinde olduğunu gösterir. Rastgele ormanlarda $h_k(X) = h(X, Q_k)$ Çok sayıda ağaç için Büyük Sayılar Kuvvetli Yasasından ve ağaç yapısından şu sonuç çıkar: Ağaçların sayısı arttıkça, neredeyse kesin olarak tüm $Q_1$,... $PE^*$ dizileri için yakınsar (Breiman, 2001).

$$P_{X,Y}\left(P_Q(h(X,Q) = Y) - max_{j\neq Y} P_Q(h(X,Q) = j) < 0\right) \tag{18}$$

(18) no'lu eşitlikte; ağaçların sayısı arttıkça, bir gözlemin doğru sınıfa atanma olasılığı ile yanlış sınıfa atanma olasılığı arasındaki farkın negatif olma olasılığına odaklanır. Rastgele ormandaki ağaçların sayısı yeterince büyük olduğunda, marj fonksiyonunun negatif olma olasılığı azalır, yani modelin doğru sınıflandırma yapma güveni artar. Bu sonuç, daha fazla ağaç eklendikçe rastgele ormanların neden aşırı uyum sağlamadığını, ancak genelleme hatasının sınırlayıcı bir değerini ürettiğini açıklamaktadır (Breiman, 2001).

**Adım 3:**

Rastgele ormanın Marj fonksiyonu (19) no'lu eşitlikte gösterilmiştir.

$$mr(X,Y) = P_Q(h(X,Q) = Y) - max_{j\neq Y}\, P_Q(h(X,Q) = j) \tag{19}$$

Sınıflandırıcılar kümesinin gücü $s$ (20) no'lu formül ile hesaplanır;

$$s = E_{X,Y}\, mr(X,Y) \tag{20}$$

ve sınıflandırıcının genel performans sınırı $PE^*$ (21) no'lu eşitlik ile belirlenir.

$$PE^* \leq \frac{var(mr)}{s^2} \tag{21}$$

$mr$ 'nin varyansı (22) no'lu formülde gösterildiği gibidir:

$$ǰ(X,Y) = argmax_{j \neq Y} P_Q(h(X,Q) = j) \quad (22)$$

$$mr(X,Y) = P_Q(h(X,Q) = Y) - P_Q\big(h(X,Q) = ǰ(X,Y)\big) \quad (23)$$

(23) no'lu eşitlik; marj fonksiyonunun hesaplanmasını gösterir. $ǰ(X,Y)$, $Y$ sınıfı dışındaki en yüksek olasılığa sahip sınıfı belirler. Marj fonksiyonu, $X$ girdisi için $Y$ sınıfının olasılığı ile en yüksek olasılığa sahip diğer sınıfın olasılığı arasındaki farkı verir (Breiman, 2001).

$$mr(X,Y) = E_Q[I(h(X,Q) = Y) - I(h(X,Q) = ǰ(X,Y)] \quad (24)$$

(24) no'lu formül, marj fonksiyonunun olasılık yerine, iki sınıf arasındaki farkı hesaplamanın beklenen değerini verir (Breiman, 2001).

**Adım 4:**

Ham marj fonksiyonu (25) no'lu formüldeki gibidir :

$$rmg(Q,X,Y) = I(h(X,Q) = Y) - I\big(h(X,Q) = ǰ(X,Y)\big) \quad (25)$$

Dolayısıyla, $mr(X,Y)$ , $rmg(Q,X,Y)$ 'nin $Q$' ya göre beklenen değeridir (Breiman, 2001).
(25) no'lu formülde belirtilen;

- ($Q$), kullanılan sınıflandırıcıyı (rastgele ormanı) temsil eder.
- ($X$), giriş özelliklerini ifade eder.
- ($Y$), gerçek sınıf etiketini ifade eder.
- $h(X,Q)$, sınıflandırıcı fonksiyonudur. $Q$, kullanılan rastgele vektörleri temsil eder.
- $ǰ(X,Y)$, $Y$ etiketi yerine en fazla marjı sağlayan etiketi ifade eder.
- $I(.)$ gösterilen olayın gösterildiği durumlarda 1, aksi halde 0 gösterge fonksiyonunu temsil eder (Breiman, 2001).

Bölüm içerisinde anlatılan uygulama kısmında kullanılacak olan yöntemlerin avantajları ve dezavantajları Tablo 2‘ de gösterilmiştir.

| YÖNTEM | AVANTAJLAR | DEZAVANTAJLAR |
|---|---|---|
| **Çoklu Doğrusal Regresyon** | - Yorumlanması ve uygulanması kolaydır.<br>- Model parametrelerinin anlamlılığı kolayca değerlendirilir.<br>- Aşırı uyum riski düşüktür. | - Doğrusal olmayan ilişkileri iyi modelleyemez.<br>- Aykırı değerlere (Outlier) duyarlıdır. |
| **Karar Ağacı** | - Karmaşık, doğrusal olmayan ilişkileri modelleyebilir.<br>- Yorumlaması kolaydır ve görselleştirme imkânı sunar.<br>- Veri ön işleme ihtiyacı düşüktür. | - Aşırı uyum (overfitting) sorunu yaşanabilir.<br>- Küçük değişiklikler ağaç yapısında büyük farklılıklara neden olabilir (yüksek varyans).<br>- Karar ağaçları büyük veri setlerinde yavaş çalışabilir. |
| **Rastgele Orman** | - Aşırı uyumu azaltmak için birçok karar ağacını bir araya getirir, genellikle daha yüksek doğruluk sağlar.<br>- Geniş veri setlerinde ve değişken sayısında iyi performans gösterir.<br>- Aykırı değerlerden daha az etkilenir. | - Yorumlaması zor olabilir, özellikle hangi değişkenlerin modelin tahminine daha fazla katkıda bulunduğunu belirlemek zor olabilir.<br>- Büyük hafıza ve işlem gücü gerektirebilir.<br>- Eğitimi uzun sürebilir. |
| **XGBoost** | - Yüksek performans ve doğruluk sunar, genellikle diğer yöntemlerden daha iyi sonuçlar verir.<br>- Aşırı uyumu önlemek için hiperparametre ayar teknikleri içerir.<br>- Hızlıdır ve büyük veri setlerinde iyi ölçeklenir. | - Yorumlaması zor olabilir, modelin iç işleyişini anlamak karmaşıktır.<br>- Parametre ayarlaması karmaşıktır ve genellikle optimizasyon gerektirir.<br>- Hafıza ve işlem gücü gereksinimi yüksektir. |

**Tablo 2 :** Çalışmada uygulanacak yöntemlerin kıyaslanması.

# DÖRDÜNCÜ BÖLÜM

# VERİ SETİ ANALİZİ

## 4.1. VERİ SETİ

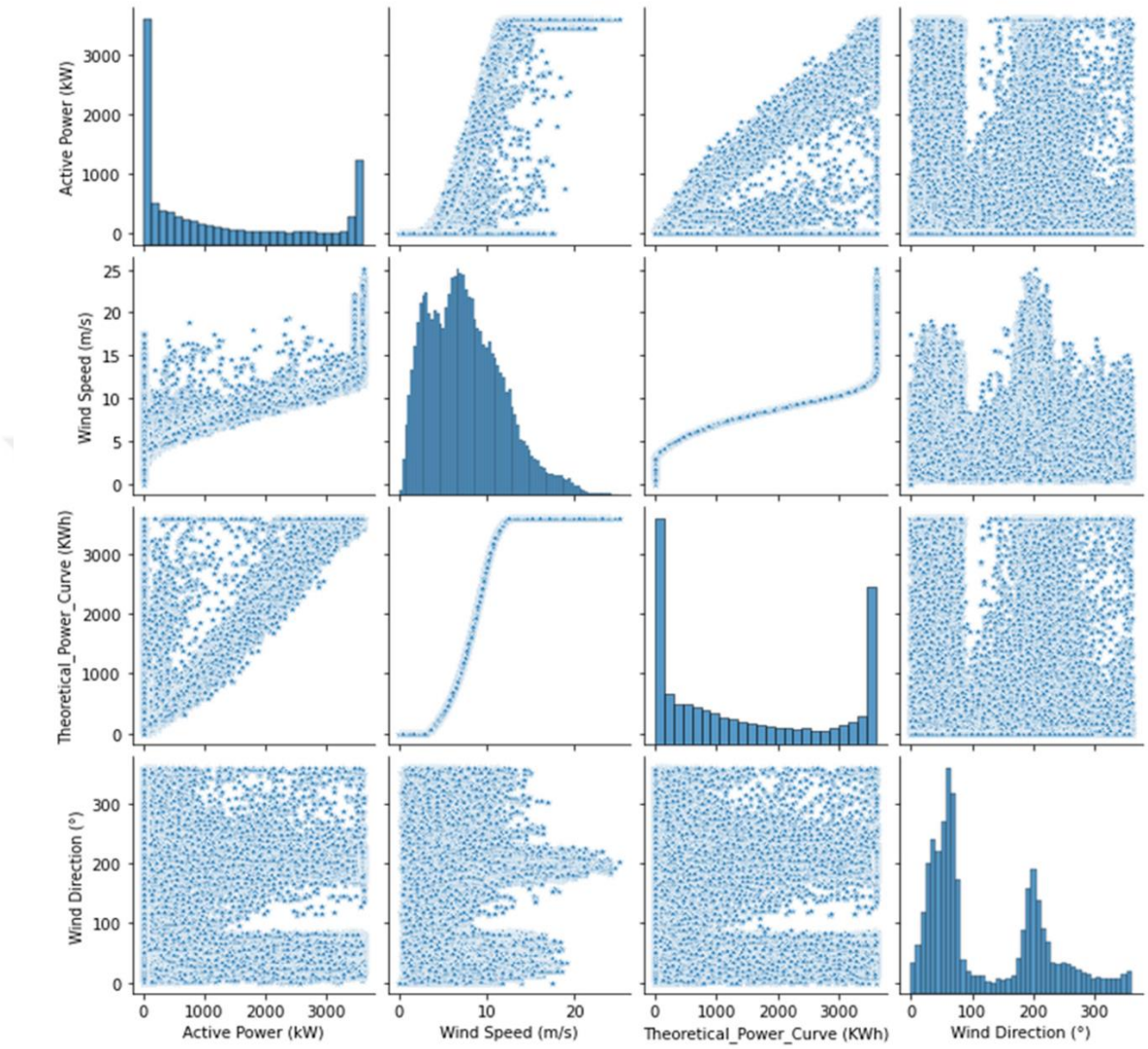
Türkiye Yalova'da bulunan bir rüzgâr türbininin SCADA sisteminden alınan veriler kullanılmıştır. Rüzgâr hızı, rüzgâr yönü, aktif güç ve teorik güç bilgileri 01.01.2018 başlangıç ve 31.12.2018 bitiş tarihleri arasındaki onar dakikalık periyotlarla SCADA sisteminde kaydedilmiş veriler ele alınmıştır (Erisen, 2018).

Aktif güç, rüzgâr türbini tarafından üretilen gücü ifade eder. Teorik güç, SCADA ünitesinin kontrol sistemi içindeki üretim formülü tarafından hesaplanan güç değeridir Aktif gücün teorik güce yakın olması yüksek verimlilikle türbinin çalışmasına bağlıdır (Fahim vd., 2022).

Veri setindeki bağımlı değişken aktif güç iken rüzgâr hızı, teorik güç ve rüzgâr yönü bağımsız değişkenlerdir.

| | **Rüzgâr Yönü(°)** | **Teorik Güç(kW)** | **Rüzgâr hızı (m/s)** | **Aktif Güç(kW)** |
|---|---|---|---|---|
| **Veri sayısı** | 50530 | 50530 | 50530 | 50530 |
| **Max** | 359.9 | 3600 | 25.20 | 3618.7 |
| **Min** | 0 | 0 | 0 | -2.471 |
| **Standart sapma** | 93.4 | 1368 | 4.22 | 1312.4 |
| **Ortalama** | 123.6 | 1492.1 | 7.55 | 1307.6 |

**Tablo 3 :**Veri setinin istatistiksel ölçümleri (Singh & Rizwan, 2022).

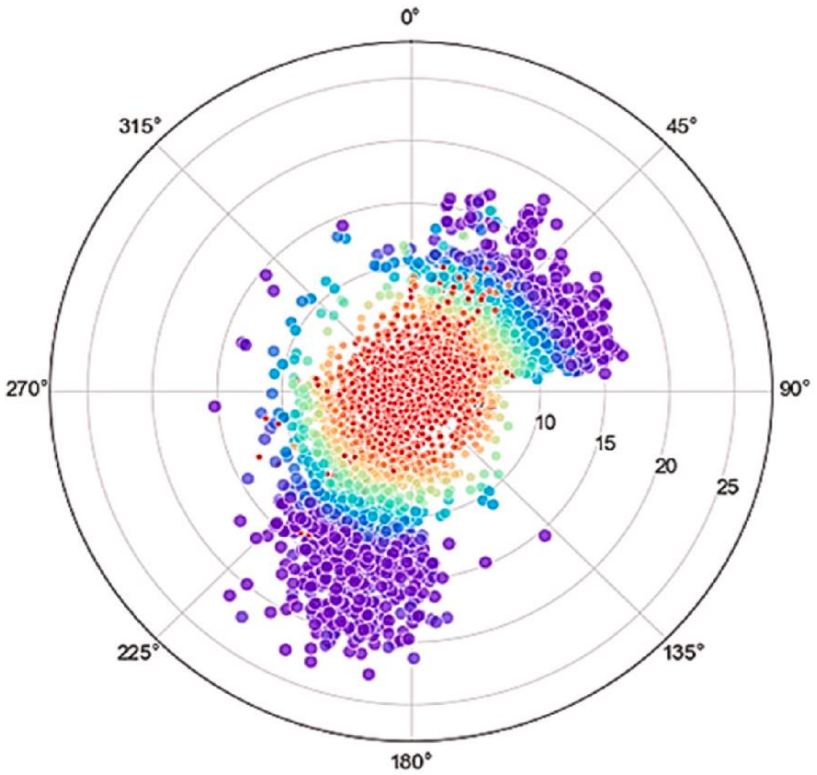



**Şekil 8 :** Veri setindeki bağımlı ve bağımsız değişkenlerin aralarındaki ilişkiyi gösteren grafikler (Singh & Rizwan, 2023).

Veri seti incelendiğinde bazı aktif güç değerlerinin negatif olduğu görülmüştür. Bunun nedenleri, rüzgâr türbini durduğunda veya rüzgâr hızı çok düşük olduğunda, türbin rotoru ters yönde çalışabilir. Bu durumda, türbinin sisteminin kendi tükettiği enerji için elektrik şebekesinden enerji çekmesi gerekir. Bu enerji, veri setinde negatif aktif güç olarak kaydedilebilir. Ayrıca sensörler veya ölçüm cihazlarında kalibrasyon hataları veya teknik sorunlar varsa, bu durum negatif güç değerlerine yol açabilir.

Aynı veri seti ile önceden yapılan çalışmalarda, bu negatif güç değerleri aykırı değer (“outlier”) olarak görüldüğü için modelleme öncesi veri setinden çıkarılmıştır. Ancak sistemin kaydettiği gerçekleşmesi mümkün bu negatif güç değerleri çalışmamızda veri setine dahil olacak şekilde modelleme yapılmıştır. Sadece rüzgâr hızı sıfır iken aktif güç değerlerinin de

sıfır alınması işlemi ile veri seti yeniden düzenlenmiştir. Bunun dışındaki tüm değerler olduğu gibi aynen modelleme sırasında kullanılmıştır.




**Şekil 9 :**Polar koordinat sisteminde rüzgâr hızı, rüzgâr yönü ve güç üretimi arasındaki ilişkiyi gösteren diyagram (Singh & Rizwan, 2023).

Polar grafik, rüzgârın estiği yüksek verimli bölgeleri belirlemek için kullanılır. Üretilen gücün rüzgâr yönü ve rüzgâr hızı ile ilişkili olarak nasıl dağıldığını gösterir. Rüzgâr 180° ile 225° ve 0° ile 90° arasındaki yönlerden estiğinde daha fazla elektrik üretir. Ayrıca, polar grafikteki küçük noktalar daha az güç üretimini, büyük noktalar ise rüzgâr türbininden daha fazla güç üretimini gösterir (Singh & Rizwan, 2023).

# BEŞİNCİ BÖLÜM

# R PROGRAMLAMA DİLİNDE UYGULAMA

## 5.1. MODEL PERFORMANSI VE HATA DEĞERLERİ

Veri setinin bir kısmı (%70 veya %80) eğitim verileri geri kalan kısmı (%20 veya%30) ise test verileri olmak üzere ikiye ayrılır. Hedef değişkenin tahmin edilmesinden önce oluşturulan model ayırdığımız test verileri üzerinde denenerek performansı ölçülür. Bu ölçüm sırasında bazı metrikler kullanılır. Bunlardan bazıları MAE, RMSE ve $R^2$ 'dir (Kuhn & Silge, 2022).

MAE, bir regresyon modelinin tahmin değerlerinin gerçek değerlerden ne kadar uzaklaştığını gösteren bir performans ölçütüdür. MAE, her tahminin gerçek değerle arasındaki mutlak farkların ortalamasının hesaplanmasıdır (Ahmadi vd., 2020).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} | y_i - \hat{y}_i | \quad (26)$$

(26) no'lu formülde; $n$ : veri noktalarının toplam sayısını, $y_i$ : gerçek değerleri ve $\hat{y}_i$ : modelin tahmin ettiği değerleri temsil eder.

MAE' nin değeri, sıfıra ne kadar yakınsa, modelin oluşturduğu tahmin değerlerinin gerçek değerlere ne kadar yakınlaştığını gösterir. MAE' nin düşük değerde olması, modelin daha iyi tahminler yaptığını ve daha az hata yaptığını ifade eder (Ahmadi vd., 2020).

RMSE, bir regresyon modelinin tahmin değerlerinin gerçek değerlerden ne kadar uzaklaştığını gösteren bir performans ölçütüdür. RMSE, her bir tahminin gerçek değerden farkının karelerinin ortalamasının kare kökü olarak hesaplanır (Ahmadi vd., 2020).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (27)$$

(27) no'lu formülde; $n$ : veri noktalarının toplam sayısını, $y_i$ : gerçek değerleri ve $\hat{y}_i$ : modelin tahmin ettiği değerleri temsil eder.

RMSE' nin değeri, sıfıra ne kadar yakınsa, modelin oluşturduğu tahmin değerlerinin gerçek değerlere ne kadar yakınlaştığını gösterir. RMSE' nin düşük olması, modelin daha iyi tahminler yaptığını ve daha az hata yaptığını ifade eder. RMSE, aykırı değerlerin etkisinin azaltılmasında ve hataların büyüklüğünün ölçülmesinde kullanışlıdır (Ahmadi vd., 2020).

R-kare($R^2$), bir regresyon modelinin ne kadar iyi uyum sağladığını ve bağımsız değişkenler ile bağımlı değişkenin arasındaki varyansı ölçen bir istatistiksel ölçümdür. R-kare değeri genellikle 0 ile 1 arasında bir değerdir ve modelin verilere ne kadar iyi uydurulduğunu gösterir (Ahmadi vd., 2020).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (28)$$

(28) no'lu formülde; $y_i$ : gerçek değerleri, $\hat{y}_i$ : modelin tahmin ettiği değerleri ve $\bar{y}$ :gerçek değerlerin ortalamasını temsil eder.

$R^2$ değeri, modelin ortaya çıkardığı varyansın toplam varyansa oranını ifade eder. $R^2$ değeri 1 değerine yaklaştıkça, modelin verilere oldukça iyi uyduğunu gösterir. Ancak, $R^2$ yalnızca bağımlı değişkenin varyansının açıklanmasını değerlendirir ve modelin doğruluğunu ölçmez. Bu nedenle, diğer performans ölçümleri RMSE veya MAE ile birlikte değerlendirilmelidir (Ahmadi vd., 2020).

### 5.1.1. Çapraz Doğrulama ile Model Performansı Değerlendirme

Makine öğrenmesi modellerinde ayrıca $k$ katlamalı çapraz doğrulama ("$k$ fold cross-validation") işlemi ile de model performansı ölçme işlemi yapılır. Çapraz doğrulama ("cross-validation"), makine öğrenimi modellerinin performansını ölçmek için kullanılır. Bu yöntem, modelin genelleştirme kapasitesini daha güvenilir bir şekilde test etmek için uygulanır ve aşırı uyum gibi sorunları tespit etmeye yardımcı olur. En yaygın kullanılan çapraz doğrulama yöntemi, $k$-katlamalı çapraz doğrulamadır. Bu yöntemde, veriler rastgele olarak $k$ eşit parçaya bölünür. Her bir $k$ parçası sırasıyla test seti olarak kullanılırken, kalan $k-1$ parça ise modelin eğitimi için kullanılır. Bu adımlar, her parçanın test seti olarak kullanıldığı bir döngü içinde tekrarlanır (Kuhn & Silge, 2022).

Eğitim verileri ile oluşturduğumuz modelimizin test verileri üzerindeki tahminleri sonucunda gerçek değerlerden ne kadar uzaklaştığımızı veya yakın değerler elde ettiğimizi ölçmek için bu açıklanan hata ölçütleri ile performans ölçümü yapıldıktan sonra modelin performansını yükseltmek ve hata oranlarını azaltmak için seçilen makine öğrenmesi yöntemlerine göre hiperparametre ayarı yapılmalıdır. Ağaç tabanlı makine öğrenmesi yöntemlerini ele alırsak buradaki hiperparametre örnekleri seçilen ağaç sayısı, ağaçların derinliği veya seçilen özellik sayısı gibi olabilmektedir (Kuhn & Silge, 2022).

Çalışmada kullanılan yöntemlerin performansları seçilen hata metriklerinden RMSE ve $R^2$ değerlerine göre karşılaştırılacaktır. Hata metriklerinin temelinde tahmin hatasının hesaplanması vardır. Bunu yapmak için ilk önce oluşturulacak modeller seçilmiş bazı verilerle

eğitilmiş, sonrasında ise oluşan modeller daha önce hiç görmediği veriler ile test edilmiştir. Çalışmada aktif güç değerleri bağımlı değişkeni oluştururken rüzgâr hızı, teorik güç ve rüzgâr yönü verileri ise bağımsız değişkenlerdir. Veri setindeki toplam veri sayısının %70'i eğitim verileri, %30'u test verileri şeklinde ayrılmıştır. Daha sonra Çoklu doğrusal regresyon yöntemi ile Makine Öğrenmesi yöntemlerinden Karar ağacı, Rastgele orman ve XGBoost algoritmaları ile R programlama dilinde modeller kurulmuş ve tahmin hataları RMSE ve $R^2$ değerleri tespit edilerek en iyi sonuç veren yöntem belirlenmiştir.

### 5.2. R PROGRAMLAMA DİLİNDE ÇOKLU DOĞRUSAL REGRESYON

```
# Gerekli kütüphaneler yüklenir, eksik olanlar indirilir ve veri seti excel dosyasından okunur.
library(pacman)
p_load(readxl,
       dplyr,
       ggplot2,
       rpart,
       caret,
       ranger,
       xgboost,
       Matrix,
       car,
       doParallel,
       lmtest)
windpower <- read_excel("WP.xlsx")

       # Veri setindeki sütunlar daha basit şekilde yeniden adlandırılır.
wp <- windpower %>%
  rename(
    active_power = "LV ActivePower (kW)",
    wind_speed = "Wind Speed (m/s)",
    power_curve = "Theoretical_Power_Curve (KWh)",
    wind_dir = "Wind Direction"
  )
       # Rüzgâr hızı 0 olduğunda aktif gücün de 0 olması için uygulanır.
wp$active_power <- ifelse(wp$wind_speed == 0, 0, wp$active_power)
```

```
        # Zaman sütunu çıkarılır.
wp <- wp %>% select(-"Date/Time")
        # Veri seti %70 eğitim, %30 test verisi olarak bölünür.
set.seed(123)
train_index <- createDataPartition(wp$active_power, p = 0.7, list = FALSE)
training_data <- wp[train_index, ]
test_data <- wp[-train_index, ]
        # Modellerin RMSE ve R2 değerlerini toplayan data frame oluşturulması
results_df <- data.frame(
  Model = character(),
  RMSE = numeric(),
  R2 = numeric(),
  stringsAsFactors = FALSE
)
        # Çoklu doğrusal regresyon
multiple_reg <- lm(active_power ~ ., data = training_data)
```

active_power bağımlı değişkeni için tüm bağımsız değişkenleri içeren eğitim verisi ile çoklu doğrusal regresyon modeli kurulur.

```
summary(multiple_reg)
-Call:
lm(formula = active_power ~ ., data = training_data)

Residuals:
    Min      1Q  Median      3Q     Max
-3427.0   -33.0    46.6   140.3   759.7

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.489e+02  6.501e+00  -38.28   <2e-16 ***
wind_speed   4.441e+01  1.538e+00   28.87   <2e-16 ***
power_curve  7.847e-01  4.768e-03  164.59   <2e-16 ***
wind_dir     4.090e-01  2.312e-02   17.70   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 403 on 35370 degrees of freedom
Multiple R-squared:  0.9059, Adjusted R-squared:  0.9059
F-statistic: 1.135e+05 on 3 and 35370 DF,  p-value: < 2.2e-16
```

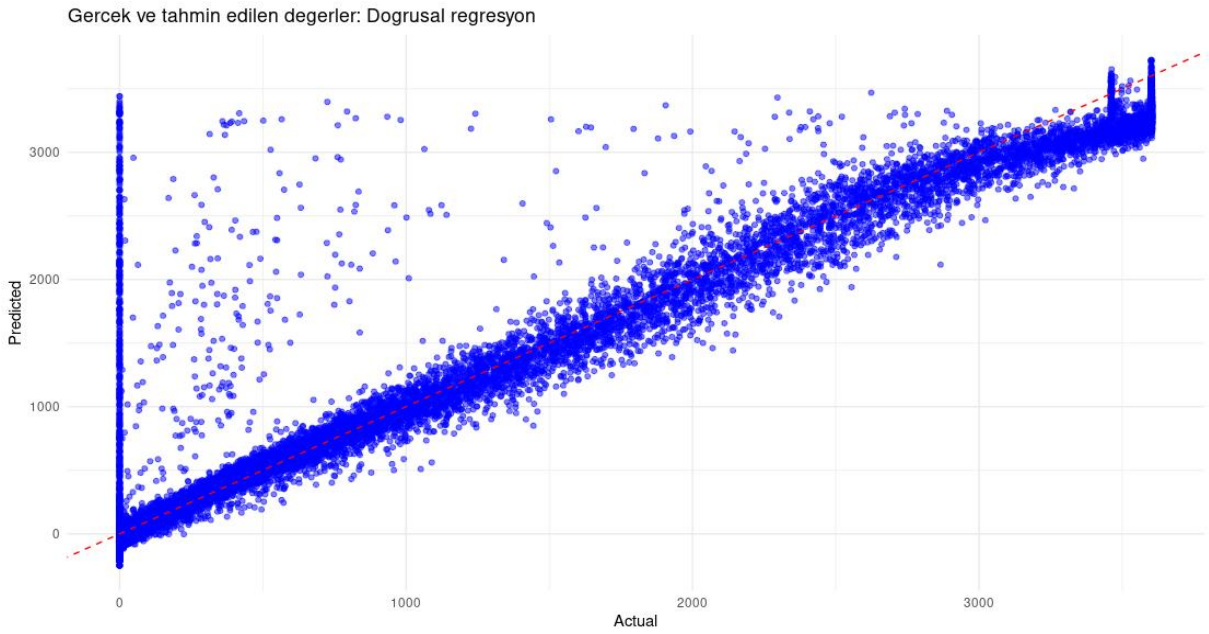
```
        # Tahminler ve Performans Değerlendirme
test_data$predicted_power <- predict(multiple_reg, newdata = test_data)
(lm_rmse <- RMSE(test_data$predicted_power, test_data$active_power))
[1] 404.4346
(lm_r2 <- R2(test_data$predicted_power, test_data$active_power))
[1] 0.9046027
```

Test verileriyle tahmin yapılır. RMSE ve R² değerleri hesaplanır.

### 5.2.1. Çoklu Doğrusal Regresyonda Gerçek ile Tahmin Değerleri Grafiği ve Varsayımların Kontrolü

```
        results_df <- rbind(results_df,
                data.frame(Model = "Dogrusal regresyon", RMSE = lm_rmse, R2 = lm_r2))
        # Çoklu Doğrusal Regresyon Modeli için Gerçek ve Tahmin Değerleri Görselleştirme
ggplot(
  data.frame(
    Actual = test_data$active_power,
    Predicted = test_data$predicted_power
  ),
  aes(x = Actual, y = Predicted)
) +
  geom_point(color = "blue", alpha = 0.5) +
  geom_abline(
    intercept = 0,
    slope = 1,
    color = "red",
    linetype = "dashed"
  ) +
  labs(title = "Gercek ve tahmin edilen degerler: Dogrusal regresyon", x = "Actual", y =
"Predicted") +
  theme_minimal()
```

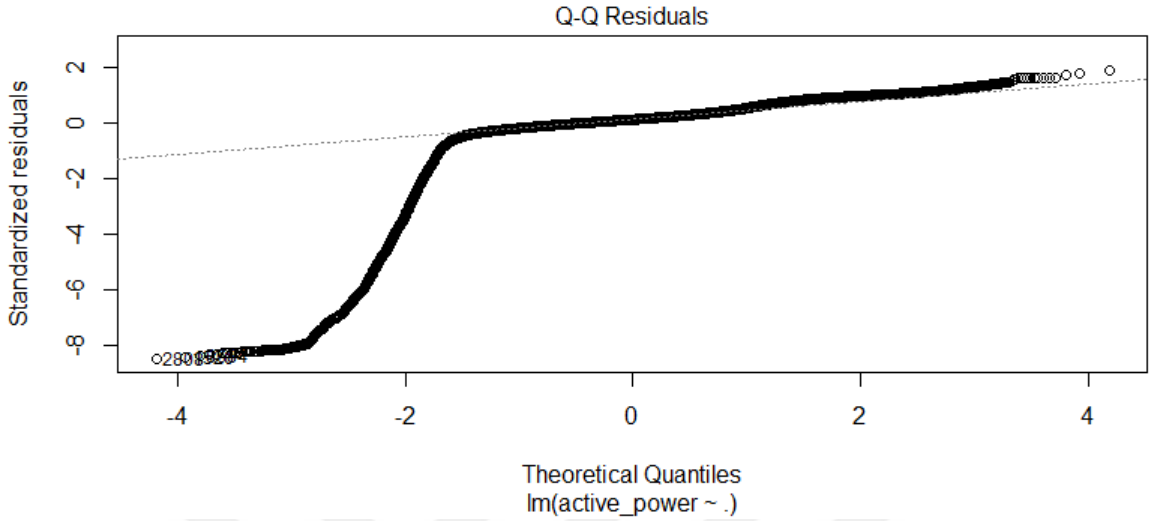
Bu komut, çoklu doğrusal regresyon modelinin gerçek değerler (aktif güç) ile tahmin edilen değerler arasındaki ilişkiyi bir noktalar grafiği (“scatter plot”) ile gösterir. Grafik üzerinde yer alan kırmızı kesikli 45 derece doğrusu, tahminlerin mükemmel olması durumunda bulunması gereken referans çizgisidir. Noktalar bu çizgiye ne kadar yakınsa, modelin tahminleri o kadar iyidir.




**Q-Q grafiği**

plot(multiple_reg, which = 2)

Çoklu doğrusal regresyon modelinin Q-Q (“Quantile-Quantile”) grafiğini oluşturur.




Bu Q-Q grafiği, artıkların normal dağılıma uyumunu değerlendirmek için oluşturulur. Oluşturduğumuz grafiğin yorumu şu şekilde yapılabilir: Orta kısımdaki noktalar referans çizgisine oldukça yakın görünüyor, bu da artıkların merkezde normal dağılıma iyi bir şekilde uyduğunu gösterir. Ancak, grafiğin başındaki ve sonundaki noktalar referans çizgisinden belirgin şekilde sapma gösteriyor. Sol tarafta noktalar referans çizgisinden aşağıya doğru sapmıştır. Bu durum, artıkların düşük değerlerde normal dağılıma uymadığını ve olası uç değerlerin bulunduğunu gösterir. Sağ tarafta, noktalar yukarıya doğru sapma göstermiş. Bu da artıkların yüksek değerlerde normal dağılıma uymadığını gösterir. Artıkların normal dağılım varsayımı tam anlamıyla karşılanmadığı grafikte görünmektedir. Özellikle uç değerler, bu sapmalara katkıda bulunabilir. Bu durum, modelin doğruluğunu etkileyebilir ve modelde iyileştirmeler yapılmasını gerektirebilir.

```
bptest(multiple_reg)
```

studentized Breusch-Pagan test

```
data:  multiple_reg
BP = 1275.8, df = 3, p-value < 2.2e-16
```

Bu komut, Breusch-Pagan testini çalıştırır. Test, multiple_reg modeli için varyans türdeşliği (“homoskedasticity”) olup olmadığını kontrol eder. Breusch-Pagan testi, modelin artıklarının varyansının türdeş olup olmadığını test eder. Eğer varyans türdeş değilse, bu regresyon modelinizin tutarsız tahminler yapmasına neden olabilir.

- BP değeri: 1275.8 (Bu test istatistiğidir. Yüksek BP değeri, değişen varyansın var olabileceğine işaret eder.)
- df: 3 (Serbestlik derecesi)
- p-değeri: 2.2e-16'den küçük (Bu çok küçük bir p-değeridir.)

Bu test sonucuna göre, değişen varyans sorunu vardır. Yani hataların varyansı türdeş değildir. Bu da modelin tahminlerinde tutarsızlığa neden olabilir ve modelin güvenilirliğini düşürebilir.

dwtest(multiple_reg)

Durbin-Watson test

data: multiple_reg

DW = 0.16484, p-value < 2.2e-16

alternative hypothesis: true autocorrelation is greater than 0 (Artıklar arasında pozitif otokorelasyon olduğu hipotezi)

Durbin-Watson testi, modelde otokorelasyon (artıklar arasında ardışık bağımlılık) olup olmadığını test eder. Durbin-Watson istatistiği, tipik olarak 0 ile 4 arasında bir değer alır:

- DW ≈ 2: Artıklar arasında bağımsızlık vardır (otokorelasyon yoktur).
- DW < 2: Pozitif otokorelasyon vardır (artıklar arasında pozitif bağımlılık).
- DW > 2: Negatif otokorelasyon vardır (artıklar arasında negatif bağımlılık).

DW = 0.16484: Bu çok düşük bir değerdir ve pozitif otokorelasyonun çok güçlü olduğunu gösterir. Yani artıklar arasında ardışık olarak pozitif bir bağımlılık vardır.

p-değeri < 2.2e-16: Bu, sıfır hipotezini (artıklar arasında otokorelasyon yoktur) reddettiğimiz anlamına gelir. p-değerinin bu kadar küçük olması, otokorelasyonun çok güçlü olduğunu destekler.

Artıklar arasında pozitif otokorelasyon vardır. Bu, modelde otokorelasyon olduğunu ve artıkların birbirleri ile ilişkili olduğunu gösterir.

Bu test sonucu, modelde pozitif otokorelasyonun var olduğunu ve bu durumun modelin güvenilirliğini etkileyebileceğini gösteriyor.

Bu nedenlerden dolayı çoklu doğrusal regresyon yöntemine göre diğer ele aldığımız makine öğrenmesi yöntemlerinin kullanılması oluşturulacak modellerin tahminler üzerindeki performansının daha iyi ve karmaşık veri yapılarına karşı daha güvenilir olabileceğini göstermektedir.

### 5.3. R PROGRAMLAMA DİLİNDE RASTGELE ORMAN

```
        # Paralel işlemcilerin oluşturulması;
num_cores <- detectCores() # Maksimum çekirdek sayısını belirler.
cl <- makeCluster(num_cores)
registerDoParallel(cl)
Paralel işlem yaparak modeli hızlandırır.
        # Hiperparametre optimizasyonu icin trainControl ayarı
train_control <- trainControl(
  method = "cv",
  # çapraz doğrulama
  number = 5,
  # 5 katli çapraz doğrulama
  verboseIter = TRUE,
  # Süreç çıktısını göster
  allowParallel = TRUE    # Paralel işlemi etkinleştir
)
        # Hiperparametre ızgarasını tanımla
tune_grid <- expand.grid(
  mtry = c(2),
 # Rastgele seçilecek değişken sayısı
 splitrule = c("variance"),
 # Bölünme kuralı
min.node.size = c(1)  # Minimum yaprak boyutu
)
        # Ranger modeli için train fonksiyonu
set.seed(123)
ranger_model <- train(
  active_power ~ wind_speed + power_curve + wind_dir,
  # Model formülü
  data = training_data,
  # Eğitim verisi
  method = "ranger",
  # ranger modeli
  trControl = train_control,
```

```
  # çapraz doğrulama kontrolü
  tuneGrid = tune_grid,
  # Hiperparametre ızgarası
  num.threads = num_cores          # Ranger paketi ile paralel işlem
)
-Aggregating results
-Fitting final model on full training set
          # Paralel işlemi kapat
stopCluster(cl)
registerDoSEQ()
          # En iyi hiperparametreler
print("En iyi parametreler:")
[1] "En iyi parametreler:"
print(ranger_model$bestTune)
  mtry splitrule min.node.size
    2   variance         1
          # Model performansı
print("Model detaylari:")
[1] "Model detaylari:"
print(ranger_model)
-Random Forest

35374 samples
    3 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 28299, 28299, 28299, 28301, 28298
Resampling results:

      RMSE     Rsquared       MAE
  404.2559  0.9054755  167.6357
Tuning parameter 'mtry' was held constant at a value of 2
Tuning parameter 'splitrule'
```

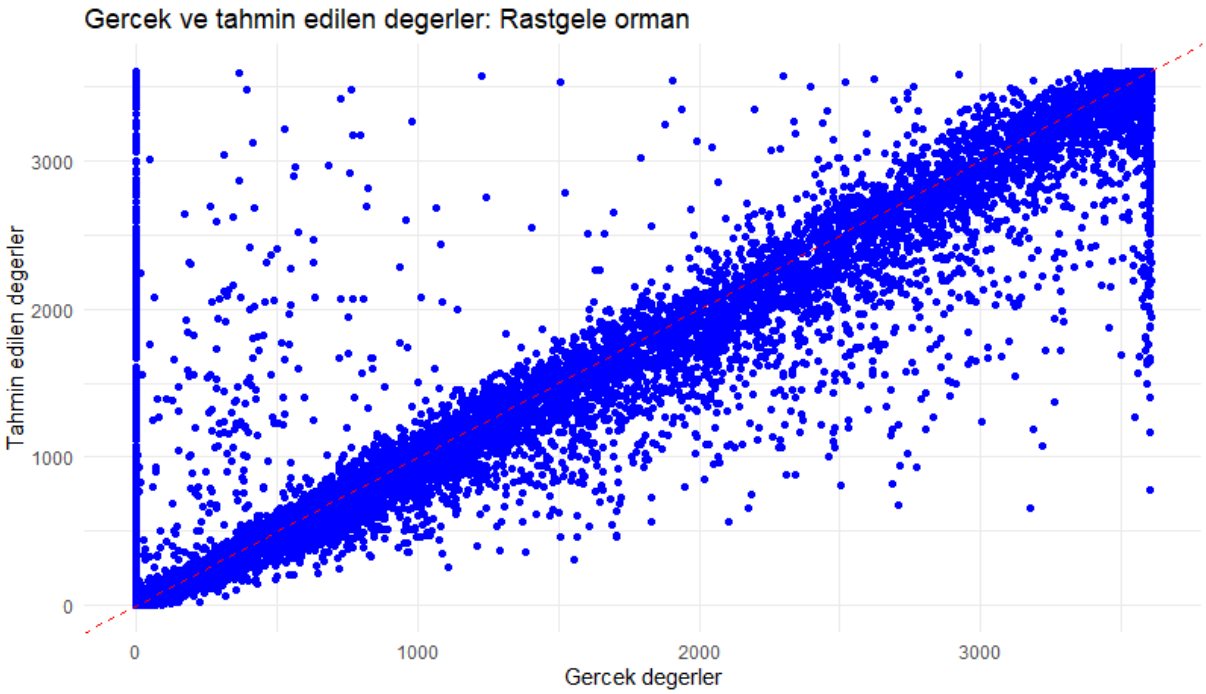
```
 was held constant at a value of variance
Tuning parameter 'min.node.size' was held
 constant at a value of 1
        # Test verisi ile tahmin
rf_predictions <- predict(ranger_model, newdata = test_data)
        # Model değerlendirme (RMSE ve R2)
(rf_rmse <- RMSE(rf_predictions, test_data$active_power))
[1] 410.7135
(rf_r2 <- R2(rf_predictions, test_data$active_power))
[1] 0.901806
cat("Paralel islem sonrasi hiperparametre pptimizasyonu:\n")
-Paralel islem sonrasi hiperparametre pptimizasyonu:
cat("RMSE:", rf_rmse, "\n")
-RMSE: 410.7135
cat("R2:", rf_r2, "\n")
-R2: 0.901806
results_df <- rbind(results_df,
+                    data.frame(Model = "Rastgele orman", RMSE = rf_rmse, R2 = rf_r2))
```

Hiperparametrelerin model için en ideal değerlerini bulmak için ızgara arama (“Grid search”) yöntemi sıkça kullanılır. Farklı değerler belirlenir ve model performansı değerlendirilir. Performansı ölçmek için genellikle doğrulama setindeki hata oranı veya k-kat çapraz doğrulama (k-fold cross-validation) kullanılır (Lantz, 2013).

### 5.3.1. Rastgele Orman Yöntemi için Gerçek ile Tahmin Değerleri Grafiği

```
results_rf <- data.frame(Gercek = test_data$active_power, Tahmin = rf_predictions)
ggplot(results_rf, aes(x = Gercek, y = Tahmin)) +
  geom_point(color = "blue") +
  geom_abline(
    slope = 1,
    intercept = 0,
    color = "red",
    linetype = "dashed"
  ) +
  labs(title = "Gercek ve tahmin edilen degerler: Rastgele orman", x = "Gercek degerler", y =
"Tahmin edilen degerler") +
```

theme_minimal()




Grafik, tahmin_gercek nesnesindeki Gercek ve Tahmin sütunlarını kullanarak oluşturulur. x eksenine gerçek değerler, y eksenine ise tahmin edilen değerler yerleştirilir.
geom_point(color = "blue"): Her veri noktası için bir mavi nokta çizer, bu noktalar tahmin edilen değerlerle gerçek değerlerin karşılaştırılmasını görselleştirir.
geom_abline(slope = 1, intercept = 0, color = "red", linetype = "dashed"): Eğim 1 ve kesişim noktası 0 olan bir kırmızı, kesikli çizgi ekler. Bu çizgi, mükemmel bir tahmin hattını temsil eder, yani tahminler tam olarak gerçek değerlere eşitse noktalar bu çizgi üzerinde yer alacaktır.
labs(...): Grafik başlığı ve eksen etiketlerini belirler.
theme_minimal(): Grafiğe sade bir tema uygular.

### 5.4. R PROGRAMLAMA DİLİNDE KARAR AĞACI

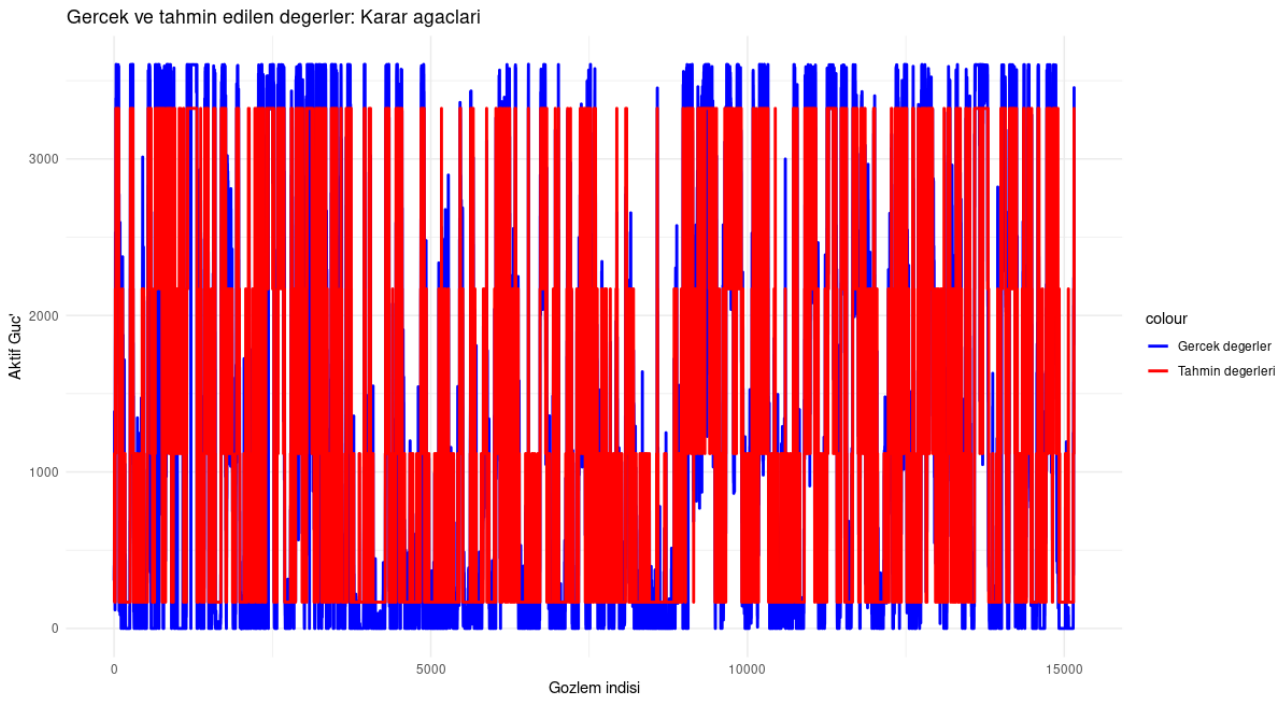
tree_model <- rpart(active_power ~ ., data = training_data, method = "anova")
tree_predictions <- predict(tree_model, newdata = test_data)
(tree_rmse <- RMSE(tree_predictions, test_data$active_power))
[1] 464.405
(tree_r2 <- R2(tree_predictions, test_data$active_power)) (Parametre optimizasyonu yapılmadan)

```
[1] 0.8742223
        # Karar ağacı optimal cp ve max depth değeri (Parametre optimizasyonu ile)
train_control <- trainControl(method = "cv", number = 5)
tune_grid <- expand.grid(cp = seq(0.01, 0.1, by = 0.01))
tuned_model <- train(
  active_power ~ wind_dir + power_curve + wind_speed,
  data = training_data,
  method = "rpart",
  trControl = train_control,
  tuneGrid = tune_grid
)
print(tuned_model$bestTune)
   cp
 0.06
train_control <- trainControl(method = "cv", number = 10)
tune_grid <- expand.grid(maxdepth = seq(1, 5, by = 1))
tuned_model <- train(
  active_power ~ wind_dir + power_curve + wind_speed,
  data = training_data,
  method = "rpart2",
  trControl = train_control,
  tuneGrid = tune_grid
)
print(tuned_model$bestTune)
 maxdepth
       3
        # En iyi cp ve max depth değerleri ile Karar Ağacı Modeli
tree_control <- rpart.control(cp = 0.06, maxdepth = 3)
tree_model_optimized <- rpart(
  active_power ~ wind_dir + power_curve + wind_speed,
  data = training_data,
  method = "anova",
  control = tree_control
)
```

```
tree_predictions_optimized <- predict(tree_model_optimized, newdata = test_data)
(tree_rmse_optimized <- RMSE(tree_predictions_optimized, test_data$active_power))
[1] 464.405
(tree_r2_optimized <- R2(tree_predictions_optimized, test_data$active_power)) (Parametre
optimizasyonu yapılmadan)
[1] 0.8742223
results_df <- rbind(
   results_df,
   data.frame(Model = "Karar agaci", RMSE = tree_rmse_optimized, R2 = tree_r2_optimized)
 )
```

#### 5.4.1. Karar Ağacı Yöntemi için Gerçek ile Tahmin Değerleri Grafiği

```
results_tree_optimized <- data.frame(
  Index = 1:nrow(test_data),
  Actual = test_data$active_power,
  Predicted = tree_predictions_optimized
)
ggplot(results_tree_optimized, aes(x = Index)) +
  geom_line(aes(y = Actual, color = "Gercek degerler"), linewidth = 1) +
  geom_line(aes(y = Predicted, color = "Tahmin degerleri"), linewidth = 1) +
  labs(title = "Gercek ve tahmin edilen degerler: Karar agaclari", x = "Gozlem indisi", y = "Aktif
Guc"') +
  scale_color_manual(values = c(
    "Gercek degerler" = "blue",
    "Tahmin degerleri" = "red"
  )) +
  theme_minimal()
```




### 5.5. R PROGRAMLAMA DİLİNDE XGBOOST

```
        # Paralel işlemciler kurulur.
cores <- detectCores()  # Maksimum işlemci sayısı
cl <- makeCluster(cores)
registerDoParallel(cl)

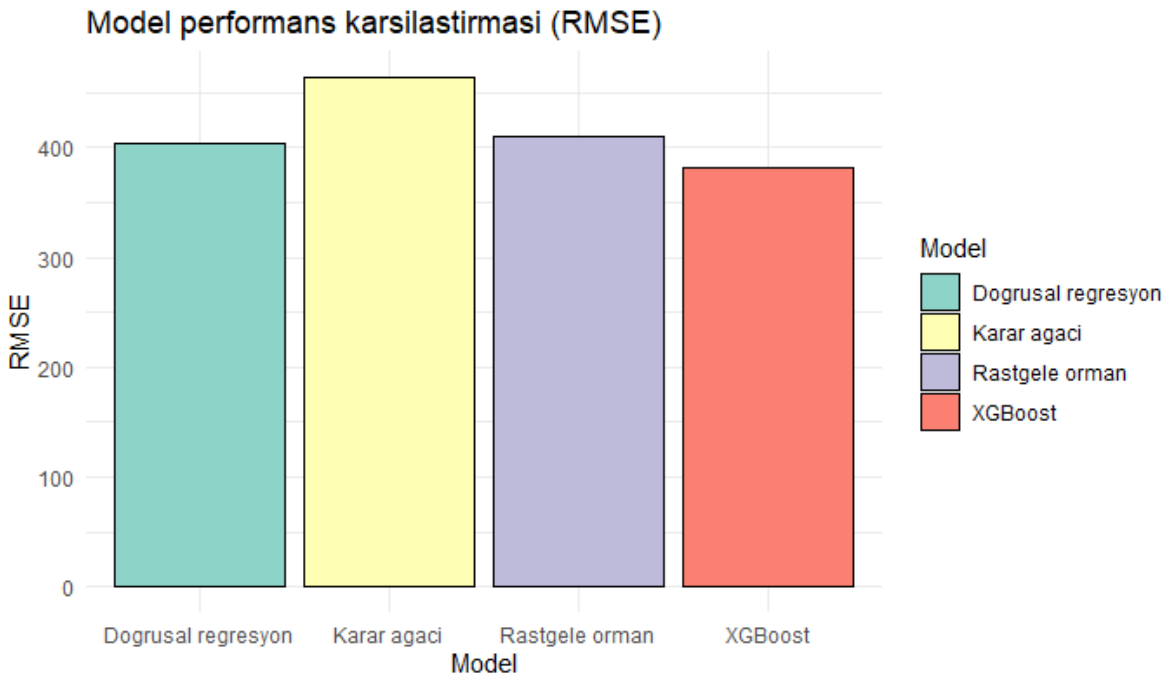
        # XGBoost için matris hazırlığı
train_matrix <- xgb.DMatrix(data = as.matrix(training_data %>% select(-active_power)),
                            label = training_data$active_power)

test_matrix <- xgb.DMatrix(data = as.matrix(test_data %>% select(-active_power)),
                           label = test_data$active_power)

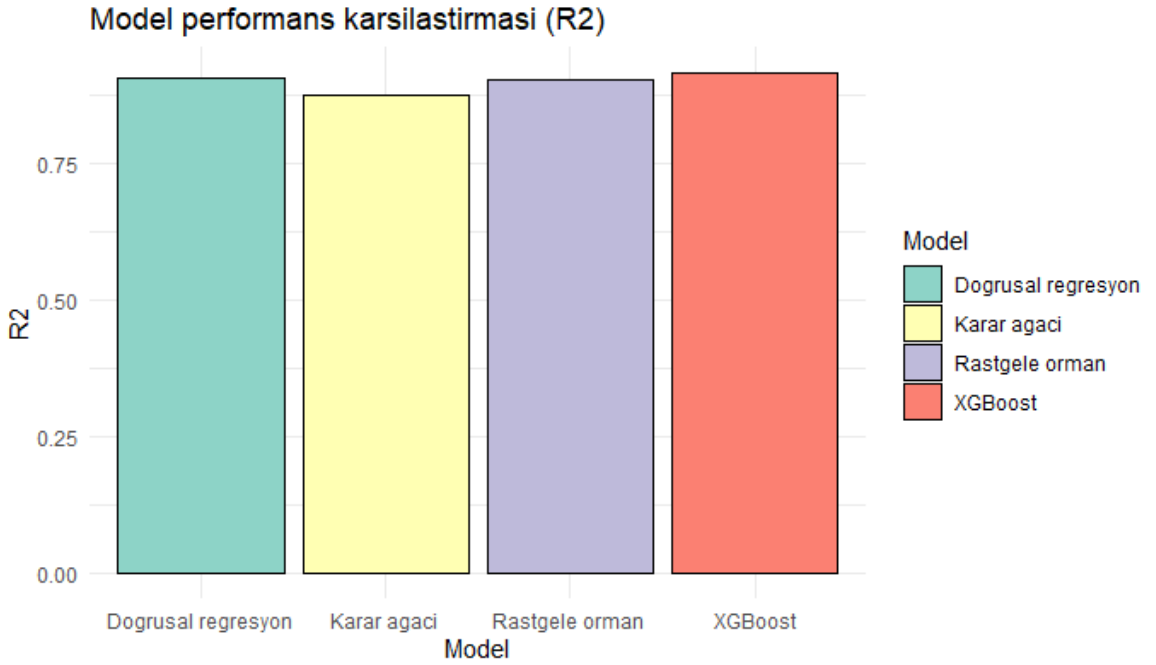
        # Hiperparametre için ızgara(“grid”) tanımlanması
grid <- expand.grid(
  nrounds = c(50),
  # Daha az iterasyon (daha hızlı işlem ve optimize etmek için)
  max_depth = c(6),
  # Maksimum derinlik
```

```
  eta = c(0.1),
  # öğrenme oranı
  gamma = c(0),
  # Minimum split loss
  colsample_bytree = c(0.8),
  # Alt örnekleme oranı (sütunlar)
  min_child_weight = c(1),
  # Minimum child weight
  subsample = c(0.8)           # Alt örnekleme oranı (satırlar)
)
          # Kontrol Parametreleri
control <- trainControl(
  method = "cv",
  # çapraz doğrulama
  number = 5,
  # Daha az katlı çapraz doğrulama hız artırır
  verboseIter = TRUE,
  # ilerlemeyi gösterir
  allowParallel = TRUE          # Paralellik etkin
)
          # Model Eğitimi ve Optimizasyon
set.seed(123)
xgb_tuned <- train(
   x = as.matrix(training_data %>% select(-active_power)),
   y = training_data$active_power,
   method = "xgbTree",
   # XGBoost algoritması
   trControl = control,
   tuneGrid = grid,
   metric = "RMSE"
 )
-Aggregating results
Fitting final model on full training set
          # Paralel işlemciler kapatılır.
```

```
stopCluster(cl)
registerDoSEQ()
        # En iyi modelin hiperparametreleri
cat("En iyi hiperparametreler:\n")
-En iyi hiperparametreler:
print(xgb_tuned$bestTune)
  nrounds max_depth eta gamma colsample_bytree min_child_weight subsample
     50         6   0.1     0              0.8                1       0.8
        # Test verisi üzerindeki performans
predictions <- predict(xgb_tuned, newdata = as.matrix(test_data %>% select(-active_power)))
(xgb_rmse <- RMSE(predictions, test_data$active_power))
[1] 381.7517
(xgb_r2 <- R2(predictions, test_data$active_power))
[1] 0.9150394
results_df <- rbind(results_df,
                    data.frame(Model = "XGBoost", RMSE = xgb_rmse, R2 = xgb_r2))
print(results_df)
         Model                RMSE        R2
1 Dogrusal regresyon     404.4346  0.9046027
2 Karar agaci                464.4050  0.8742223
3 Rastgele orman          410.7135  0.9018060
4 XGBoost                    381.7517  0.9150394
        # Modellerin performans karşılaştırması için görselleştirme
ggplot(results_df, aes(x = Model, y = RMSE, fill = Model)) +
  geom_bar(stat = "identity", color = "black") +
  labs(title = "Model performans karsilastirmasi (RMSE)", x = "Model", y = "RMSE") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set3")
```

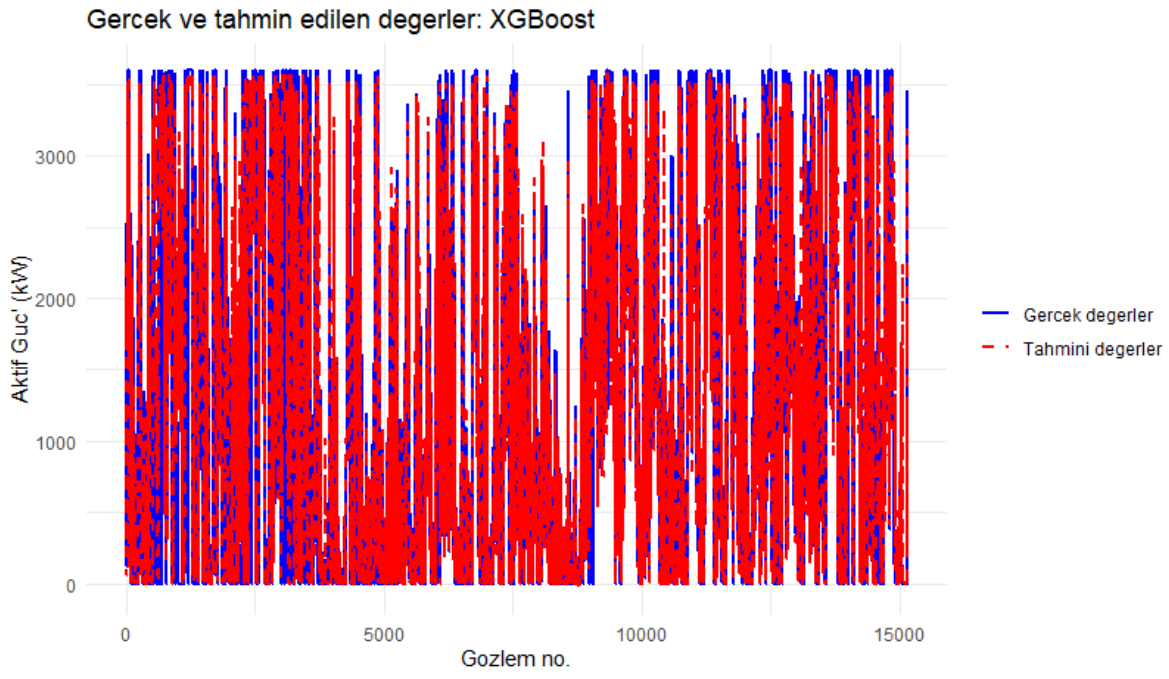



```
ggplot(results_df, aes(x = Model, y = R2, fill = Model)) +
  geom_bar(stat = "identity", color = "black") +
  labs(title = "Model performans karsilastirmasi (R2)", x = "Model", y = "R2") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set3")
```



### 5.5.1. XGBoost Yöntemi için Gerçek ile Tahmin Değerleri Grafiği

```
# XGBoost Modeli için Gerçek ve Tahmin Değerleri Görselleştirme
results_xgb <- data.frame(
  Index = 1:length(test_data$active_power),
  Gercek = test_data$active_power,
  Tahmin = predictions
)
ggplot(results_xgb, aes(x = Index)) +
  geom_line(aes(y = Gercek, color = "Gercek degerler"), linewidth = 1) +
  geom_line(aes(y = Tahmin, color = "Tahmini degerler"),
            linewidth = 1,
            linetype = "dashed") +
  labs(title = "Gercek ve tahmin edilen degerler: XGBoost", x = "Gozlem no.", y = "Aktif Guc'
(kW)") +
  scale_color_manual(values = c(
    "Gercek degerler" = "blue",
    "Tahmini degerler" = "red"
  )) +
  theme_minimal() +
  theme(legend.title = element_blank())
```



Eğitim veri seti ile oluşturulan modellerin test verileri üzerindeki tahmin performansları sonucu oluşan hata tahmin değerleri ve model performans değerleri aşağıdaki gibi sonuçlanmıştır.

| **MODEL** | **RMSE** | $(\mathbf{R^2})$ |
|---|---|---|
| ÇOKLU DOĞRUSAL REGRESYON | 404,4346 | 0.9046 |
| RASTGELE ORMAN | 410,7135 | 0.9018 |
| KARAR AĞACI | 464,4050 | 0.8742 |
| XGBOOST | 381,7517 | 0.9150 |

**Tablo 4 :**Uygulanan modellerin hata değerleri sonuç tablosu.

Aynı veri seti ele alınarak daha önce yapılmış çalışmalarda; 3497 adet aktif güç verisi sıfır ve negatif değerler aldığı için aykırı değer olarak değerlendirilmiş ve modelleme öncesi veri setinden çıkarılmıştır.

Singh vd. (2021) araştırmalarında aynı veri setini aykırı değerleri çıkararak ele almış ve yöntem olarak rastgele orman ve karar ağacı yöntemleri ile rüzgâr gücü üretim tahmini modelleri kurarak sonuçta; Rastgele Orman modeli için RMSE= 243,176 ve $R^2$= 0.9651 ve Karar Ağacı modeli için RMSE = 319,893 ve $R^2$= 0.9497 değerleri bulunmuştur (Singh vd., 2021).

Singh ve Rizwan (2022) araştırmalarında ele aldığımız veri seti ile benzer aykırı değerleri çıkarma işlemi sonrasında Rastgele Orman ve XGBoost yöntemleri ile rüzgâr gücü üretim tahmini modelleri kurarak sonuçta; Rastgele Orman modeli için RMSE= 242,815 ve $R^2$= 0.9550 ve XGBoost modeli için RMSE = 238,834 ve $R^2$= 0.9650 değerleri bulunmuştur (Singh & Rizwan, 2022).

Singh ve Rizwan (2022) bir başka çalışmalarında aynı işlemleri tekrarlayarak modelin hiperparametre ayarlanması ile tahmin için XGBoost modeli kurarak RMSE = 231,597 ve $R^2$= 0.9690 değerlerini elde etmişlerdir (Singh & Rizwan, 2022).

Görüldüğü üzere aynı veri seti ele alınarak sıfır ve negatif güç değerleri veri setinden aykırı değer olarak çıkarılmış ve aynı makine öğrenmesi yöntemleri ile elde edilen sonuçlar neticesinde XGBoost yöntemi en düşük RMSE = 231,597 ve en yüksek $R^2$= 0.9690 değerini vermiştir.

# SONUÇ

Rüzgâr Gücü Üretim tahmininin önemi rüzgâr türbininde üretilen elektrik enerjisinin tüketiminde, rüzgâr türbininin bakım ve işletme planlama çalışmalarında anlaşılmaktadır. Bu bakımdan hem kısa vadede hem de uzun vadede rüzgâr gücü üretiminin doğru tahmin edilmesi oldukça önemlidir.

Rüzgâr enerjisinin değişken yapısı nedeniyle enerji sistemlerindeki belirsizliğin azaltılması ve işletmecilerin gelirlerini arttırabilmesi için, rüzgâr santrallerindeki enerji üretimi tahminlerinin daha güvenilir olması için yüksek performanslı tahmin modellerine ihtiyaç vardır.

Bu çalışmada, rüzgâr gücü üretim tahmini için dört farklı regresyon modeli değerlendirilmiştir. Bunlar Çoklu Doğrusal Regresyon, Rastgele Orman, Karar Ağacı ve XGBoost. Modellerin performansını değerlendirmek için RMSE ve $R^2$ hata değerleri ele alınmıştır. Bu hata değerleri, modelin tahminlerinin doğruluğunu ve bağımlı değişkenin (aktif güç) bağımsız değişkenlerle ne kadar iyi açıklandığını göstermektedir.

Çoklu Doğrusal Regresyon modeli, RMSE = 404,4346 ve $R^2$ = 0.9046 değerleriyle oldukça iyi bir performans sergilemiştir. Ancak çoklu doğrusal regresyonun varsayımlarının kontrolü sırasında uygulanan testler sonucunda modelin güvenilirliğinin düşük olduğu ve değişkenler arasında doğrusal olmayan ilişkiler olduğu tespit edilmiştir. Bu nedenlerden dolayı çoklu doğrusal regresyon yöntemine göre diğer ele aldığımız makine öğrenmesi yöntemlerinin kullanılması oluşturulacak modellerin tahminler üzerindeki performansının daha iyi ve karmaşık veri yapılarına karşı daha güvenilir olabileceğini göstermiştir.

Çoklu Doğrusal Regresyon modelinin iyileştirilmesi ve daha doğru tahminler yapılabilmesi için gelecek çalışmalarda bazı yöntemler önerilebilir. Örneğin, Temel Bileşenler Analizi gibi bir yöntem kullanılarak bağımsız değişkenler arasındaki ilişkiler incelenebilir. Bu analiz, birbirleriyle yüksek korelasyona sahip olan bağımsız değişkenlerin, daha az sayıda ancak daha anlamlı yeni değişkenler (temel bileşenler) olarak yeniden ifade edilmesini sağlar.

Bu yeni oluşturulan bileşenler, orijinal bağımsız değişkenlerin kombinasyonlarından oluşur ve genellikle değişkenler arasındaki çoklu korelasyon sorununu ortadan kaldırarak modelin daha doğru tahminler yapmasına yardımcı olur. Böylece, modelin karmaşıklığı azalırken, tahmin gücü ve genel performansı artabilir.

Sonuç olarak, bu tür bir yaklaşım, gelecekteki çalışmalarda ele alınarak mevcut modelin geliştirilmesine ve farklı modelleme yöntemleriyle karşılaştırmalı analizlerin yapılmasına olanak sağlayabilir. Bu tür bir katkı hem teorik hem de uygulamalı çalışmalar için değerli olabilir.

Rastgele Orman modelinde RMSE = 410,7135 ve $R^2$ = 0.9018 değerleri elde edilmiştir. Rastgele Orman modelinin bu yüksek performansı, modelin doğrusal olmayan ilişkileri yakalama kapasitesinin bir göstergesidir.

Karar Ağacı modelinde RMSE = 464,4050 ve $R^2$ = 0.8742 değerleriyle en düşük model performansı sergilemiştir. Bu modelin daha yüksek bir RMSE değeri ve daha düşük $R^2$ değeri, tek bir karar ağacının karmaşık ilişki ve desenleri yakalamakta yetersiz kaldığını göstermektedir. Karar ağacı modeli, model basitliği ve kolay yorumlanması açısından avantajlı olsa da tahmin doğruluğu bakımından diğer yöntemlere kıyasla daha düşük kalmıştır.

XGBoost modelinde RMSE = 381,7517 ve $R^2$ = 0.9150 değerleriyle en yüksek model performansı göstermiştir. XGBoost'un en düşük RMSE değerine ve en yüksek $R^2$ değerine sahip olması, bu modelin rüzgâr gücü üretim tahminlerinde diğer ele alınan yöntemlere göre daha doğru ve güvenilir sonuçlar elde ettiğini ortaya koymaktadır. XGBoost' un bu başarısı, modelin karmaşık veri yapılarını öğrenme kapasitesinden ve aşırı uyumu kontrol etme yeteneğinden kaynaklanmaktadır.

Elde edilen sonuçlar, rüzgâr gücü üretim tahminlerinde makine öğrenmesi yöntemlerinin etkinliğini ortaya koymaktadır. XGBoost, tahminlerde en düşük hata oranı ile en başarılı model olarak öne çıkmıştır.

Rüzgâr gücü üretim tahminlerinde makine öğrenmesi tekniklerinin, özellikle XGBoost gibi güçlü modellerin kullanımının rüzgâr enerjisi üretimi tahminlerinin doğruluğunu artırabileceğini, enerji ve bakım-işletme planlaması süreçlerine önemli katkılar sağlayabileceğini göstermektedir.

Gelecekteki çalışmalar, model performansını daha da artırmak için özellik seçimi, özellik oluşturma, özellik dönüştürme, eksik değerlerin işlenmesi, modeldeki hiperparametrelerin optimizasyonu ve daha büyük veri setleri seçilmesi gibi farklı stratejilerle genişletilebilir.

# KAYNAKÇA


Ahlström, A. : 2002 — “Simulating Dynamical Behaviour of Wind Power Structures” , **Lisans Tezi,** Kungliga Tekniska Högskolan Üniversitesi Teknoloji Enstitüsü.

Ahmadi, A., Nabipour, M., Mohammadi-Ivatloo B., Amani, A. M., Rho, S., Piran, M. J. : 2020 — “Long-term wind power forecasting using tree-based learning algorithms”, **IEEE Access**, 8, 151511-151522. https://ieeexplore.ieee.org/abstract/document/9169867/

Alkesaiberi, A., Harrou, F., Sun, Y. : 2022 — “Efficient Wind Power Prediction Using Machine Learning Methods:A Comparative Study”, **Energies**, 15(7), 2327. https://doi.org/10.3390/en15072327

Breiman, L. : 1996 — “Bagging predictors” , **Machine Learning**, 24(2), 123-140. https://doi.org/10.1007/BF00058655

Breiman, L. : 2001 — “Random forests” , **Machine Learning**, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Chandran, V., Patil, C. K., Merline Manoharan, A., Ghosh, A., Sumithra, M. G., Karthick, A., Arun, K. : 2021 — “Wind power forecasting based on time series model using deep machine learning algorithms” , **Materials Today: Proceedings**, 47(1), 115-126. https://doi.org/10.1016/j.matpr.2021.03.728

Chatterjee, Samprit., Hadi, A. S. : 2012 — “Regression Analysis by Example Fifth Edition” , **Wiley**.

Chen, T., Guestrin, C. : 2016 — “XGBoost: A scalable tree boosting system” , **Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, 785-794.

Chopra, S., Meindl, P. : 2007 — “Supply Chain Management Strategy, Planning & Operation”, **Das Summa Summarum des Management**, 265-275.

Delgado, I., Fahim, M.: 2021 "Wind turbine data analysis and lstm-based prediction in scada system", **Energies**, 14(1).

Demir, S., Sahin, E. K. : 2023 "An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost" , **Neural Computing and Applications**, 35(4), 3173-3190. https://doi.org/10.1007/S00521-022-07856-4/METRICS

Dukpa, A., Duggal, I., Venkatesh, B., Chang, L. : 2010 "Optimal participation and risk mitigation of wind generators in an electricity market" , **IET Renewable Power Generation**, 4(2), 165-175. https://doi.org/10.1049/IET-RPG.2009.0016

Elibüyük, U., Üçgül, İ. : 2014 "Rüzgâr türbinleri, çeşitleri ve rüzgâr enerjisi depolama yöntemleri" , **Yekarum**, 2(3). https://dergipark.org.tr/en/pub/yekarum/issue/21891/235308

Erisen, B. : 2018 "Wind Turbine Scada Dataset" https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset

Fahim, M., Sharma, V., Cao, T. V., Canberk, B., Duong, T. Q. : 2022 "Machine Learning-Based Digital Twin for Predictive Modeling in Wind Turbines" , **IEEE Access**, 10, 14184-14194. https://doi.org/10.1109/ACCESS.2022.3147602

González-Sopeña, J. M., Pakrashi, V., Ghosh, B. : 2021 "An overview of performance evaluation metrics for short-term statistical wind power forecasting" , **Renewable and Sustainable Energy Reviews**, 138, 110515. https://doi.org/10.1016/J.RSER.2020.110515

Hanifi, S., Liu, X., Lin, Z., Lotfian, S. : 2020 "A Critical Review of Wind Power Forecasting Methods-Past, Present and Future" , **Energies**, 13(15), 3764. https://doi.org/10.3390/en13153764

Heizer J., Render B., Munson C. : 2020 "Operations management : sustainability and supply chain management", **Pearson.**

James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J. : 2023 "An Introduction to Statistical Learning" , **Springer**. https://doi.org/10.1007/978-3-031-38747-0

Jankowski, N., Duch, W., Grąbczewski, K. : 2011 "Meta-Learning in Computational Intelligence" , **Studies in Computational Intelligence**, 358. https://doi.org/10.1007/978-3-642-20980-2

Karık, F., Sözen, A., İzgeç, M.: 2017 "Rüzgâr Gücü Tahminlerinin Önemi:Türkiye Elektrik Piyasasında Bir Uygulama" , **Politeknik Dergisi**, 20(4), 851-861. https://doi.org/10.2339/POLITEKNIK.369038

Kuhn, M., Silge, J. : 2022 "Tidy modeling with R" , **O'Reilly Media**.

Lantz, B. : 2013 "Machine Learning with R" , **Packt Pub Ltd.**

Lin, W., Wang, P., Chao, K., Lin, H., Yang, Z., Lai, Y. : 2021 "Wind power forecasting with deep learning networks:Time-series forecasting" , **Applied Sciences**, 11(21), 10335.

Nasteski, V. : 2017 "An overview of the supervised machine learning methods" , **Horizons B.**, 4, 51-62. https://doi.org/10.20544/HORIZONS.B.04.1.17.P05

Patel, H. H., Prajapati, P. : 2018 "Study and Analysis of Decision Tree Based Classification Algorithms" , **International Journal of Computer Sciences and Engineering**, 6(10), 74-78. https://doi.org/10.26438/ijcse/v6i10.7478

Peng, C., Zhang, Y., Zhang, B., Song, D., Lyu, Y., Tsoi, A. : 2023 "A novel ultra-short-term wind power prediction method based on XA mechanism" , **Applied Energy**, 351, 121905. https://www.sciencedirect.com/science/article/pii/S0306261923012692

Ragheb, M., Ragheb A. M.: 2011 “Wind turbines theory-the betz equation and optimal rotor tip speed ratio” , **Fundamental and advanced topics in wind power**, books.google.com.

Roungkvist, J. S., Enevoldsen, P.:2020 “Timescale classification in wind forecasting: A review of the state-of-the-art” , **Journal of Forecasting**, 39(5), 757-768. https://doi.org/10.1002/FOR.2657

T.C. Enerji ve Tabii Kaynaklar Bakanlığı : 2022 “https://enerji.gov.tr/bilgi-merkezi-enerji-ruzgar”

Singh, U., Rizwan, M. : 2022 “SCADA system dataset exploration and machine learning based forecast for windturbines”,**ResultsinEngineering**,16,100640. https://doi.org/10.1016/j.rineng.2022.100640

Singh, U., Rizwan, M., Alaraj, M., Alsaidan I. : 2021 “A machine learning-based gradient boosting regression approach for wind power production forecasting:A step towards smart grid environments” , **Energies**, 14(16), 5196. https://www.mdpi.com/1996-1073/14/16/5196

Singh, U., Rizwan, M. : 2023 “Analysis of wind turbine dataset and machine learning based forecasting in SCADA-system” , **Journal of Ambient Intelligence and Humanized Computing**, 14(6), 8035-8044. https://doi.org/10.1007/S12652-022-03878-X

Swetapadma, A., Yadav. A. : 2016 “A novel decision tree regression-based fault distance estimation scheme for transmission lines” , **IEEE Transactions on Power Delivery**, 32(1), 234-245. https://ieeexplore.ieee.org/abstract/document/7536195/

Şenel, M. C., Koç, E. : 2015 “Dünyada ve Türkiye’de Rüzgâr Enerjisi Durumu-Genel Değerlendirme” , **Mühendis ve Makina**, 56(663), 46-56.

Tayali, H. A. : 2023 “Introduction to mathematical models in operations planning”, **CRC Press.** https://doi.org/10.1201/9781003258094

Wang, X., Guo, P., Huang, X. : 2011 “A Review of Wind Power Forecasting Models” , **Energy Procedia**, 12, 770-778. https://doi.org/10.1016/j.egypro.2011.10.103

Williams G. : 2011 “Data Mining with Rattle and R: The Art of Excavating Data for Knowledge”. **Springer**.