

T.C.  
SAKARYA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

**MAKİNE ÖĞRENME Sİ YÖNTEMLERİ  
KULLANILARAK KISA DÖNEM RÜZGAR GÜCÜ  
TAHMİNİ**

**YÜKSEK LİSANS TEZİ**

**Kübra YAZICI**

**Enstitü Anabilim Dalı : ENDÜSTRİ MÜHENDİSLİĞİ**

**Tez Danışmanı : Prof. Dr. Semra BORAN**

**Ağustos 2021**

T.C.  
SAKARYA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

**MAKİNE ÖĞRENMESİ YÖNTEMLERİ  
KULLANILARAK KISA DÖNEM RÜZGAR GÜCÜ  
TAHMİNİ**

**YÜKSEK LİSANS TEZİ**

**Kübra YAZICI**

**Enstitü Anabilim Dalı : ENDÜSTRİ MÜHENDİSLİĞİ**

**Bu tez 18.08.2021 tarihinde aşağıdaki jüri tarafından oybirliği / oyçokluğu ile kabul edilmiştir.**

**Jüri Başkanı**

**Üye**

**Üye**

## **BEYAN**

Tez içindeki tüm verilerin akademik kurallar çerçevesinde tarafımdan elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun şekilde sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunduğunu, tezde yer alan verilerin bu üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim.

Kübra YAZICI

18.08.2021

## TEŐEKKÜR

Yüksek lisans tezimin gerçekleştirilmesinde değerli bilgilerini benimle paylaşan, çalışmanın her aşamasında desteğini sunan, danışmanım Prof. Dr. Semra BORAN'a sonsuz teşekkürlerimi sunarım.

Emeklerinin karşılığını ödeyemeyeceğim başta annem Selma YAZICI olmak üzere her zaman yanımda olan tüm aileme gösterdikleri sabır, anlayış ve destekleri için sonsuz teşekkür ederim.

# İÇİNDEKİLER

TEŞEKKÜR.....	i
İÇİNDEKİLER .....	ii
SİMGELER VE KISALTMALAR LİSTESİ.....	v
ŞEKİLLER LİSTESİ.....	viii
TABLolar LİSTESİ.....	x
ÖZET .....	xi
SUMMARY.....	xii

## BÖLÜM 1.

GİRİŞ .....	1
1.1. Tezin Konusu, Amacı ve Yöntem.....	3

## BÖLÜM 2.

LİTERATÜR ÖZETİ.....	7
----------------------	---

## BÖLÜM 3.

MATERYAL VE YÖNTEM.....	21
3.1. Rüzgar Enerjisi.....	21
3.1.1. Dünyada rüzgar enerjisi.....	21
3.1.2. Türkiye’de rüzgar enerjisi.....	23
3.1.3. Rüzgar gücü tahminin önemi.....	25
3.1.4. Rüzgar tahmin çeşitleri.....	26
3.1.4.1. Zaman ölçekli sınıflandırma.....	26
3.1.4.2. Tahmin yapısına göre sınıflandırma.....	28
3.2. Makine Öğrenmesi Algoritmaları.....	29
3.2.1. Gauss süreç regresyonu.....	29

3.2.2. Destek vektör regresyonu.....	31
3.2.3. Karar ağacı.....	34
3.2.4. Uzun-kısa süreli hafıza (LSTM).....	37
3.2.5. Topluluk öğrenmesi algoritmaları.....	39
3.2.5.1. Torbalama yöntemi.....	39
3.2.5.2. İstifleme yöntemi.....	40
3.3. Optimizasyon Algoritmaları.....	41
3.3.1. Bayes optimizasyon algoritması.....	41
3.3.2. Adam optimizasyonu algoritması.....	42
3.4. Özellik Seçimi Algoritması.....	43
3.4.1. Komşuluk bileşen analizi (NCA).....	44
3.5. Eksik Veri İşleme Algoritması.....	46
3.5.1. K-NN algoritması.....	46
3.6. Performans Kriterleri.....	47

#### BÖLÜM 4.

ÖNERİLEN KISA DÖNEM RÜZGÂR GÜCÜ TAHMİN MODELLERİ.....	48
4.1. Önerilen Model 1.....	48
4.2. Önerilen Model 2.....	52

#### BÖLÜM 5.

UYGULAMA.....	55
5.1. Problem Tanımı.....	55
5.2. Önerilen Modellerin Uygulanması.....	55
5.2.1. Önerilen model 1'in uygulama adımları.....	55
5.2.1.1. Verilerin toplanması.....	55
5.2.1.2. Verilerin düzenlenmesi.....	58
5.2.1.3. Makine öğrenmesi algoritmalarının uygulanması.....	64
5.2.1.4. Makine öğrenmesi algoritmalarının karşılaştırılması.....	70
5.2.1.5. Makine öğrenmesi algoritmalarının topluluk öğrenmesi ile birleştirilmesi.....	74

5.2.1.6. Tahmin modellerinin performanslarının değerlendirilmesi.....	78
5.2.2. Önerilen model 2'nin uygulama adımları.....	80
5.2.2.1. Verilerin toplanması.....	80
5.2.2.2. Verilerin düzenlenmesi.....	80
5.2.2.3. LSTM algoritmasının uygulanması.....	83
5.2.2.4. LSTM modellerinin performanslarının değerlendirilmesi.....	85
5.2.3. Önerilen modellerin karşılaştırılması.....	90
BÖLÜM 6.	
SONUÇ VE ÖNERİLER.....	93
KAYNAKLAR.....	97
ÖZGEÇMİŞ	
.....	1055

## SİMGELER VE KISALTMALAR LİSTESİ

ABC	: Yapay arı kolonisi
ANFIS	: Uyarlamalı ağ tabanlı bulanık çıkarım sistemi
BO	: Bayes optimizasyonu algoritması
BO-DT	: Bayes optimizasyonu ile optimize edilmiş karar ağacı
BO-GPR	: Bayes optimizasyonu ile optimize edilmiş gauss süreç regresyonu
BO-SVR	: Bayes optimizasyonu ile optimize edilmiş destek vektör regresyonu
C	: Düzenleme sabiti
CART	: Regresyon ağacı
CNN	: Evrimsel sinir ağları
CS	: Guguk kuşu arama algoritması
DA	: Yusufçuk algoritması
DT	: Karar ağacı
ELM	: Aşırı öğrenme makineleri
EMD	: Ampirik mod dönüşümü
FFANN	: İleri beslemeli yapay sinir ağı
GPR	: Gauss süreç regresyonu
GRU	: Geçitli tekrarlayan ünite
GWO-BP	: Gri kurt optimizasyonu ile optimize edilmiş geri yayımlı ağ
IBFGS	: Sınırlı bellekli BFGS
ICSO	: İyileştirilmiş guguk kuşu arama algoritması
IDA	: İyileştirilmiş yusufçuk algoritması
K-NN	: K-En yakın komşu algoritması
LSTM	: Uzun kısa süreli hafıza
MAD	: Ortalama mutlak sapma



MAE	: Ortalama mutlak hata
MAPE	: Ortalama mutlak yüzde hata
MGM	: Meteoroloji genel müdürlüğü
MLR	: Çoklu lineer regresyon
MSE	: Ortalama kare hatası
NARX	: Harici girişli doğrusal olmayan otoregresif ağ
NCA	: Komşuluk bileşen analizi
NCA-LSTM	: Komşuluk bileşen analizi ile indirgenmiş LSTM
NMAE	: Normalize mutlak ortalama hata
NRMSE	: Normalize ortalama hata kareleri kökü
NOAA	: Ulusal okyanus ve atmosfer dairesi
NWP	: Sayısal hava tahmini
PCA	: Temel bileşenler analizi
RBFFNN	: Radyal temelli fonksiyon ağı
RE	: Bağlı hata
RES	: Rüzgâr enerji santrali
RİTM	: Rüzgar gücü izleme ve tahmin merkezi
RMSE	: Ortalama hata kareleri kökü
SCADA	: Merkezi denetleme kontrol ve veri toplama sistemi
SFS	: Sıralı özellik seçimi
SGD	: Stokastik gradyan inişi
SVR	: Destek vektör regresyonu
SWD	: Durağan dalgacık dönüşümü
YSA	: Yapay sinir ağı
XGBoost	: Aşırı gradyan artırma
$a_i, a_i^*$	: $i$ . örneğe ait Lagrange çarpanları
$b$	: Yanlılık
$C$	: Düzenleme sabiti
$\tilde{C}_t$	: tanh katmanı ile hücre durumunun birleşiminden oluşan vektör
$D_w$	: Ağırlıklandırılmış mesafe

$f_t$	: Unutma kapısı katmanının çıktı değeri
GW	: Gigavat
GWh	: Gigavat-saat
$h_t$	: LSTM algoritması çıktı değeri
I	: Birim matris
$l_i$	: Çıktı değeri ile regresyon modeli tahmin değeri arasındaki uyumsuzluğu ölçen kayıp fonksiyonu
$i_t$	: Giriş kapısı katmanının çıktı değeri
K	: Türbin kapasitesi
kW	: kilovat
kWh	: kilovat-saat
$P_{ij}$	: $x_j$ noktasının $x_i$ için referans noktası olarak seçilme olasılığı
$R^2$	: Determinasyon katsayısı
TWh	: Teravat-saat
$v_k$	: $x$ girdi değişkeninin en yakın komşusu
W	: Ağırlıklandırılmış özellik vektörü
$w_k$	: $k$ 'inci komşunun ağırlığı
$w_m$	: $m$ . özelliğe atanan ağırlık değeri
$y_i$	: Çıktı değeri
$\tilde{y}_i$	: Regresyon modeli tahmin değeri
$x_i$	: Girdi vektörü
$\tilde{x}_j$	: K-NN algoritması ile tamamlanan veri
$\zeta_1, \zeta_1^*$	: Yapay değişkenler
$\varepsilon$	: Epsilon
$\emptyset(x)$	: Yüksek boyutlu uzay
$\varphi(\cdot)$	: Olasılık yoğunluk fonksiyonu
$\Phi(\cdot)$	: Normal dağılımın kümülatif dağılım fonksiyonu
$\sigma_f$	: Sinyal standart sapması
$\sigma_l$	: Özellik verilerinin uzunluğu
$\lambda$	: Düzeltme parametresi
$\Sigma$	: Sigmoid katmanı

## ŞEKİLLER LİSTESİ

Şekil 1.1. 1985-2020 yılları arasında üretilen elektriğin kaynaklara göre dağılımı.....	2
Şekil 3.1. Dünya rüzgar kurulu gücü .....	22
Şekil 3.2. 1998-2020 yılları Türkiye rüzgar enerji santralleri kurulu güç kapasitesi .....	25
Şekil 3.3. Rüzgar enerjisi yıllık elektrik üretimi ve oranları .....	25
Şekil 3.4. Doğrusal olmayan regresyon fonksiyonu için $\epsilon$ tüpü.....	33
Şekil 3.5. Regresyon için basit bir karar ağacı örneği.....	35
Şekil 3.6. Tekrarlayan sinir ağı yapısı .....	37
Şekil 3.7. Tek katmanlı tekrarlayan sinir ağındaki tekrarlayan bir birimin yapısı..	38
Şekil 3.8. LSTM biriminin yapısı .....	38
Şekil 4.1. Model 1 akış şeması.....	51
Şekil 4.2. Model 2 akış şeması.....	54
Şekil 5.1. Veri seti .....	57
Şekil 5.2. Girdi parametrelerine ait kutu diyagramları.....	58
Şekil 5.3. Ele alınan türbinin teorik güç eğrisi.....	60
Şekil 5.4. Ocak ayı rüzgar güç eğrisi .....	61
Şekil 5.5. Nisan ayı rüzgar güç eğrisi .....	61
Şekil 5.6. Temmuz ayı rüzgar güç eğrisi.....	62
Şekil 5.7. Ekim ayı rüzgar güç eğrisi .....	63
Şekil 5.8. Test verisine ait üretilen güç, DT ve BO-DT rüzgar gücü tahmin verisi .....	71
Şekil 5.9. Test verisine ait üretilen güç, GPR ve BO-GPR rüzgâr gücü tahmin verisi .....	72
Şekil 5.10. Test verisine ait üretilen güç, SVR ve BO-SVR rüzgar gücü tahmin verisi .....	73
Şekil 5.11. Torbalama modeli .....	74

Şekil 5.12. İstifleme 1 modeli .....	75
Şekil 5.13. İstifleme 2 modeli .....	76
Şekil 5.14. İstifleme 3 modeli .....	77
Şekil 5.15. LSTM ağlarının eğitim ve test verisi .....	81
Şekil 5.16. NCA algoritması ile belirlenen özellik ağırlıkları.....	82
Şekil 5.17. Çalışmada kullanılan LSTM ağ mimarisi .....	84
Şekil 5.18. Test verisine ait üretilen güç, LSTM-72 saat ve NCA-LSTM-72 saat rüzgar gücü tahmin verisi .....	86
Şekil 5.19. Test verisine ait üretilen güç, LSTM-120 saat ve NCA-LSTM-120 saat rüzgar gücü tahmin verisi .....	87
Şekil 5.20. Test verisine ait üretilen güç, LSTM-240 saat ve NCA-LSTM-240 saat rüzgar gücü tahmin verisi .....	88
Şekil 5.21. Test verisine ait üretilen güç, torbalama ve NCA-LSTM-240 saat rüzgar gücü tahmin verisi.....	91

## TABLULAR LİSTESİ

Tablo 2.1. Kısa dönem rüzgar gücü tahmin çalışmaları .....	16
Tablo 3.1. Zaman ölçeğine ve uygulamalarına göre rüzgar gücü tahmini .....	27
Tablo 5.1. Aylara ait rüzgar hızı- rüzgar gücü korelasyon değerleri.....	59
Tablo 5.2. Çalışmada kullanılan algoritma ve algoritmalara ait hiper parametre değer aralıkları .....	65
Tablo 5.3. DT algoritması için Bayes optimizasyonu iterasyonları .....	65
Tablo 5.4. DT algoritması varsayılan ve optimize edilmiş hiperparametre değerleri .....	67
Tablo 5.5. GPR algoritması için Bayes optimizasyonu iterasyonları.....	67
Tablo 5.6. GPR algoritması varsayılan ve optimize edilmiş hiperparametre değerleri .....	68
Tablo 5.7. SVR algoritması için Bayes optimizasyonu iterasyonları .....	69
Tablo 5.8. SVR algoritması varsayılan ve optimize edilmiş hiperparametre değerleri .....	70
Tablo 5.9. Makine öğrenmesi algoritmalarının performans kriterleri .....	73
Tablo 5.10. Topluluk öğrenmesi modellerinin parametre değerleri .....	77
Tablo 5.11. Tahmin modellerinin performans kriterlerinin karşılaştırılması .....	79
Tablo 5.12. LSTM tahmin modellerinin performans kriterlerinin karşılaştırılması .....	89
Tablo 5.13. Torbalama ve NCA-LSTM tahmin modellerinin karşılaştırılması.....	92

## ÖZET

**Anahtar Kelimeler:** Kısa dönem rüzgar gücü tahmini, makine öğrenmesi, topluluk öğrenmesi, bayes optimizasyonu, komşuluk bileşen analizi

Rüzgar enerjisinin değişken yapısından kaynaklanan sistem belirsizliğinin azaltılması ve santrallerin daha doğru üretim tahminleri yaparak gelirlerini arttırması için daha doğru ve kararlı tahmin modellerine ihtiyaç duyulmaktadır. Çalışmada bu amaçla kısa dönem rüzgar gücünün tahmini için diğer çalışmalardan farklı olarak hem statik hem dinamik yapıda ve birbirinden farklı makine öğrenmesi algoritmalarına dayanan iki yeni hibrit model geliştirilmiştir.

Geliştirilen statik modelde her mevsimden veriyi içeren sabit bir eğitim seti kullanılmıştır. Statik modelde tahmin algoritmaları olarak kullanılan Destek vektör regresyonu (SVR), Karar ağacı (DT), Gauss süreç regresyonu (GPR)'nin optimum hiper parametre değerlerinin araştırılmasında Bayes optimizasyonu algoritmasından yararlanılmıştır. Optimize edilen makine öğrenmesi algoritmalarının tahmin performanslarını arttırmak amacıyla birden çok modelin birleştirilmesine dayanan topluluk öğrenmesi algoritmalarından torbalama ve istifleme kullanılarak topluluk öğrenmesi modelleri oluşturulmuştur. Geliştirilen dinamik modelde ise zaman serisi analizinde başarılı sonuçlar veren Uzun kısa süreli hafıza (LSTM) sinir ağlarından yararlanılmıştır. Statik modelden farklı olarak eğitim veri seti zamana bağlı değişen yapıda olup test günlerinin önceki 3 gün (72 saat), 5 gün (120 saat), 10 gününe (240 saat) ait verilerle oluşturulmuştur. Geçmiş rüzgar gücü ile beraber kullanılacak girdi parametresinin seçilmesi amacı ile Komşuluk bileşen analizi (NCA) algoritmasından yararlanılmıştır. Dinamik model olması sebebiyle her adımda ağ önceki zamana ait gerçekleşen güç verileri ile güncellenmiştir.

Geliştirilen modeller Türkiye'de bulunan bir rüzgar türbinin rüzgar gücünün mevsim bazlı tahmininde 24 saatlik zaman ufku için uygulanmıştır. Statik yapıdaki torbalama modeli %11,045 Normalize Ortalama Hata Kareleri Kökü (NRMSE), %4,880 Normalize Mutlak Ortalama Hata (NMAE) ve 0,899 Determinasyon katsayısı ( $R^2$ ) değerlerini alarak dinamik NCA-LSTM modeline oranla daha başarılı olmuştur. Dinamik model ani değişimlerin ve dalgalanmaların yaşandığı zaman dilimlerinde daha başarılı olmasına rağmen statik yapıdaki torbalama modeli her mevsimden sayıca çok daha fazla veri ile eğitildiği için genel olarak gerçek güç değerlerine daha yakın sonuçlar vermiştir.

# SHORT-TERM WIND FORECAST USING MACHINE LEARNING METHODS

## SUMMARY

Keywords: Short term wind power prediction, machine learning, ensemble learning, Bayesian optimization, Neighborhood Component Analysis

More accurate and stable forecasting models are needed to reduce the system uncertainty caused by the variable nature of wind energy and to increase the revenues of the power plants by making more accurate production forecasts. For this purpose, two new hybrid models, both static and dynamic, and based on different machine learning algorithms, have been developed for the forecasting of short-term wind power in this study, unlike other studies.

In the developed static model, a fixed training set containing data from all seasons was used. The Bayesian optimization algorithm was used to investigate the optimum hyper parameter values of Support vector regression (SVR), Decision tree (DT), Gaussian process regression (GPR), which are used as estimation algorithms in the static model. In order to increase the prediction performance of the optimized machine learning algorithms, ensemble learning models were created using bagging and stacking from ensemble learning algorithms based on combining multiple models. Long short-term memory (LSTM) neural networks, which gives successful results in time series analysis, was used in the developed dynamic model. Unlike the static model, the training dataset has a time-varying structure and was created with data from the previous 3 days (72 hours), 5 days (120 hours), 10 days (240 hours) of the test days. Neighborhood component analysis (NCA) algorithm was used to select the input parameter to be used with the historical wind power. Since it is a dynamic model, the network is updated with the actual power data of the previous time at each step.

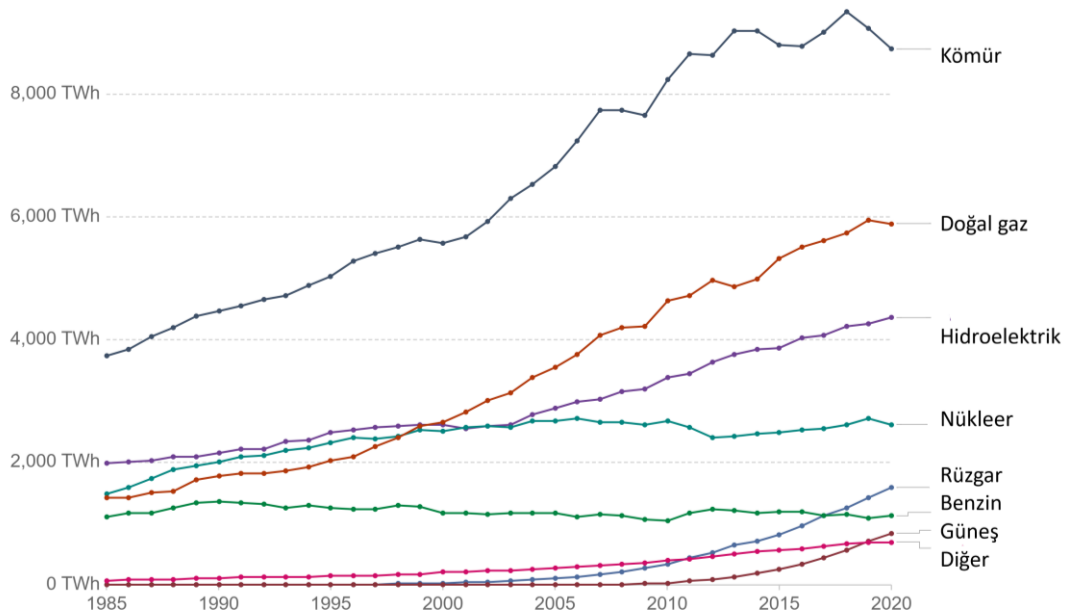
The developed models are applied for the 24-hour time horizon in the seasonal estimation of the wind power of a wind turbine in Turkey. The static bagging model was more successful than the dynamic NCA-LSTM model by taking the values of 11.045% Normalized Root of Mean Squares of Error (NRMSE), 4.880% Normalized Absolute Mean Error (NMAE) and 0.899 Coefficient of Determination ( $R^2$ ). Although the dynamic model was more successful in times of sudden changes and fluctuations, the static bagging model generally gave results closer to the real power values since it was trained with much more data than each season.

## **BÖLÜM 1. GİRİŞ**

Nüfus artışının yanı sıra son yıllarda gelişen teknoloji ve endüstrileşme nedeniyle enerjiye olan ihtiyaç gün geçtikçe artmakta ve enerji ihtiyacı günümüzün en büyük problemlerinden biri haline gelmektedir. Türkiye’de 2019 yılında elektrik enerjisi tüketimi 290,4 milyar kilovatsaate ulaşmış ve bu değer 2023 yılında yıllık ortalama %4,8 artışla 375,8 TWh'ye ulaşması beklenmektedir (ETKB, 2019). Enerji tüketimindeki artış dışa bağımlılığı artırmakta ve tükettikleri enerjiyi üretemeyen ve bu ihtiyacı ithalatla karşılayan ülkeler için cari açığa neden olmaktadır. Türkiye'nin cari açığına bakıldığında, enerji ithalatı bu açığın büyük kısmını oluşturmaktadır (Bağcı, 2019). Dünyada elektrik enerjisi tüketimi son 20 yılda %45 artarak 2019 yılında 140000 TWh'nin üzerine çıkmış ve önümüzdeki 20 yıllık dönemde de %39 artması beklenmektedir (Smil, 2017).

Şekil 1.1.'de Dünyada üretilen elektriğin 1985-2020 yılları arasında kaynaklara göre değişimi yer almaktadır.





Şekil 1.1. 1985-2020 yılları arasında üretilen elektriğin kaynaklara göre dağılımı (Our World in Data, 2021)

Son 35 yılda genel anlamda kömür ve doğal gazdan üretilen elektrikte artış görülürken nükleer enerjide büyük bir değişim görülmemiştir. Elektrik üretiminde kömür (%33) ve doğal gazın (%22) son yıllarda kullanım oranları azalmış buna karşın hidroelektrik, güneş ve rüzgar enerjisi santrallerinden sağlanan elektrik üretimi artmıştır.

Türkiye’de ise 2000 yılında 27300 MW seviyesinde olan elektrik üretim kapasitesi 2011-2019 yılları arasında her yıl ortalama 4600 MW kurulu güç eklenmesi ile 2020 yılında 95890 MW’a ulaşmıştır. 2018 yılında termik santrallerin toplam kurulu gücü yaklaşık 2000 MW azalmış, güneş ve rüzgar enerji santrallerinin toplam kurulu gücünde ise 21000 MW artış yaşanmıştır. 2020 yılının ilk dokuz ayında 1912 MW kurulu güç kapasitesine sahip yenilenebilir enerji santralleri kurulmuştur. Bu kapasitenin 1.263 MW’lık bölümünü hidroelektrik santralleri, 374 MW’lık kısmı rüzgâr enerjisi santrallerinden (RES), 237 MW’lık kısmı ise güneş enerjisi santrallerinden sağlanmıştır. Bu verilere göre yenilenebilir enerji kaynaklarının kurulu gücünün toplam elektrik santralleri kurulu gücündeki oranı %50,4 ‘e ulaşmıştır.

Yenilenebilir enerji santrallerinin kurulu gücünde artış yaşanırken doğal gaz ve çok yakıtlılar ile elektrik üreten santrallerin toplam kurulu gücünde 291 MW azalış

gerçekleşmiştir. On Birinci Kalkınma Planı'na göre 2023 yılı itibari ile Türkiye'nin toplam kurulu elektrik gücünün 109,5 GW'ye ulaşması beklenmektedir. 2020 yılının ilk dokuz ayı dikkate alınarak enerji kaynaklarının elektrik üretimindeki payları şu şekildedir: ithal ve yerli kömür santralleri (%34), hidroelektrik santralleri (%29), doğal gaz santralleri (%19), RES'ler (%8), güneş enerji santralleri (%4), jeotermal enerji santralleri (%3), biyokütle enerji santralleri (%2) (TSKB, 2020). Kaynak dağılımına bakıldığında son yıllarda kömür ve doğal gazın kullanım oranları azalsa da hem küresel anlamda hem de ülkemizde elektrik üretiminin neredeyse yarısının fosil yakıtlar olarak bilinen petrol, kömür ve doğal gazdan sağlanması ciddi bir sorundur. Enerji ihtiyacını karşılamada yoğun bir şekilde kullanılan fosil yakıt kaynaklarının sonlu olması, fiyatlarındaki dalgalanma ve doğaya verdiği zarar sebebi ile dünya hızla sürdürülebilir ve temiz enerji kaynakları olarak bilinen yenilenebilir enerji kaynaklarına yönelmektedir.

Önümüzdeki 10 yıl içinde özellikle alternatif enerji kullanımı güneş, rüzgar vb. kaynaklara bağlı olacak ve enerji üretim endüstrisindeki önemli gelişmeleri de beraberinde getirecektir (Öz ve Alyürük, 2020). Yenilenebilir enerji kaynakları içerisinde rüzgar enerjisinin yerli, sürekli, doğrudan kullanılabilir olması özelliklerinden dolayı dünyada kullanım oranı her geçen gün artmaktadır. Rüzgar enerjisi ayrıca gaz salınımını azaltma, hammadde maliyetinin olmaması ve türbinlerin uzun dönem kullanılabilmesi gibi avantajlara da sahiptir (Bayraç, 2011). Rüzgâr türbinlerinin ilk kurulum maliyetlerinin yüksek olmasına karşın hammaddeye ihtiyaç duymadan çalışabilmeleri işletme maliyetlerini düşürmektedir (Şenel ve Koç, 2015). Rüzgar enerjisi, küresel olarak yenilenebilir enerji kaynakları arasında en hızlı büyüyen ve son 6 yılda en çok yatırım yapılan enerji türüdür. 2019 yılında Avrupa'da elektrik talebinin yaklaşık %15'i, Türkiye'de yaklaşık %7'si rüzgar enerji santrallerinden sağlanmaktadır (YEKDEM, 2020).

### **1.1. Tezin Konusu, Amacı ve Yöntem**

Elektrik enerjisinin büyük ölçekte depolanamaması sebebiyle üretilen elektrik aynı zamanda tüketilmek durumundadır. Üretim ve tüketim arasındaki fark şebekeye olumsuz yansımakta bu nedenle üretim ve tüketim arasındaki farkın azaltılıp

dengelenmesine çalışılmaktadır. Özellikle rüzgar enerjisi gibi yenilenebilir enerji kaynakları ile elektrik üretimi değişken bir yapıya sahip olduğundan kontrol edilmesi geleneksel elektrik üretimine göre daha zordur.

Bu noktada klasik makine öğrenmesi yöntemlerinin yetersiz kalabildiği ve makine öğrenmesine dayalı hibrit modellerin rüzgar enerjisi üretiminin elektrik şebekelerine başarılı bir şekilde entegre edilmesinde büyük rol oynadığı görülmektedir. Hibrit tahmin modelleri ile yöntemlerinin avantajları birleştirilerek daha doğru tahmin sonuçları elde edilebilmektedir. Literatürdeki rüzgar gücü tahmin çalışmalarına bakıldığında dinamik veya statik yapıda tahmin modeli geliştirildiği görülmüş ve her iki yapıda tahmin modelinin geliştirildiği kapsamlı bir çalışmaya rastlanmamıştır. Bu nedenle çalışmada, elektrik piyasalarında doğru üretim planlaması ve doğru fiyat teklifi için sistem operatörlerine ve üreticilere yardımcı olabilecek kısa dönem rüzgar tahmini için makine öğrenme yöntemlerini esas alan statik ve dinamik yapıda iki yeni hibrit model geliştirilmiştir. Önerilen modeller Türkiye’de faaliyet gösteren bir rüzgar türbininin kısa dönem rüzgar enerjisi tahmini için kullanılmıştır. Birinci modelde test verisi Ocak, Nisan, Temmuz ve Ekim aylarına ait rastgele seçilen 15’er günlük veriler olup geri kalanı eğitim verisi kabul edilmiştir. Veri setindeki eksik veriler düşük hesaplama maliyetine sahip olan ve makine öğrenmesi yöntemlerinin veri ön işleminde sıklıkla kullanılan K-NN algoritması kullanılarak tamamlanmıştır (Gao ve ark., 2020). Burada eğitim ve test veri setlerinin her mevsimden veriyi içermesi amaçlanmış ve verilerin geçmiş değerleri ile arasındaki ilişki dikkate alınmamıştır. Uygulamada birden fazla makine öğrenme algoritmasını birleştirilerek daha yüksek performanslı modeller elde etme imkanı sağladığı için topluluk öğrenme algoritmaları tercih edilmiştir. Makine öğrenmesi yöntemlerinin parametre optimizasyonunda diğer yöntemlerden farklı olarak önceki deneylerden elde edilen bilgileri kullanan ve hızlı çalışan Bayes optimizasyonu algoritması (BO) kullanılmıştır. Burada hiper parametrelerin aldığı değerlerin model performansına olan etkisinin incelenmesi amaçlanmıştır. Makine öğrenmesi algoritmaları olarak literatürde kısa süreli rüzgar gücü tahmininde sıklıkla kullanılan ve başarılı sonuçlar veren Karar ağacı (DT), Gauss süreç regresyonu (GPR) ve Destek vektör regresyonu (SVR) kullanılmıştır. Optimize edilmiş algoritmalar arasında, Bayes optimizasyon algoritması ile optimize edilmiş

GPR (BO-GPR) ve Bayes optimizasyon algoritması ile optimize edilmiş DT (BO-DT), NRMSE, NMAE ve  $R^2$  performans kriterleri açısından daha iyi tahmin performansına sahip oldukları için topluluk öğrenmesi algoritması için temel öğrenen olarak seçildi. Topluluk öğrenmesi yöntemleri olan torbalama ve istifleme yöntemlerinde BO-GPR ve BO-DT'nin farklı kombinasyonları denenmiş ve sonuçlar Normalize Ortalama Hata Kareleri Kökü (NRMSE), Normalize Mutlak Ortalama Hata (NMAE) ve determinasyon katsayısı ( $R^2$ ) performans kriterleri 24 saatlik zaman ufkunda değerlendirilmiştir. En iyi performans kriterlerine sahip olan BO-DT ve torbalama algoritmasının birleşiminden oluşan model, %11,416 NRMSE, % 4,880 ve NMAE % 0,899  $R^2$  ile kısa dönem rüzgar enerjisinde oluşturulan statik modeller içerisinde en etkili model olduğunu göstermiştir.

İkinci modelde ilk modelde test için seçilen günlere ait gelecek 24 saatlik rüzgar gücünü tahmin etmek için eğitim verisi ve kullanılan tahmin yöntemi değiştirilmiştir. İlk modelin aksine her adımda gerçekleşen üretim değerleri modele beslenerek dinamik tahmin yaklaşımı uygulanmıştır. Bu nedenle seçilen test günleri için eğitim verileri geçmiş 3 gün (72 saat), 5 gün (120 saat) ve 10 gün (240 saat) olarak belirlenmiş ve eğitim veri setinin uzunluğunun tahmin performansı üzerindeki etkisi araştırılmıştır. Tahmin yöntemi olarak ise dinamik modelin yapısına uygun olarak zamana bağlı problemlerin modellenmesinde başarılı sonuçlar veren ve çok değişkenli veriler ile çok adımlı tahminlerde bulunabilen uzun kısa süreli hafıza (LSTM) metodu kullanılmıştır. LSTM yönteminin çıktı değeri ile gerçek değer arasındaki farkı minimuma indirmek ve öğrenme parametrelerini optimize etmek için Adam optimizasyon algoritması uygulanmıştır. Ayrıca veri setindeki girdi değişkenlerinin tahmin performansına etkisini araştırmak ve yalnızca rüzgar gücü ile güçlü ilişkiye sahip girdi değişkenlerinin seçilmesi için eğitim veri setine Komşuluk Bileşen Analizi (NCA) yöntemi uygulanmıştır. İkinci model sonuçlarına göre test verisinin önceki 10 gününe ait verilerle eğitilen ağ %16,220 NRMSE, %9,549 NMAE ve 0,796  $R^2$  ile en yüksek performansa sahip dinamik ağ olmuştur.

Tez kapsamında oluşturulan modeller ve hata kriterleri hesaplamaları MATLAB2020a programı ile yapılmıştır.

Tezin birinci bölümünde artan elektrik ihtiyacı, dünyada ve Türkiye'deki elektrik üretiminin kaynaklara göre dağılımı ve önerilen kısa dönem rüzgar gücü tahmin modellerine değinilmiştir. Bölüm 2, kısa dönem rüzgar enerjisi tahmin çalışmaları ile ilgili literatürü özetlemektedir. Bölüm 3, Dünya'daki ve Türkiye'deki rüzgar enerjisinin gelişimi, rüzgar gücü tahminin önemi, rüzgar tahmin türleri ve önerilen rüzgar tahmin modellerinde kullanılan yöntemleri anlatmaktadır. Bölüm 4, çalışmada kullanılan veri seti ve önerilen modellerin uygulama aşamalarını, belirlenen rüzgar türbini üzerindeki tahmin sonuçlarını ve önerilen modellerin karşılaştırılmasını göstermektedir. Bölüm 5'te, önerilen kısa dönem tahmin modellerinin sonuçları ve gelecek çalışmalar için öneriler yer almaktadır.

## BÖLÜM 2. LİTERATÜR ÖZETİ

Literatürde istikrarlı bir rüzgar enerjisi sağlamak amacıyla geliştirilen makine öğrenmesi ve topluluk öğrenmesi algoritmalarına dayanan birçok model bulunmaktadır.

Lee ve Baldick (2014) 48 saatlik zaman ufkunda rüzgar gücünü tahmin etmek amacıyla yapay sinir ağı (YSA) ve GPR algoritmalarını esas alan bir model geliştirmiştir. YSA alt modelleri geçmiş ve tahmini rüzgar gücü verilerinden hareketle 48 saatlik süre boyunca gelecek rüzgar gücünü tahmin eder. YSA'ya paralel olarak GPR ise sadece geçmiş rüzgar gücü verilerinden hareketle tahmin yapmaktadır. Aynı saat için oluşan birden fazla tahmin değerinden en uygun tahmin karar süreci tarafından belirlenmektedir.

Chen ve ark., (2014) kısa dönem rüzgar gücünü tahmin etmek amacıyla NWP tahmin modeli ve GPR modelleri birleştirilmiştir. Önerilen modelde gelen rüzgar hızı verileri gaussian süreçleri ile düzeltildikten sonra düzeltilen veriler arasındaki ilişkiden hareketle rüzgar gücü tahmini yapılmıştır. Modelin uygulandığı veri seti rüzgar hızı, rüzgar yönü, sıcaklık, basınç, nem ve rüzgar gücünden oluşmakta ve 10 dakikalık zaman periyodundadır. Önerilen model çok katmanlı algılayıcı (MLP), bütünleşik otoregresif hareketli ortalama (ARIMA) gibi klasik rüzgar tahmin yöntemleri ile karşılaştırılmış ve YSA'ya oranla %9 ile %14 arasında iyileşme sağlanmıştır.

Xiaoyun ve ark., (2016) 100 yükseklikte 5 dakika aralıklarla alınan rüzgar hızı, rüzgar yönü, sıcaklık, basınç ve hava yoğunluğu girdilerini içeren veri setine ait rüzgar gücünü tahmin etmek için LSTM yöntemini kullanmıştır. Oluşturulan LSTM modelinde geçmiş 1 aya ait veri eğitim, gelecek 24 saatlik veri ise test verisi olarak

kullanılmıştır. Rüzgar gücü ile güçlü ilişkiye sahip olmayan parametreler gereksiz bilgiye yol açarak tahmin performansını kötüleştirebileceği için özellik seçim yöntemlerinden olan Temel bileşenler analizi (PCA) yöntemi uygulanmış ve 5 parametreden rüzgar hızı ve rüzgar yönü girdi parametresi olarak seçilmiştir. Tahmin hatası NRMSE ve NMAE değerlerine göre Temel bileşen analizi ile indirgenmiş LSTM modeli BP'ye oranla %4 ve SVM'ye oranla %7 daha az tahmin hatasına sahiptir.

Rüzgar gücü tahmininde topluluk öğrenmesi yöntemleri uygulamalarına örnek çalışmada Heinermann ve Kramer (2016), K-En yakın komşu algoritması (K-NN), DT ve SVR kullanarak kısa süreli rüzgar gücü tahmin edilmiştir. Daha az hata ile tahmin eden, DT ve SVR yöntemleri heterojen topluluk öğrenmesi ile birleştirilerek tahmin doğruluğu arttırılmıştır.

Li ve ark., (2018) dalgacık dönüşümü ve guguk kuşu arama (CS) algoritması yöntemleri kullanılarak destek vektör makinesine (SVM) dayalı bir hibrit model önerilmiştir. Önerilen modelde eksik ve hatalı verilerle başa çıkmak için fourier, gaussian ve polinomal eğri uydurma kullanılmış ardından dalgacık dönüşümü ile orjinal sinyalde eleme yapılmıştır. CS optimizasyonu ile SVM'nin çekirdek fonksiyonu ve ceza faktörü optimize edilerek tahmin doğruluğu arttırılmıştır. Oluşturulan model Çin'de bulunan rüzgar santraline ait saatlik rüzgar hızı, rüzgar yönü ve rüzgar gücünden oluşan veri seti üzerinde uygulanmıştır. Geçmiş 60 gün (1440 saat) verisi eğitim için gelecek 3 gün (72 saat) verisi ise test için kullanılmıştır. Sonuçlara göre önerilen model ARIMA, SVR ve geri yayımlı sinir ağına oranla daha az hata değerlerine sahiptir.

Tahir ve ark., (2018) kısa dönem rüzgar gücünü tahmini için rassal orman, SVR ve radyal temelli fonksiyon ağı (RBFNN) yöntemlerinden oluşan iki katmanlı yığılmış topluluk tabanlı bir model önerilmiştir. Modelin ilk katmanında bulunan SVR, rassal orman ve RBFNN modelleri ile elde edilen çıktı değerleri ikinci katmandaki SVR ile birleştirilerek rüzgar gücü tahmin edilmektedir. Sonuçlara göre önerilen yığılmış

topluluk tabanlı model kısa dönem rüzgar gücü tahmininde klasik SVR, lineer regresyon, regresyon ağacı, rassal orman gibi yöntemlerden daha başarılı olmuştur.

Dong ve ark., (2018) tarafından yapılan çalışmada uzun tarihsel verileri taşıyabilen ve kendi kendini güncelleyen LSTM yöntemine dayalı bir kısa dönem rüzgar gücü tahmini modeli önerilmiştir. Çalışmada normalize edilen saatlik rüzgar hızı ve rüzgar gücü verileri ile 1,6,12,24 ve 48 saat gibi farklı zaman ufuklarını tahmin etmek için seçilen test verisinin geçmiş 60, 120 ve 240 saatlik verisi eğitim verisi olarak kullanılmıştır. Yapılan deneylerde zaman ufukunun uzamasıyla Ortalama Kare Hatası (MSE)'nin arttığı görülmüş ve tüm zaman ufukları için en düşük MSE değerinin geçmiş 120 saat veri ile modelin eğitilmesi ile elde edildiği görülmüştür. Model son olarak literatürde rüzgar gücü tahmin çalışmalarında sıklıkla kullanılan doğrusal regresyon, rassal orman, gradyan arttırma yöntemleri ile farklı zaman ufukları için karşılaştırılmıştır. LSTM modeli 1 saatlik zaman ufukunda lineer regresyon ile yakın MSE değerine sahip olmasına rağmen zaman ufukunun artmasıyla diğer yöntemlerden daha başarılı olmuştur.

Fu ve ark., (2019) iyileştirilmiş tavuk sürüsü optimizasyonu (ICSO) algoritması yöntemi ile optimize edilmiş SVR model saatlik rüzgar hızı, rüzgar yönü, sıcaklık ve rüzgar gücünden oluşan veri seti üzerinde uygulanmıştır. Klasik tavuk sürüsü optimizasyonu (CSO) algoritması yöntemi büyük veri setlerinde yerel optimuma kolay yakalanmasının önüne geçmek için iyileştirilmiştir. Önerilen model biri 500 diğeri 400 eğitim verisinden oluşan iki veri setinde uygulanmış ve 40 veri ile test edilmiştir. Sonuçlara göre önerilen model klasik CSO ile optimize edilmiş SVR' e göre daha iyi sonuçlar vermiş ve eğitim veri sayısının tahmin üzerinde etkili olduğu gösterilmiştir.

Ma ve Zhai (2019) dalgacık dönüşümü, ileri beslemeli yapay sinir ağı (FFANN) ve karınca koloni optimizasyon yöntemlerinden oluşan 2 aşamalı hibrit bir model önermiştir. Modelin ilk aşamasında rüzgar yönü, sıcaklık, basınç gibi meteorolojik parametreler ile rüzgar hızı tahmini yapılmıştır. İkinci aşamada ise ilk aşamadan gelen rüzgar hızı tahminleri karınca kolonisi ile optimize edilmiş YSA ile rüzgar gücü alt serileri tahmin edilmiştir. Tahmin edilen alt seriler dalgacık dönüşümü uygulanarak



rüzgar gücü tahmin sonuçları elde edilmiştir. Önerilen model gelecek 24 saat rüzgar gücü tahmininde genetik algoritmalarla optimize edilmiş YSA, parçacık sürü optimizasyonu ile optimize edilmiş YSA gibi yöntemlere oranla daha başarılı olmuştur.

Son ve ark., (2019) yaptıkları çalışmada 3 farklı bölgeye ait kısa dönem rüzgar gücünü tahmin etmek için 10 dakikalık rüzgar gücü, rüzgar hızı ve rüzgar yönü değişkenlerinin kombinasyonlarından oluşan 4 farklı veri seti ve LSTM yöntemi ile hibrit model oluşturmuşlardır. Oluşturulan veri setlerinin girdi parametreleri sırasıyla Model 1 için rüzgar gücü, Model 2 rüzgar gücü ve rüzgar yönü, Model 3 rüzgar gücü ve rüzgar hızı ve Model 4 için rüzgar gücü, rüzgar yönü, rüzgar hızıdır. Tahmin sonuçları kısa vadeli rüzgar gücü tahmininde Model 1'in ilk aşamada Model 3'ün ise orta ve geç aşamalarda diğer modellere oranla çok daha iyi performansa sahip olduğunu göstermiştir. Bunun nedeni kısa zaman ufkunda rüzgar gücünün en önemli parametre olması ve zaman ufkunun uzamasıyla sadece güç verisinin yeterli olmayışı ve meteorolojik verilere ihtiyaç duyulmasıdır. Çalışmada bu nedenle ilk aşamada Model 1, orta ve geç aşamada Model 3'ün kullanılması ile hibrit bir LSTM modeli uygulanarak tahmin hataları büyük ölçüde azaltılmıştır.

Cali ve Sharma (2019) LSTM ve duyarlılık analizi yöntemlerini kullanarak Sotevanto rüzgar çiftliği verileri ile 24 saatlik zaman ufkunda rüzgar gücü tahmini yapmışlardır. Çalışmada 1 yıla ait saatlik veri setinin ilk 9 ayı tahmin verisi kalan 3 ay ise test verisi olarak kullanılmıştır. Tahmin performansını arttırmak için gecikmeli değişken olarak tarihsel rüzgar enerjisi otokolerasyon fonksiyonu ile belirlenen geçmiş 24. saate ait rüzgar gücü verisi eklenmiştir. Daha sonra tahmin modelinin doğruluğunu arttırmak ve etkili değişkenleri tespit etmek için duyarlılık analizi uygulanmıştır. Veri setinin girdi parametrelerinin kombinasyonları ile 9 farklı tahmin modeli oluşturulmuş ve NRMSE değerleri hesaplanmıştır. Sonuçlara göre sıcaklık, rüzgar hızı (10 metre, 35 metre, 100 metre ve 170 metre yükseklikte ölçülen), rüzgar yönü (35 metre ve 170 metre yükseklikte ölçülen) parametrelerinden oluşan ikinci model %10,8 NRMSE değeri sonraki deneyler için temel model seçilmiştir. İkinci modelin 1 saat zaman

ufkunda NRMSE ve NMAE deęerleri %4,23 ve %3,01 olarak 24 saat zaman ufkunda ise %10,43 ve %7,65 olarak ölçülmüştür.

Li ve ark., (2020) kısa dönem rüzgar gücü tahmininde SVR ve iyileştirilmiş yusufçuk algoritması (IDA) yöntemlerinden oluşan bir hibrit model önerilmiştir. IDA, SVR'nin optimum parametrelerini seçmek için kullanılmıştır. Önerilen model Fransa'da bulunan bir rüzgar çiftliğinin 2017 yılı saatlik rüzgar hızı, rüzgar yönü ve rüzgar gücünden oluşan veri seti üzerinde test edilmiştir. Kısa dönem rüzgar tahmini için önerilen hibrit modelde geçmiş 6 gün verisi (144 saat) eğitim verisi, gelecek 2 gün (48 saat) verisi ise test verisi olarak belirlenmiştir. Önerilen model, kısa dönem rüzgar gücü tahmininde YSA ve GPR yöntemlerine kıyasla daha iyi sonuç vermiştir.

Banik ve ark., (2020) yükseltme, gradyan artırma ağacı ve aşırı gradyan artırma (XGBoost) yöntemlerinin tahmin edici olarak kullanıldığı çalışmada 2014 yılına ait saatlik verilerden oluşan veri setinde aykırı deęerler kaldırılmış ve rüzgar gücü ve meteorolojik parametreler arasındaki ilişki Pearson korelasyon ısı haritası ile belirlenmiştir. Sonuçlara göre rüzgar gücü ile rüzgar hızı, rüzgar yönü, sıcaklık ve nem arasında güçlü bir korelasyon bulunmuş ve dięer parametreler modele dahil edilmemiştir. Ardından modellere 5 kat doğrulama uygulanmış ve 1 saatlik ve 1 haftalık olmak üzere iki farklı zaman ufkunda tahminler yapılmıştır.  $R^2$  ve Ortalama Hata Kareleri Kökü (RMSE) ölçütlerine göre topluluk aşırı gradyan artırma modeli dięer yöntemlere oranla en iyi sonucu vermiştir.

Liu ve ark., (2021) Texas'da bulunan bir rüzgar çiftliğine ait 5 dakika zaman adımına sahip 1 yıllık veri setini kullanarak LSTM ve yusufçuk algoritması (DA)'dan oluşan bir tahmin modeli önermiştir. Önerilen modelde toplanan veriler ilk olarak 7:3 oranında eğitim ve test verisine ayrılmış ve normalize edilmiştir. Ardından DA, LSTM yönteminin öğrenme oranı, eşik deęeri ve döngü sayısı parametrelerinin optimize edilmesi için uygulanmıştır. Çalışmada kullanılan bir giriş katmanı, iki gizli katman ve bir çıkış katmanından oluşan LSTM modelinin dięer parametreleri ise Adam optimizasyonu ile optimize edilmiştir. Kısa dönem rüzgar gücü tahminindeki başarısını göstermek için DA ile optimize edilmiş LSTM

modeli (DA-LSTM), Aşırı öğrenme makineleri (ELM), Gri kurt optimizasyonu ile optimize edilmiş geri yayımlı ağ (GWO-BP) ve LSTM gibi farklı sinir ağları ile RMSE, Ortalama Mutlak Hata (MAE),  $R^2$  hata kriterleri açısından karşılaştırılmıştır. GWO-BP ve ELM modellerine kıyasla, LSTM modelinin RMSE değeri % 24,58 ve % 39,69, MAE değeri ise % 29,69 ve % 43,39 daha azdır. Ayrıca LSTM modeliyle karşılaştırıldığında, DA-LSTM modelinin RMSE değeri % 38,49 ve MAE % 34,07 azalmıştır. Deneysel sonuçlar, ELM, GWO-BP ve LSTM modellerine kıyasla DA-LSTM modelinin kısa vadeli rüzgar gücü tahmini için daha uygun olduğunu ve daha yüksek tahmin doğruluğu ve güvenilirliğine sahip olduğunu göstermektedir.

Türkiye’de yapılan rüzgar gücü tahmin çalışmalarına bakıldığında YSA yöntemi başta olmak üzere makine öğrenmesi yöntemlerinin sıklıkla kullanıldığı görülmektedir.

Köksoy (2015) bireysel tahmin yöntemlerinin tahmin performansını artırmak için yöntemlerin kombinasyonlarına dayanan Lp-norm tabanlı, ağaç tabanlı ve bulanık küme (FSS) tabanlı 12 kombinasyon modeli sunmuştur. Oluşturulan kombinasyon modellerini doğrulamak için 14 pilot RES seçilmiş ve Türkiye Rüzgar Gücü İzleme ve Tahmin Merkezi'nin (RİTM) tahmin verileri üzerinde test edilmiştir. Çalışmada 4 aylık veri seti kullanılmış ve gelecek 48 saatlik rüzgar gücü geçmiş 720 saatten hareketle gerçekleştirilmiştir. Analiz sonuçlarına göre Lp-norm tabanlı kombinasyon modelleri sekiz RES için, regresyon ağacı tabanlı kombinasyon modelleri beş RES için ve Bulanık küme tabanlı kombinasyon modelleri ise yalnızca bir RES için minimum NMAE değerini vermiştir. Çalışmanın sonucunda kombinasyon modelleri ile 14 RES genelinde en iyi bireysel yönteme kıyasla %1 oranında iyileştirme sağlanmıştır.

Aksoy (2018) makine öğrenmesi ve derin öğrenme yöntemlerini kullanarak Bursa’da bulunan bir RES’in 24 saatlik rüzgar gücünü tahmin etmiştir. Çalışmada çoklu lineer regresyon (MLR), SVR, K-NN ve DT regresyon yöntemleri ve derin öğrenme yöntemlerinden YSA, rüzgar hızı, sıcaklık ve rüzgar gücünden oluşan veri seti üzerinde uygulanmıştır. Çalışmada ilk olarak 10 dakikalık çözünürlüğe sahip

veri saatlik veriye dönüştürülmüş ardında verilerin %80'i eğitim %20'si ise test için ayrılmış ve ardından veriler normalize edilmiştir. Model performansları karşılaştırılırken MAE, MSE ve RMSE hata ölçütleri kullanılmış ve yalnızca veri setinin son 24 saati için hata değerleri hesaplanmıştır. Çalışmada sıcaklık, rüzgar hızı ve yalnızca rüzgar hızı olmak üzere iki farklı girdi seti ile 5 yöntem için seçilen bir türbinin tahmin sonuçları elde edilmiştir. Tahmin sonuçlarına göre sıcaklık ve rüzgar hızı girdilerinin kullanılması ile SVR, K-NN ve ANN yöntemleri için hata değeri daha düşük, MLR için daha yüksek ve DT için ise aynı kalmıştır. Bu nedenle girdi olarak sıcaklık ve rüzgar hızının seçilmesine karar verilmiş ve tüm türbinler için tahminler gerçekleştirilmiştir. MLR yöntemi yöntemler içerisinde en yüksek tahmin hatasına sahip olurken DT ise en düşük hataya sahip olmuştur.

Cantürk (2018) Türkiye'de bulunan bir rüzgar çiftliğine ait 10 dakika aralıklarla alınan rüzgar hızı, sıcaklık ve rüzgar gücü parametrelerinden oluşan veri setinde YSA yöntemini uygulamıştır. Çalışmada farklı özelliklere sahip her mevsimden bir gün test verisi olarak seçilmiş ve YSA yöntemi ile statik ve dinamik olmak üzere iki model uygulamasına yer verilmiştir. Statik modelde eğitim verisi olarak test verisinin bulunduğu aylara ait birer hafta seçilmiş ve FFANN kullanılmıştır. Dinamik modelde ise Harici Girişli Doğrusal Olmayan Otoregresif Ağ (NARX) ile test gününün önceki 2 gün ve önceki 1 haftasına ait veriler eğitim verisi olarak seçilerek statik modelin aksine daha az sayıda veri kullanılmıştır. Çalışmada dinamik model kabul edilebilir sonuçlar vermesine rağmen en düşük hata değeri genel olarak statik yapıdaki ağ ile sağlanmıştır. Ayrıca çalışmada sıcaklık parametresi, aktivasyon fonksiyonu ve eğitim süresinin uzunluğunun model performansına etkisi de araştırılmıştır. Sonuçlara göre sıcaklık parametresinin ve aktivasyon fonksiyonunun ağ performansı üzerinde büyük ölçüde etkisi olmadığı görülmüştür. Ayrıca dinamik modelde eğitim süresinin arttıkça tahmin performansının düştüğü görülmüştür.

Çevik (2019) yaptığı çalışmada yapay zeka yöntemleri ile bir sonraki gün, gelecek 24 saat ve gelecek 1 saat gibi üç farklı zaman dilimi için rüzgar tahmin modelleri geliştirmiştir. Bir sonraki gün üretilecek toplam gücün tahmini için rüzgar hızı,

rüzgar yönü, sıcaklık, nem, basınç, mevsim ve geçmiş 3 güne ait rüzgar gücü girdi parametrelerinden oluşan veri seti Uyarlamalı Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS) yöntemi ile Türkiye’de bulunan iki santral için tahmin yapılmıştır. Gelecek 24 saat rüzgar tahmini için önerilen iki modelin ilkinde ANFIS ve SVR yöntemleri ikincisinde ise YSA ve Karmaşık YSA yöntemleri kullanılmıştır. Modelin ilk aşamasında sayısal hava tahminlerinin indirgenmesi için korelasyon, YSA ve Yapay arı kolonisi (ABC) uygulanmış ikinci adımda ANFIS ile güç tahmini yapılmıştır. Tahmin sonuçları veri setinin alındığı santralin tahmin sonuçları ile karşılaştırılmış ve önerilen model daha az tahmin hatasına sahip olmuştur. Gelecek saatteki rüzgar gücünü tahmin etmek için Sotevonte rüzgar santraline ait saatlik çözünürlükteki 6 yıllık rüzgar yönü, rüzgar hızı ve rüzgar gücü verilerinin 4 yılı eğitim, 2 yılı test için kullanılmıştır. İlk adımda parametrelerin her birinin bir sonraki saat için YSA, SVR ve ANFIS yöntemleri ile tahmini yapılmış ve elde edilen tahmin sonuçları birleştirilerek bir sonraki saat için rüzgar gücü tahmini yapılmıştır. İlk adımda ayrıca bileşenlere ön işlem olarak Durağan dalgacık dönüşümü (SWD) ve Ampirik mod dönüşümü (EMD) uygulanmıştır. Son aşamada ise bir önceki saate ait tahmin hatası ile 0-1 arasında rastgele sayı üreterek tahmin değeri güncellenmiş ve nihai tahmin değeri elde edilmiştir. Ayrıca modelde faydalanılan yöntemlerin sonuçlarının ortalaması alınarak toplu sonuç da elde edilmiştir.

Görgel ve Kavlak (2020) yaptıkları çalışmada Evrışimsel sinir ağları (CNN) ve LSTM yöntemlerini birleştirerek hibrit bir kısa dönem rüzgar gücü tahmin yaklaşımı önermişlerdir. 1 yıla ait saatlik rüzgar hızı ve rüzgar gücü verilerinden oluşan veri seti ilk olarak normalize edilmiş ardından 300 veri eğitim kalan veri ise test için kullanılmıştır. Analiz sonuçlarına göre CNN modeli MSE değeri 12,8 olarak hesaplanmıştır. Aynı verileri tahmin etmek için 3 gizli katmandan oluşan LSTM modeli kurulmuş ve modelde 100 iterasyon 5 kez tekrar ettirildiğinde MSE değeri 8,4 çıkmıştır. Önerilen hibrit modelde CNN katmanlarına LSTM katmanı eklenerek MSE değeri 2,009’a kadar düşürülmüştür. Önerilen model 1 yıllık saatlik çözünürlükte olan veri seti ile test edilmiş ve Geçitli tekrarlayan ünite (GRU), Sarimax, Prophet, CNN ve LSTM yöntemleri ile karşılaştırılmıştır. Önerilen model

1,17 MSE deęeri ile yöntemler arasında en düşük MSE deęerine sahip yöntem olmuştur.

Literatürde oluşturulan kısa dönem rüzgar gücü tahmin modelleri ve özellikleri Tablo 2.1.'de verilmiştir.

Tablo 2.1. Kısa dönem rüzgar gücü tahmin çalışmaları

Çalışma	Girdi değişkenleri	Eğitim ve test verisi	Zaman adımı ve zaman ufku	Tahmin Yöntemi	Ön işleme	Hata Kriteri
Lee ve Baldick (2014)	Rüzgar hızı, rüzgar yönü, koordinat bilgisi	4:1	1 saat-48 saat	YSA, GPR	Normalizasyon	RMSE
Chen ve ark., (2014)	Rüzgar hızı, rüzgar yönü, sıcaklık, basınç, nem	45 gün – 30 gün	10 dakika – 24 saat	NWP, GPR	-	MAE, RMSE, NRMSE, MAPE
Köksoy (2015)	Rüzgar hızı ve rüzgar yönü	720 saat-48 saat	1 saat- 48 saat	YSA, SVM, fiziksel model, kümeleme	PCA	NMB, NMAE, NRMSE
Xiaoyun ve ark., (2016)	Rüzgar hızı, rüzgar yönü, sıcaklık, basınç ve hava yoğunluğu	1 ay- 24 saat	5 dakika -24 saat	LSTM	PCA	NMAE, NRMSE
Heinermann ve Kramer (2016)	Rüzgar hızı	18 ay- 18 ay	10 dakika-	K-NN, DT, SVR	-	MSE
Cantürk (2018)	Rüzgar gücü	28 gün- 4 gün	10 dakika- 1,3,6,24 saat	FFANN	Normalizasyon	NMAE
	Rüzgar hızı, sıcaklık, rüzgar gücü	2 gün, 1 hafta- 1,3,6,24 saat	10 dakika- 1,3,6,24 saat	NARX	Normalizasyon	NMAE

Tablo 2.1. (Devamı)

Çalışma	Girdi değişkenleri	Eğitim ve test verisi	Zaman adımı ve zaman ufku	Tahmin Yöntemi	Ön işleme	Hata Kriteri
Aksoy (2018)	Rüzgar hızı, sıcaklık	4:1	1 saat- 24 saat	MLR, SVR, DT, K-NN, YSA	Normalizasyon	MAE, RMSE, MSE
Dong ve ark., (2018)	Rüzgar hızı	60, 120, 240 saat-1, 6, 12, 24, 48 saat	1 saat- 1, 6, 12, 24, 48 saat	LSTM	Normalizasyon	MSE
Tahir ve ark., (2018)	Jeneratör yatak sıcaklığı, jeneratör hızı, rüzgar hızı, rüzgar yönü, ortam sıcaklığı	3106 veri- 1332 veri	10 dakika- 1 saat	RBFNN, SVR, rassal orman, rastgele arama	-	RRMSE, MAPE, SSE, R <sup>2</sup>
Li ve ark., (2018)	Rüzgar hızı ve rüzgar yönü	1440 saat- 72 saat	1 saat- 72 saat	SVR, CS	Dalgacık dönüşümü, normalizasyon, Fourier Dönüşümü	MAE, MSE, MAPE
Çevik (2019)	Rüzgar hızı, rüzgar yönü, sıcaklık, nem, basınç, mevsim ve geçmiş 3 güne ait rüzgar gücü	305 gün-60 gün	1 gün- 60 gün	ANFIS	-	NMAE



Tablo 2.1. (Devamı)

Çalışma	Girdi değişkenleri	Eğitim ve test verisi	Zaman adımı ve zaman ufku	Tahmin Yöntemi	Ön işleme	Hata Kriteri
Çevik (2019)	Rüzgar hızı, rüzgar yönü, sıcaklık	Her ayın son üç günü test geri kalanı eğitim verisi	1 saat- 24 saat	ANFIS, SVR	Korelasyon, ABC	NRMSE
Çevik (2019)	Rüzgar hızı, rüzgar yönü	4 yıl- 1 yıl	1 saat- 1 yıl	ANFIS, YSA, SVR	EMD, SWD	NRMSE
Cali ve Sharma (2019)	Rüzgar hızı, sıcaklık, rüzgar yönü, rüzgar gücü	9 ay- 3 ay	1 saat- 24 saat	LSTM	Duyarlılık analizi, otokorelasyon	NRMSE
Son ve ark., (2019)	Rüzgar hızı, rüzgar yönü, rüzgar gücü	1080 veri-1080 veri, 1008 veri-432 veri, 1440 veri-720 veri	10 dakika- 180 saat, 72 saat, 120 saat	LSTM	Normalizasyon	RMSE, MAPE
Ma ve Zhai (2019)	Rüzgar hızı, rüzgar yönü, sıcaklık, basınç, bağıl nem, çiy noktası sıcaklığı	2 yıl- 1 yıl	10 dakika-24 saat	FFANN, karınca kolonisi optimizasyonu	Dalgacık dönüşümü	MAPE, RMSE, MAE
Fu ve ark., (2019)	Rüzgar hızı, rüzgar yönü, sıcaklık	400, 500 saat- 40 saat	1 saat- 40 saat	SVR, ICSO	Normalizasyon	RMSE, RE
Li ve ark. (2020)	Rüzgar hızı, rüzgar yönü	144 saat- 48 saat	1 saat- 48 saat	SVR, IDA	-	NRMSE, NMAE, MAPE, R <sup>2</sup>

Tablo 2.1. (Devamı)

Çalışma	Girdi değişkenleri	Eğitim ve test verisi	Zaman adımı ve zaman ufku	Tahmin Yöntemi	Ön işleme	Hata Kriteri
Görgel ve Kavlak (2020)	Rüzgar hızı	300 gün-65 gün	1 gün-65 gün	CNN, LSTM	Normalizasyon	MSE
Banik ve ark., (2020)	Rüzgar hızı, rüzgar yönü, sıcaklık, nem	5:1	1 saat- 1gün, 1 hafta	Yükseltme, gradyan artırma ağacı, XGBoost	Pearson korelasyon	R <sup>2</sup> , RMSE
Liu ve ark., (2021)	Rüzgar gücü	7:3	5 dakika	LSTM, DA	Normalizasyon	RMSE, MAE, R <sup>2</sup>

ABC: Yapay arı kolonisi, ANFIS: Uyarlamalı Ağ Tabanlı Bulanık Çıkarım Sistemi, CNN: Evrimsel sinir ağları, CS: Guguk kuşu arama algoritması, DA: Yusufçuk algoritması, DT: Karar ağacı, EMD: Ampirik mod dönüşümü, FFANN: İleri beslemeli yapay sinir ağı, GPR: Gauss süreç regresyonu, ICSO: İyileştirilmiş guguk kuşu arama algoritması, IDA : İyileştirilmiş yusufçuk algoritması, K-NN: K-En yakın komşu algoritması, LSTM: Uzun Kısa Süreli Hafıza, MAPE: Ortalama Mutlak Yüzde Hata, MLR: Çoklu lineer regresyon, MSE: NARX: Harici Girişli Doğrusal Olmayan Otoregresif Ağ, NWP: Sayısal hava tahmini, PCA: Temel bileşenler analizi, RBFNN: Radyal temelli fonksiyon ağı, RE: Bağlı hata SVR: Destek vektör regresyonu, SWD: Durağan dalgacık dönüşümü, YSA: Yapay sinir ağı, XGBoost: Aşırı gradyan artırma

Bu tez çalışmasında rüzgar enerjisinin olasılıksal doğasından kaynaklanan elektrik piyasasındaki dengesizliklerin önüne geçmek için kısa dönem rüzgar gücünün tahmininde kullanılacak iki yeni model geliştirilmesi amaçlanmıştır. Literatürdeki kısa dönem rüzgar tahmin çalışmalarına bakıldığında hibrit yöntemlerin sıklıkla uygulandığı ve başarılı sonuçlar verdiği görülmektedir. Bu nedenle bu tez çalışmasında makine öğrenmesi yöntemlerine dayanan statik ve dinamik yapıda iki yeni hibrit model geliştirilmiştir. Kısa dönem rüzgar gücü tahmini için yapılan çalışmalar incelendiğinde genellikle tek bir tahmin yapısı üzerinde durulmuş, hem statik hem de dinamik yapıda model geliştirilen kapsamlı bir çalışmaya rastlanmamıştır. Birbirinden farklı makine öğrenmesi yöntemlerinden oluşan statik ve dinamik yapıda oluşturulan iki model Türkiye’de bulunan bir rüzgar türbinin kısa dönem rüzgar gücü tahmini için uygulanmış ve tahmin sonuçlarının farklılığına değinilmiştir.

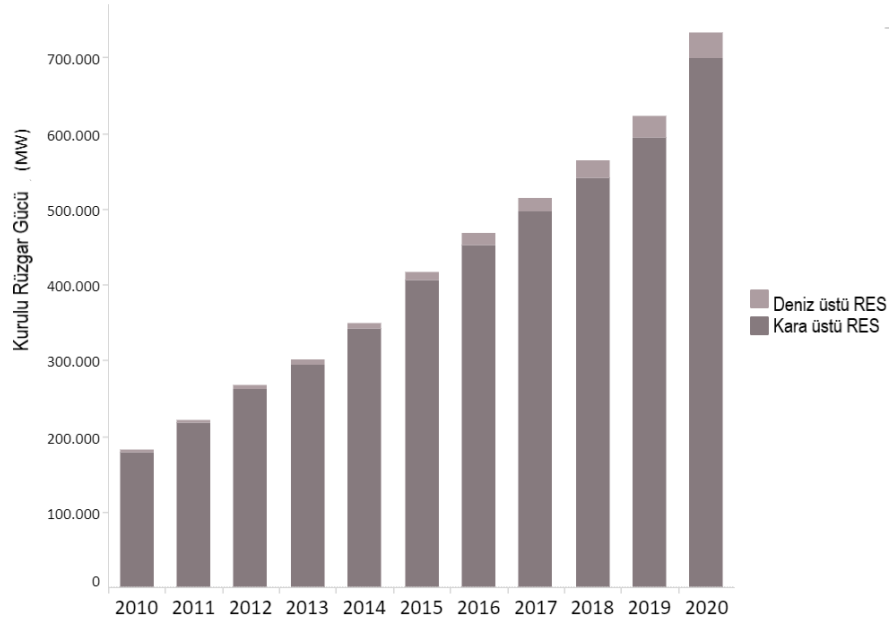
## **BÖLÜM 3. MATERYAL VE YÖNTEM**

### **3.1. Rüzgar Enerjisi**

#### **3.1.1. Dünyada rüzgar enerjisi**

Dünyadaki yenilenebilir enerji kaynakları içerisinde hidroelektrik santrallerinden sonra en fazla elektrik üretimi RES'ten sağlanmaktadır (BP, 2021). Yenilenebilir enerji kaynaklarının tüm kaynaklar içerisindeki payı %11 iken bunun yaklaşık %5' ini rüzgar enerjisi oluşturmaktadır (<https://ourworldindata.org/renewable-energy>, Erişim tarihi: 20.02.2021). Rüzgar enerjisi endüstrisi hızla gelişerek toplam kurulu kapasitesini on yılda yaklaşık beş kat arttırmıştır. 2020 yılında yeni küresel rüzgar enerjisi kurulumları 2019 yılına göre % 53 artarak 90 GW'yi geçmiştir.

Şekil 3.1.' de kara üstü ve deniz üstü RES'lerin 2010-2020 yılları için kurulu güç miktarları verilmiştir. 2020 yılı itibari ile dünyada kurulu rüzgar gücü 733276 MW olup bunun 698909 MW'yi kara üzeri RES, 34367 MW'yi ise deniz üstü RES'ler oluşturmaktadır.



Şekil 3.1. Dünya rüzgar kurulu gücü (<https://irena.org/wind>, Erişim tarihi: 30.01.2021)

2017 yılında, dünya çapında 90'dan fazla ülke ve bölge rüzgar enerjisi endüstrisinin gelişimine katılmış, 30 ülke 1 GW'yi aşan ve 9 ülke 10 GW'yi aşan kurulu kapasiteye ulaşmıştır (GWEC, 2017). 2020 yılında deniz üstü rüzgar endüstrisinde 6 GW'lik kapasiteye sahip yeni santraller kurularak büyük ölçüde büyüme sağlanmıştır. Çin, 3 GW'nin üzerinde deniz üstü rüzgar kapasitesiyle yıllık deniz üstü RES kurulumlarında dünyaya liderlik etmiştir. Geçtiğimiz yıl 1,5 GW'lik deniz üstü RES kuran Hollanda Çin'den sonra en büyük ikinci pazar haline gelmiştir.

Hollanda'nın ardından Belçika (706 MW), İngiltere (483 MW) ve Almanya (237 MW) kapasite ile 2020 yılında deniz üstü rüzgar endüstrisine en çok katkı sağlayan ülkelerdir. 2020 yılında 86,9 GW kara üstü rüzgar kapasitesi eklenerek, yıllık% 59 büyüme sağlanmıştır. 2020'deki bu büyük artış, dünyanın en büyük iki rüzgar enerjisi piyasasına sahip olan Çin ve Amerika Birleşik Devletleri'ndeki hızlı büyümeden kaynaklanmıştır. ABD kara rüzgar sektörü 2020'de bugüne kadarki en yüksek yeni kurulumu bildirmiş, yaklaşık 17 GW'lik kara üstü RES devreye alınmış ve toplam kurulu gücü 120 GW'nin üzerine çıkarmıştır.

Kara üstü ve deniz üstü rüzgar enerjisindeki yeni kurulumlar ile birlikte 2020 yılı hem karada hem de açık denizde yeni rüzgar kurulumları açısından tarihin en yüksek kurulum yapılan yılı olmuştur. Toplam kurulu güç açısından ise 2020 yılı için ilk 5 ülke değişmemiştir. Bu ülkeler şunlardır: Çin, ABD, Almanya, Hindistan ve İspanya, hepsi birlikte dünyanın toplam rüzgar enerjisi tesislerinin % 73'ünü oluşturuyor. 2020 yılı verilerine göre Türkiye'nin dünyadaki kurulu güce oranı ise %1,26'dır (GWEC, 2021).

Avrupa'da 2020 ile 2030 yılları arasında 110 GW kapasiteye sahip deniz üstü RES'lerin işletmeye gireceği tahmin edilmektedir. Bu sayede deniz üstü RES kurulu gücü 2030 yılında 150 GW'ye ulaşacak ve 315 milyon tonluk CO2 emisyonu engellenebilecektir (EWEA, 2011). Ayrıca Ekim 2019'da Pekin'de sunulan "Rüzgarın Geleceği" başlıklı bir rapora göre, 2050 yılına kadar dünyada rüzgar enerjisinin on kat büyüyeceği ve 6.000 GW'yi aşacağı düşünülmektedir. Bu sayede dünyanın elektrik ihtiyacının üçte biri rüzgar enerjisi tarafından sağlanabilecektir (IRENA, 2019).

### **3.1.2. Türkiye'de rüzgar enerjisi**

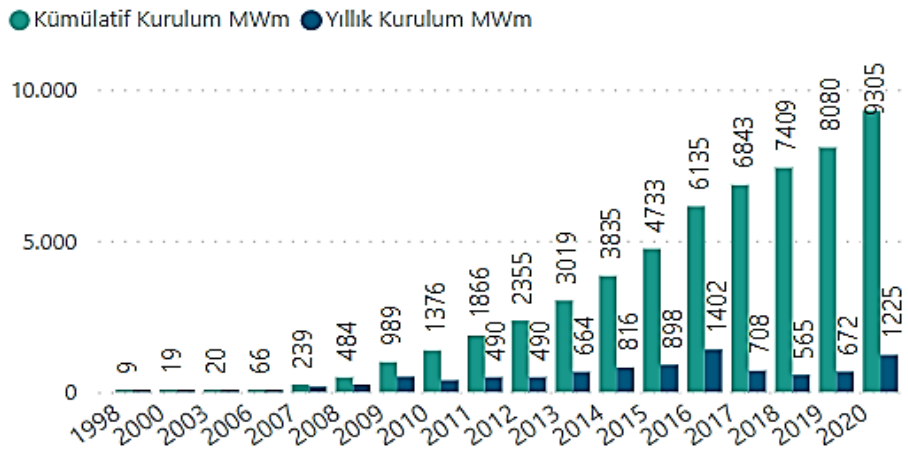
Son yıllarda yenilenebilir enerji kaynaklarına verilen önemin artması ve ülkemizin rüzgar enerjisi potansiyelinin fazla oluşu gelecekte Türkiye'nin enerji ihtiyacının karşılanmasında rüzgar enerjisinin önemli bir rol oynayacağına işaret etmektedir (Yaniktepe ve ark., 2013). Rüzgar gücü rüzgar hızının küpü ve kullanılan türbinin rotor süpürme alanı ile doğru orantılı olarak artmaktadır. Bu sebeple rüzgar potansiyeli rüzgar hızı ve sürekliliğine göre bölgeden bölgeye değişkenlik göstermektedir.

Türkiye'nin rüzgar potansiyelini ölçmek için Elektrik İşleri Etüt İdaresi tarafından 13 farklı gözlem istasyonu kurularak bir çalışma yapılmış ve 2006 yılında Türkiye Rüzgar Potansiyeli Atlası hazırlanmıştır. Meteoroloji Genel Müdürlüğü (MGM) tarafından Türkiye'de homojen olarak dağılmış 45 adet meteoroloji istasyonu ile 1989-1998 yıllarına ait saatlik rüzgar verileri ölçülerek Türkiye Rüzgar Atlası oluşturulmuştur (MGM, 2002). 50 m yükseklikte yapılan rüzgar hızı ölçümlerine göre Türkiye kara rüzgar potansiyeli, rüzgar hızı 6,5 m/s'nin üzerindeki bölgeler dikkate alındığında,

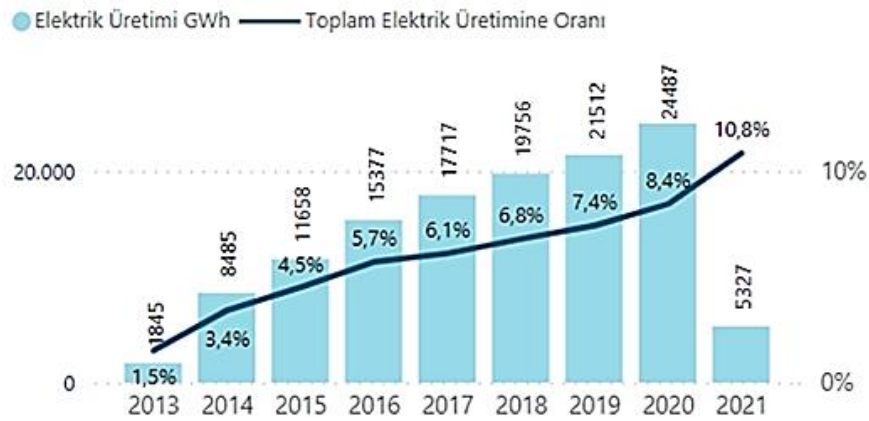
131756,40 MW; 7,0 m/s'nin üzerinde olduğu bölgeler değerlendirildiğinde ise 48000 MW olarak belirlenmiştir. Türkiye deniz rüzgar potansiyeli ise rüzgar hızının 6,5 m/s'den fazla olduğu alanlarda 17393,20 MW olarak hesaplanmıştır. Türkiye'nin yıllık ortalama rüzgâr hızı 2,58 (m/s) ve yıllık ortalama rüzgâr yoğunluğu 25,82 ( $W/m^2$ ) olarak ölçülmüştür (Ozturk ve Yuksel, 2016). Türkiye'nin yıllık ortalama rüzgar hızlarına göre en fazla rüzgar alan bölgeleri Ege Bölgesi, Marmara Bölgesi ve Güney Doğu Anadolu Bölgesidir. Yıllık ortalama rüzgar hızı Marmara, Ege ve Güney Doğu Anadolu bölgesi için sırasıyla 3,29 m/s, 2,69 m/s ve 2,65 m/s'dir. Yıllık ortalama rüzgar yoğunluğu değerleri ise aynı bölgeler için sırasıyla 51,91  $W/m^2$ , 29,33  $W/m^2$  ve 23,47  $W/m^2$ 'dir.

Türkiye Rüzgâr Enerji Birliği tarafından Ocak 2021' de yayınlanan rapora göre 2020 yılında 1224 MW kurulu güç devreye alınarak Türkiye'nin kurulu rüzgar gücü 9305 MW'a ulaşmıştır. İşletmedeki RES'lerin bölgelere göre dağılımı Ege %38,46, Marmara %34,71, Akdeniz %12,36, İç Anadolu %9,48, Karadeniz %3,69, Güneydoğu Anadolu %1,15 ve Doğu Anadolu %0,15'tir (TÜREB, 2021a). En fazla kurulu güce sahip il ise 1549, 5 MWm ile İzmir'dir. 2021 yılı itibari ile 239 santral, 3591 türbin ile 9305,02 MWm kurulu güce ve 5326,78 MWm elektrik üretimi seviyesine ulaşılmıştır. Türkiye'de halen 1.872,13 MW'lık 41 adet RES'in inşası devam etmektedir.

Şekil 3.2.' de Türkiye'de Rüzgar Enerji Santralleri için son on iki yılda (1998-2020) toplam kurulu kapasite gelişimi verilmiştir. Grafikte de görüldüğü gibi kurulu rüzgar gücü son 10 yılda 10 kat artarak 8 GW'yi aşmıştır. Son 10 yıl içerisinde en yüksek kapasite artışı 2015-2016 yıllarında gerçekleşmiş olup (4733' den 6135 MWm'e) artış oranı %29,62'dir.



Şekil 3.2. 1998-2020 yılları Türkiye rüzgar enerji santralleri kurulu güç kapasitesi (TÜREB, 2021b)



Şekil 3.3. Rüzgar enerjisi yıllık elektrik üretimi ve oranları (TÜREB, 2021b)

Şekil 3.3.'te de görüldüğü gibi 2020 yılında rüzgar enerjisinden 24487 GWh elektrik üretilmiş ve bu miktar Türkiye'de üretilen toplam elektriğin %8,44'ünü oluşturmuştur. Bu değer 2023 yılında %20 düzeyine yükseltilmesi hedeflenmektedir.

### 3.1.3. Rüzgar gücü tahminin önemi

Rüzgar enerjisiyle elektrik üretimi oldukça değişken bir profile sahiptir ve bu durum elektrik sistemlerinde belirsizliğe neden olmaktadır. Elektrik piyasalarında katılımcılardan geleceğe dair üretim ve tüketim tahminleri ve fiyat teklifleri istenmektedir. Bildirdikleri miktarda üretim yapamayan katılımcılar ürettikleri ve



tahmin deęerleri arasındaki fark ile orantılı olarak ceza ödemektedir. Tahmin hataları nedeniyle yanlış verilen teklifler ve cezalar rüzgar santrallerinin gelirlerinin yaklaşık %10'unu oluşturmaktadır. (Dukpa ve ark., 2010).

Elektrik piyasaları ile alınan üretim tahminleri, tüketim tahminleri ve üretim tahminine dair fiyat teklifleri gibi deęerler şebekedeki elektriğin fiyatını en az yapacak şekilde santrallerin çalışma programlarının oluşturulmasında kullanılmaktadır. Üretim ve tüketim dengesinin sağlanması ile elektrik sistemlerindeki belirsizlik azaltılarak elektrik fiyatının uygun hale getirilmesi ile sistem verimlilięi artırılmış olmaktadır. Bu nedenle rüzgar tahminine dayalı enerji tahmin modelleri rüzgar enerjisi kaynaklarının güvenilir, ekonomik ve kaliteli çalışmasında önemli bir yere sahiptir. Rüzgar enerjisi tahmin modelleri elektrik piyasasında güç sistemleri planlamasında, rezerv planlamasında, bakım ve onarım planlamasında ve ihalelerde kullanılmaktadır.

Tahmin modelleri sayesinde santraller, gün öncesi piyasasında tahmin hatalarını azaltarak gelirlerini artırabilmekte, böylece elektrik piyasasındaki enerji dengesizliklerini ve dolayısıyla maliyetleri azaltabilmektedir (Karık ve ark., 2017). Özellikle kısa dönem rüzgar gücü tahmini, gün öncesi elektrik ticaretinde, gün öncesi elektrik sisteminin planlanmasında, gerekli rezerv miktarının belirlenmesinde ve birim taahhüt kararlarının alınmasında önemli rol oynamaktadır (Kerem, 2018). Kısa vadeli rüzgar tahmini sayesinde, aşırı üretim planlaması ve fazla rezerv tahsisi gibi problemler önlenebilir, işletme maliyetlerinin düşürülmesi ve sisteme daha fazla rüzgar enerjisi entegre edilmesi sağlanabilir.

### **3.1.4. Rüzgar tahmin çeşitleri**

#### **3.1.4.1. Zaman ölçekli sınıflandırma**

Zaman ölçekli sınıflandırmada zaman ufku belirleyici olmaktadır. Zaman ufku gelecekte tahmin edilecek zamanın toplam uzunluęudur. Rüzgar enerjisi tahmin yöntemlerinin zaman ölçekli sınıflandırmasında farklı tanımlamalar mevcut olup Rounkvist ve Enevoldsen (2020), yaptıkları çalışmada rüzgar gücü tahmin

çalışmalarını zaman ölçeğine göre zaman ufku ve kullanım alanları açısından Tablo 3.1.'de görüldüğü gibi sınıflandırmıştır.

Tablo 3.1. Zaman ölçeğine ve uygulamalarına göre rüzgar gücü tahmini (Roungkvist ve Enevoldsen (2020))

Zaman Ölçeği	Zaman Ufku	Uygulamalar
Çok Kısa Dönem	1 saate kadar	-elektrik piyasası takas fiyatı - yük takibi -güç sistemi frekans kontrolü -gerçek zamanlı şebeke işlemleri -düzenleme eylemleri -türbin kontrolü
Kısa Dönem	1 saatten 6 saate kadar	-ekonomik yük dağıtım planlaması -yük artırma / azaltma kararları -doğru yükleme kararları -elektrik piyasasında operasyonel güvenlik -önyükleme paylaşımı -gerçek zamanlı şebeke işlemleri -düzenleme eylemleri
Orta Dönem	6 saatten 72 saate kadar	-gün öncesi elektrik piyasası -ekonomik sevkیات -elektrik ticareti -üretici çevrimiçi / çevrimdışı kararlar -doğru yükleme kararları -elde edilecek bakım planlaması -optimum işletim maliyeti -operasyon yönetimi -gün öncesi elektrik piyasasında operasyonel güvenlik -elektrik piyasasında operasyonel güvenlik -ihtiyaç rezervi kararları

Tablo 3.1. (Devamı)

Zaman Ölçeği	Zaman Ufku	Uygulamalar
Uzun Dönem	72 saatten 1 yıla kadar	-rüzgar çiftliği tasarımı için fizibilite çalışması -şebeke bakımı -bakım planlaması -operasyon yönetimi -optimum işletim maliyeti -güç sistemi planlaması -ihtiyaç rezervi -birim taahhüt kararları -rüzgar enerjisi planlaması

Literatürde rüzgar gücü tahminlerinde farklı zaman ölçeği sınıflandırmaları mevcuttur. Soman ve ark., (2010) çok kısa dönem zaman ufkunu 0-30 dakika, kısa dönem zaman ufkunu 30 dakika- 6 saat, orta dönem zaman ufkunu 6 saat-24 saat ve uzun dönem zaman ufkunu ise 1 gün- 1 hafta şeklinde tanımlamıştır. Wang ve ark., (2011) zaman ufuklarını çok kısa dönem için 0-8 saat, kısa dönem için 8-24 saat ve uzun dönem için 1 günden birkaç güne kadar olacak şekilde ele almıştır.

### 3.1.4.2. Tahmin yapısına göre sınıflandırma

Rüzgar gücü tahminleri metodolojilerine göre üç kategoride sınıflandırılmaktadır:

- İstatistiksel: İstatistiksel yöntemler, meteoroloji dikkate alınmadan büyük miktarda tarihsel veriye dayanmakta ve ölçülen güç verilerinin ilişkisini bulmayı amaçlamaktadır. İstatistiksel modellerin modellenmesi kolaydır ve diğer modellere kıyasla geliştirilmesi daha ucuzdur. Temel olarak, istatistiksel yöntem kısa dönem rüzgar gücü tahminleri için daha uygundur. Bu yöntemin dezavantajı ise tahmin süresi arttıkça tahmin hatasının artmasıdır. Genellikle yapay zeka (sinir ağları, bulanık sinir ağı) ve zaman serisi analiz yaklaşımlarını içermektedir (Garcia ve De-La-Torre-Vega, 2009; Giebel ve ark., 2011).

- Fiziksel: Fiziksel yöntem veya deterministik yöntem, sıcaklık, basınç, yüzey pürüzlülüğü gibi hava tahmini verilerini kullanan daha düşük atmosfer veya sayısal hava tahminine (NWP) dayanır. Genel olarak, yerel meteoroloji hizmetinden elde edilen ve rüzgar çiftliğinde rüzgar türbinlerine dönüştürülen rüzgar hızı rüzgar gücüne dönüştürülür (Lange ve Focken, 2008). Fiziksel yöntemler, çok sayıda hesaplama ihtiyacı duydukları için süper bilgisayarlarda oluşturulur. Fiziksel sistemler, en iyi tahmin hassasiyetine ulaşmak için atmosferin ayrıntılı bir fiziksel tanımına dayalı parametrelendirmeleri kullanır.
- Hibrit: Fiziksel yöntemleri ve istatistiksel yöntemleri birleştiren hibrit yöntem özellikle hava tahminlerini ve zaman serisi analizini kullanır. Hibrit modellerin amacı, her modelin avantajlarından yararlanmak ve optimal bir tahmin performansı elde etmektir (Wu ve Hon, 2007). Bireysel tahmin yönteminde yer alan bilgiler sınırlı olduğundan, hibrit yöntem mevcut bilgileri maksimize edebilir, bireysel model bilgilerini entegre edebilir ve çoklu tahmin yöntemlerinin avantajlarından en iyi şekilde faydalanabilir, böylece tahmin doğruluğunu artırabilir (Dongmei ve ark., 2011).

### **3.2. Makine Öğrenmesi Algoritmaları**

Çalışmada oluşturulan hibrit modellerde tahmin algoritmalarından Gauss süreç regresyonu, Karar ağacı, Destek vektör regresyonu ve Uzun kısa süreli hafıza yöntemleri kullanılmıştır.

#### **3.2.1. Gauss süreç regresyonu**

Gauss Süreç Regresyonu (GPR) temel ilkeleri Rasmussen ve Williams tarafından belirlenen çekirdek tabanlı parametrik olmayan bir olasılıklı modeldir. Gauss süreci (GP), ön olasılık dağılımını varsayarak son olasılık dağılımını tahmin eder ve eğitim

verilerine dayanarak ön olasılık dağılımını günceller (Williams ve Rasmussen, 2006).

Regresyon modeli fonksiyonu Denklem 3.1 ile ifade edilir.

$$y = f(x) + \varepsilon \quad \varepsilon \approx N(0, \sigma_n^2) \quad (3.1)$$

Denklem 3.1' de  $y$  yanıt değişkeni,  $x$  ise girdi değişkenlerdir. Toplanır gürültü  $\varepsilon$  ortalaması 0 olup normal dağıldığı kabul edilmektedir (Wan ve Sapsis, 2017). Çıktıdaki gürültü, gözlem hataları gibi  $x$  üzerindeki harici faktörlerden kaynaklanan belirsizliği ifade eder. Bir Gauss prosesi ortalama fonksiyonu  $m(x)$  ve kovaryans fonksiyonu  $k(x, x')$  ile ifade edilir (Denklem 3.2, Denklem 3.3).

$$m(x) = E[f(x)] \quad (3.2)$$

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x')))] \quad (3.3)$$

Ortalama fonksiyonu  $m(x)$  ve kovaryans fonksiyonu  $k(x, x')$ , Gauss sürecinin hiperparametreleri olarak adlandırılmaktadır (Arthur ve ark., 2020). Ortalama fonksiyonu  $m(x)$ ,  $x$  girdi değişkenine sahip  $f(x)$  fonksiyonunun beklenen değerini gösterir ve genel olarak notasyonel basitlik için sıfır olarak atanmaktadır. Kovaryans fonksiyonu  $k(x, x')$  ise ortalama fonksiyonu  $m(x)$  için güven düzeyinin bir ölçüsünü ifade etmektedir.

Model buradan hareketle Denklem 3.4 ile tanımlanır.

$$f(x) \approx GP(m(x), k(x, x')) \quad (3.4)$$

$D = \{(x_i, y_i) \mid i = 1, 2, \dots, N\}$   $x \in R^d$ ,  $y \in R$  ile Gauss modelinin eğitim veri seti temsil edildiğini varsayarsak, burada  $x$  girdi  $y$  çıktı,  $R^d$ ,  $d$  boyutlu  $R$  ise tek boyutlu vektör uzayıdır. Gauss değişkenlerinin doğrusal bir kombinasyonu olan  $D$  ile Gauss dağılımı Denklem 3.5'te verilmiştir.

$$y \approx GP(m(x), k(x, x') + \sigma_n^2 I_n) \quad (3.5)$$

Burada  $I$  birim matrisi ifade etmektedir. Kovaryans fonksiyonu Gauss süreç regresyon modelinde merkezi bileşendir bu nedenle fonksiyon seçimi oldukça önemlidir. Doğrusal, üstel, Matérn, rasyonel kuadratik, karesi alınmış üstel gibi farklı fonksiyonlardan literatürde en çok kullanılan karesel üsteldir. Bunun sebebi fonksiyonun sonsuz derecede türev alınabilmesi ve bu sebeple düzgün dağılmasıyla açıklanabilir. Bu özelliklerinden dolayı, karesi alınmış üstel çekirdek fonksiyonu, enerji tahmini ile ilgili çalışmalarda sıkça kullanılmaktadır (Heo ve Zavala, 2012). Bu çalışmada da kovaryans fonksiyonu olarak kullanılan karesi alınmış üstel fonksiyon Denklem 3.6'da gösterilmiştir.

$$k(x_i, x_j | \theta) = \sigma_f^2 \exp \left[ -\frac{1}{2} \frac{(x_i - x_j)^t (x_i - x_j)}{\sigma_l^2} \right] \quad (3.6)$$

Denklem 3.6'da  $\sigma_f$  sinyal standart sapmasını,  $\sigma_l$  ise özellik verilerinin uzunluğunu ifade eder. GPR yönteminin performansı modelin hiperparametreleri olan kovaryans fonksiyon türüne ve kovaryans fonksiyonu uzunluk ölçeği parametresine ( $\sigma_l$ ) ve sinyal standart sapmasına ( $\sigma_f$ ) bağlıdır. Bu nedenle kovaryans fonksiyon türünün ve özelliklerinin doğru seçilmesi model performansı için oldukça önemlidir.

### 3.2.2. Destek vektör regresyonu

Destek vektör makinesi yenilenebilir enerji üretimi tahmininde en sık kullanılan yöntemlerden biridir. Destek vektör makineleri veri kümesinden öğrenen ve hem regresyon hem de sınıflandırma için kullanılabilen denetimli öğrenme yöntemidir. Algoritmanın amacı eğitim verilerindeki herhangi bir noktadan maksimum derecede uzak olan iki sınıf arasında bir karar sınırı bulmaktır. Destek vektör makinesi, bir hiper düzlem kullanarak iki sınıfı en iyi şekilde ayıran bir sınırdır (Vapnik, 2013). Yöntem model karmaşıklığı üzerinde esnek bir kontrol sağlamaktadır bu sayede doğrusal olmayan problemleri çözmede küçük bir eğitim setinde bile başarılı sonuçlar vermektedir (Deng ve ark., 2018). Sadece eğitim hatasını değil aynı zamanda bir güven aralığının toplamından oluşan genelleme hatasının da üst sınırını en aza indirmeye çalışır. Regresyonda ise sınıflandırmanın aksine hiper düzlem sınırları içerisinde

maksimum veri kalacak şekilde hatanın belli bir eşik değeri içerisinde kalması sağlanarak hata minimizasyonu gerçekleştirilir.

Destek vektörü regresyonu, başlangıç verilerini daha yüksek boyutlu uzaya eşitlemek için çekirdek fonksiyonlarını kullanır, böylece doğrusal olmayan modeller doğrusal bir probleme dönüştürülebilir (Qi ve ark., 2019). Destek vektör regresyonu performansını çekirdek fonksiyonunun seçimi büyük ölçüde etkilemektedir (Zhong ve ark, 2019). Doğrusal, Polinomial, Sigmoid ve Gaussian Radyal tabanlı RBF çekirdek fonksiyonları literatürde en sık kullanılanlardandır (Yu ve ark., 2006).

$x_i$  girdi parametrelerinden oluşan vektör ve  $y_i$  çıktı değerini temsil etmektedir ( $i$  veri kümesindeki  $i$ . veri noktasını ifade etmektedir). Bu durumda örnek veri seti  $\{(x_i, y_i)\}_{i=1}^n$  olarak tanımlanabilir, burada  $n$  toplam örnek sayısını göstermektedir.

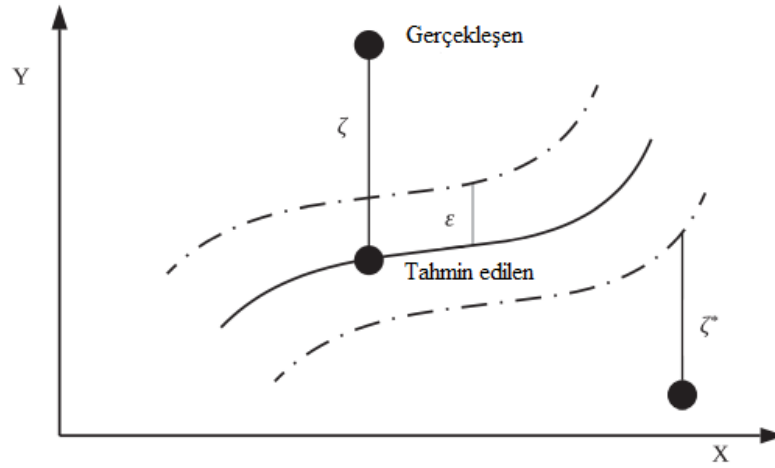
$$y = f(x) = W \cdot \phi(x) + b \quad (3.7)$$

Denklem 3.7' de  $W$  destek vektörlerinin bir kombinasyonundan oluşan ağırlıklandırılmış özellik vektörünü,  $b$  sapmayı,  $\phi(x)$  ise yüksek boyutlu uzayı ifade etmektedir. Destek vektör regresyonunda amaç eğitim verisi  $x_i$  değeri ile önceden belirlenen hata toleransından ( $\varepsilon$ ) küçük olan bir uzaklıktaki  $y_i$  tahmin değerini hesaplayabilen  $f(x)$  fonksiyonu oluşturmaktır (Li ve ark., 2009). Buna göre regresyon fonksiyonu Denklem 3.8'in minimize edilmesiyle bulunur.

$$\text{Minimum} = \frac{1}{2} |W|^2 + C \frac{1}{N} \sum_{i=1}^N L_{\varepsilon}(y_i, f(x_i)) \quad (3.8)$$

$$L_{\varepsilon}(Y_i, f(x_i)) = \begin{cases} 0, & |y_i - f(x_i)| \leq \varepsilon \\ |y_i - f(x_i)| - \varepsilon, & \text{diğer} \end{cases} \quad (3.9)$$

Denklem 3.8'in ikinci terimi ampirik hatadır ve Denklem 3.9'da verilen  $\varepsilon$  duyarsız kayıp fonksiyonu ile ölçülür. Şekil 3.4.'te doğrusal olmayan bir regresyon fonksiyonu için  $\varepsilon$  tüpü gösterilmiştir. Tahmin edilen değer  $\varepsilon$  tüpü içerisinde yer alırsa kayıp fonksiyonu sifıra eşit olmaktadır. Aksi takdirde kayıp,  $\varepsilon$  ile tahmin edilen değer arasındaki farkın büyüklüğüdür (Li ve ark., 2009).



Şekil 3.4. Doğrusal olmayan regresyon fonksiyonu için  $\epsilon$  tüpü (Dong ve ark., 2005)

Tüm veri setindeki  $x_i$  'ler için  $\epsilon$ 'den küçük uzaklıktaki  $y_i$  tahmin değerini veren  $f(x)$  bulmak mümkün olmadığı için regresyon fonksiyonu ve kısıtlara her nokta için yapay değişkenler eklenir (Denklem 3.10, Denklem 3.11).

$$\text{Minimum } \zeta_1 \zeta_1^* \quad W \quad b: \frac{1}{2} |W|^2 + C \frac{1}{N} \sum_{i=1}^N (\zeta_1 + \zeta_1^*) \quad (3.10)$$

$$\begin{aligned} y_i - W \cdot \phi x_i - b &\leq \epsilon + \zeta_1 \\ W \cdot \phi x_i + b &\leq \epsilon + \zeta_1^*, \quad i = 1, 2, \dots, N \\ \zeta_1 &\geq 0 \quad \zeta_1^* \geq 0 \end{aligned} \quad (3.11)$$

Burada  $\zeta_1$  ve  $\zeta_1^*$  yapay değişkenleri ifade etmektedir. Diğer kısıtlara ek olarak hata fonksiyonunu minimuma indirmek için Lagrange çarpanları eklenmektedir. Denklem 3.12 ve Denklem 3.13 'de  $a_i$  ve  $a_i^*$  Lagrange çarpanlarını,  $i$  ve  $j$  ise farklı örnekleri temsil etmektedir. Ayrıca doğrusal olmayan modellerin doğrusal hale getirilmesi için kullanılan çekirdek fonksiyonu  $K(x_i, x_j)$  regresyon fonksiyonuna eklenir (Denklem 3.12).

$$\begin{aligned} \text{Minimum } \{a_i\} \{a_i^*\} &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (a_i - a_i^*) \cdot (a_j - a_j^*) \cdot K(x_i, x_j) - \epsilon \sum_{i=1}^N (a_i - a_i^*) + \\ &\sum_{i=1}^N y_i (a_i - a_i^*) \end{aligned} \quad (3.12)$$

$$\begin{aligned} \sum_{i=1}^N (a_i - a_i^*) &= 0 \\ (a_i - a_i^*) &\in [0, C] \end{aligned} \quad (3.13)$$



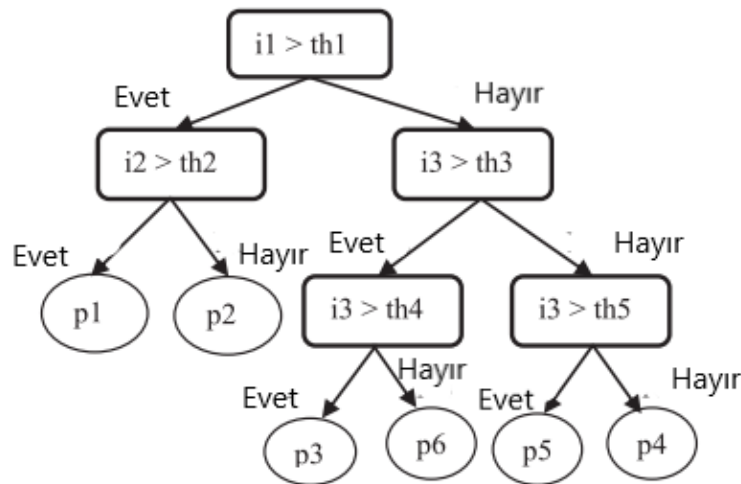
Bu durumda nihai tahmin fonksiyonu Denklem 3.14 ile ifade edilir.

$$y = f(x) = \sum_{i=1}^N (a_i - a_i^*) K(x_i, x_j) + b \quad (3.14)$$

Başarılı bir SVR modeli için hiper parametreler dikkatli bir şekilde ayarlanmalıdır, aksi takdirde aşırı öğrenme veya eksik öğrenmeye neden olabilir. Denklem 3.10 'daki kutu kısıtlaması olarak da bilinen düzenleme sabiti (C), SVR modelinin karmaşıklığını (fonksiyonun düzlüğünü) belirlemek için kullanılır. Epsilon ( $\epsilon$ ) ise regresyon fonksiyonlarının karmaşıklığını kontrol eder. Yapay değişkenler  $\zeta_1$  ve  $\zeta_1^*$ , sırasıyla epsilonun üstündeki ve altındaki mutlak sapmaları temsil eder. Düşük C değerleri düz işlemlere neden olabilirken, yüksek C değerleri aşırı öğrenmeye neden olabilir. Diğer yandan, yüksek  $\epsilon$  değeri regresyon modelini çok basit hale getirir, bu da düşük tahmin doğruluğu ile sonuçlanırken, düşük  $\epsilon$  değeri SVR modelini çok karmaşık hale getirir ve tahmin başarısını azaltır. Bu nedenle, doğru C ve  $\epsilon$  değerlerinin seçilmesi, tahmin problemlerinde önemlidir (Ito ve Nakano, 2003).

### 3.2.3. Karar ağacı

Karar ağaçları hem hesaplama maliyetlerinin düşük oluşu hem de kolay yorumlanabilir olması sebebi ile sınıflandırma ve regresyon problemlerinde sıkça kullanılmaktadır. Karar ağacı, karar verme için basit ve anlaşılır bir yapıya sahiptir. Bir karar ağacı, kök, dallar, yapraklar ve düğümlerden oluşmaktadır (Vega ve ark., 2009). Karar ağacı sürecinde, bir parametre kök veya birinci düğüm olarak seçilir ve ardından birinci düğüm, bir dizi özelliğe dayalı olarak birkaç dahili düğüme bölünür. Kök, dal ve düğümden oluşan bir zincirin ucu yaprak olarak adlandırılır (Tiryaki, 2008). Karar ağacında bölme işlemi, girdi parametrelerinden biri tarafından uygulanır ve bunların aralığı, MSE'nin en aza indirilmesine göre seçilir. Modelin çıktısı ayrı bir değerler kümesi olduğunda, sınıflandırma ağacı olarak adlandırılır ve sistemin çıktısı gerçek bir değerler kümesi olduğunda regresyon ağacı olarak adlandırılır (Lewis, 2000). Şekil 3.2. bir karar ağacının yapısını şematik olarak göstermektedir.



Şekil 3.5. Regresyon için basit bir karar ağacı örneği (Swetapadma ve Yadav, 2016)

Şekil 3.5.'te  $I = \{i_1, i_2, i_3\}$  girdi parametrelerini,  $Th = \{th_1, th_2, th_3, th_4, th_5\}$  eşik değerlerini ve  $P = \{p_1, p_2, p_3, p_4, p_5, p_6\}$  ise tahmin değerlerini temsil etmektedir. Tahminler girdi parametrelerine uygulanan Eğer - İse kurallarına göre yapılmaktadır.

Karar ağaçlarını belirlemeye yönelik tüm algoritmalar arasında, sınıflandırma ve regresyon ağacı (CART) en sık kullanılanlardandır. Kısa dönem rüzgar güç tahmini bir regresyon problemi olduğu için çalışmada regresyon ağaçlarının prensibine odaklanılmıştır.

CART algoritması, minimum kare hatasına dayalı olarak en uygun özelliklerin ve bölme noktalarının seçilmesi ile tekrarlayan ağaç oluşturma sürecidir. Ağacın veri setinin özelliklerine göre ayarlanabilme özelliği sayesinde fonksiyon yapısını önceden ayarlanmasına gerek kalmamakta, hem kesikli hem de sürekli değişkenlerle çalışılabilmektedir. Karar ağacı kategorik, sayısal ve eksik veriler ile çalışabilme, çoklu yanıt değerlerini tahmin edebilme ve basit bir yapıya sahip olması gibi avantajlara sahip olduğu gibi bazı dezavantajları da beraberinde getirmektedir (De'ath ve Fabricius, 2000). Karar ağacı regresyonu, verilerdeki küçük değişiklikler farklı bir ağaçla sonuçlanabileceğinden kararsız olabilir. Deneylerde ideal bir doğrulama için küçük değişikliklerden kaçınılmalıdır aksi takdirde aşırı öğrenmeye sebep olabilir. Bu

sorunu önlemek için karar ağacı regresyon parametrelerinin ayarlanması gerekmektedir. Karar ağacı regresyonunun, tahminin kalitesini etkileyebilecek parametrelerden bazıları uygunluk fonksiyonu, ağacın derinliği, maksimum dal düğüm sayısı, minimum yaprak örnek sayısı ve özellik (girdi parametresi) sayısı olarak sıralanabilir. Uygunluk fonksiyonu, tahmin edilen değer ile gözlemlenen değer arasındaki hatayı en aza indirmeyi amaçlar. Ağacın derinliği, ağacın ne kadar büyüyebileceğini belirler. Ağaç ne kadar derin olursa, veriler hakkında o kadar fazla bilgi edinir. Karar ağacı algoritması maksimum dal düğüm sayısına eşit ya da daha az sayıda dal düğümünü böler. Dal düğümü sayısı maksimum dal düğüm sayısından fazla olursa algoritma maksimum dal düğüm sayısına ulaşana kadar katman katman dal düğümlerini böler. Minimum yaprak örnek sayısı, yaprak düğümünde olması gereken minimum örnek sayısıdır. Yapraklardan biri bir örneğe sahip ise ve minimum yaprak örnek sayısı ikiye eşitse, bu yaprak nedeniyle ayrılma yapılmaz. Minimum yaprak örnek sayısı ve maksimum dal düğümü aşırı öğrenmeyi kontrol etmektedir. Özellik sayısı ise en iyi bölünmeyi arayan özellik sayısını ifade eder ve yüksek değerleri aşırı öğrenmeye neden olabilmektedir.

CART algoritmasında veri kümesindeki bölünmeleri ve dallardaki özelliklerin değerlendirilmesi için Gini katsayısı kullanılmaktadır (Breiman ve ark.,1984).

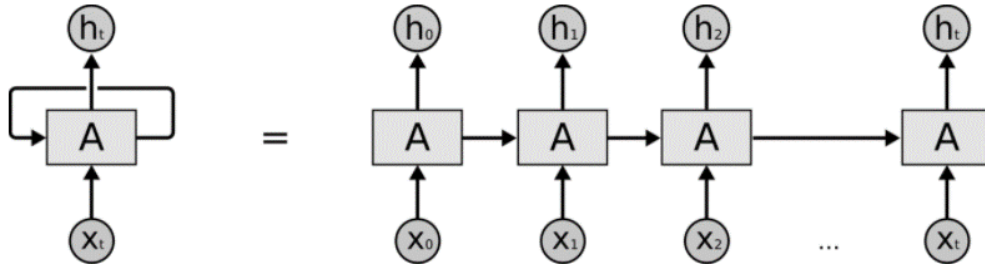
$$Gini\ katsayısı = \frac{f(C_i,t)}{|T|} \quad (3.15)$$

Denklem 3.15'te verilen gini katsayısı denkleminde T eğitim verisini  $C_i$  i. sınıfı ifade etmektedir.

Regresyon ağacı algoritması özetlenecek olursa tek bir düğümle başlar ve sonraki adımda her girdi parametresi için uygunluk fonksiyonu değeri hesaplanır. Minimum uygunluk değerini veren bölünme seçildikten sonra üçüncü adımda ikinci adıma dönülerek durdurma kriteri sağlanana kadar döngü tekrarlanır.

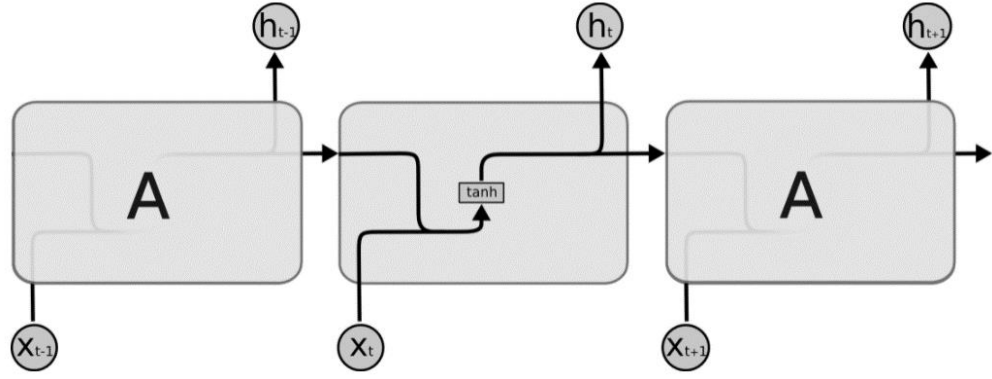
### 3.2.4. Uzun-kısa süreli hafıza (LSTM)

LSTM, Hochreiter tarafından önerilen sınırlı veri örneklerindeki uzun vadeli bağımlılıkları öğrenme yeteneğine sahip tekrarlayan sinir ağlarının bir türüdür (Hochreiter ve Schmidhuber, 1997). LSTM, zaman serileri analizinde deneyimlerden öğrenme konusunda tekrarlayan sinir ağına göre daha fazla güce sahiptir (Yu ve ark., 2019). Rüzgar enerjisi kayıtları kronolojiktir, dolayısıyla bilginin kalıcı olması beklenir bu sayede güç eğrisi daha doğru sonuçların tahmin edilmesine yardımcı olur. Bu nedenle, tekrarlayan sinir ağları rüzgar gücü tahmin modellerinde tercih edilmektedir (Cao ve ark., 2012). Tekrarlayan sinir ağlarının içindeki geri bildirim döngüleri, modelin zaman içinde bilgileri tutmasına izin vermektedir. Şekil 3.6.'daki eşitliğin işaretinin sol tarafı, bir tekrarlayan sinir ağı hücresinin yapısını göstermektedir. A bloğu içindeki verileri yinelemeli olarak kullanılır ve bir kez genişletildiğinde yapısı eşitliğin sağ tarafında gösterilen zincire eşdeğerdir.

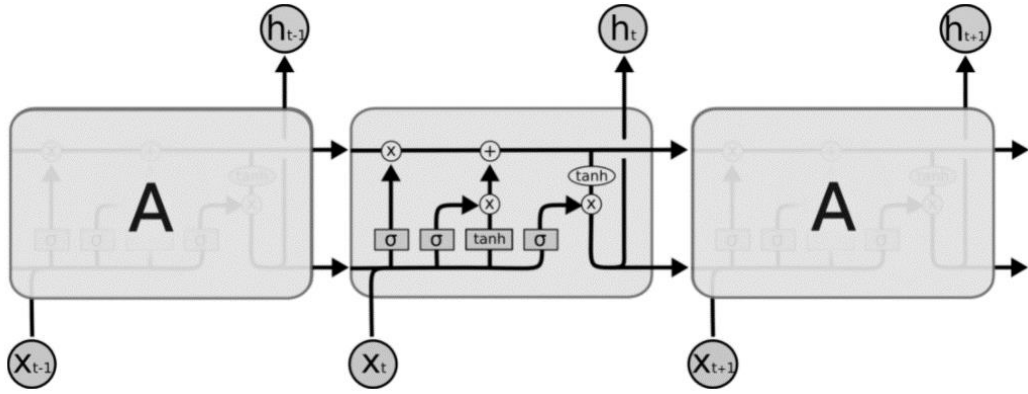


Şekil 3.6. Tekrarlayan sinir ağı yapısı (Olah, 2015)

Bununla birlikte, tekrarlayan sinir ağı modeli uzun süreli geçmişi hatırlamakta sorun yaşamakta ve veri seti çok büyük hale geldiğinde bu sorun daha açık bir şekilde görünmektedir. LSTM, tekrarlayan sinir ağının bir uzantısıdır ve uzun süreli bilgileri hatırlamak için tasarlanmış bir bellek hücresi içerir. Tekrarlayan sinir ağının her hücresinde yalnızca tek bir tanh katmanı vardır (Şekil 3.6.). LSTM modelinde, her bir analize giren ve çıkan verileri kontrol etmek ve bu bilgilerin hücre durumuna eklenip eklenmeyeceğini belirlemek için kullanılan bir dizi katman bulunmaktadır (Şekil 3.7.).



Şekil 3.7. Tek katmanlı tekrarlayan sinir ağındaki tekrarlayan bir birimin yapısı (Olah, 2015)



Şekil 3.8. LSTM biriminin yapısı (Olah, 2015)

Sonuç olarak, LSTM'nin bellek hücresi, bir sigmoid katmanını ve bir çarpandan oluşan hücre durumuna bilgi ekleme ve kaldırma yeteneğine sahip kapıları birleştirir. Sigmoid katmanı, bilgiyi 0 ile 1 arasında bir sayıya dönüştürerek bileşenlere hangi bilgilerin verileceğini açıklar. Çıktı 1 ise, tüm bilgiler geçmekte 0 ise hiçbir bilginin geçmesine izin verilmemektedir. LSTM'deki bellek hücresi üç farklı kapıdan oluşmaktadır. Bunlar bilgilerin saklanıp saklanmayacağına karar veren unutma kapısı, son hücreden güncellenen bilgileri etkinleştiren giriş kapısı ve sonraki hücreye hangi bilgilerin çıkarılacağını seçmek için bir çıkış kapısıdır. LSTM'de bütünlüğü uzun süre korumak için bilgi değişmeden kalırken bilginin zincir boyunca akmasına izin verilmektedir. Şekil 3.8.' de gösterildiği gibi  $C$  yatay olarak işleyen diyagramın üstündeki hücre durumudur. LSTM,  $h_t$  çıktı verisini  $x_t$  giriş verisine bağlı olarak aşağıdaki gibi hesaplar:

$$f_t = \sigma(W_t \cdot [h_{t-1}, x_t] + b_f) \quad (3.16)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.17)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3.18)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (3.19)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.20)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (3.21)$$

Denklem 3.16, Denklem 3.17, Denklem 3.20' de  $f_t$  unutmama kapısı katmanının,  $i_t$  giriş kapısı katmanının ve  $o_t$  çıkış kapısı katmanının çıktısını ifade etmektedir.  $W$  ağırlık matrisini,  $b$  yanlılığı,  $C$  ise hücre durumunu temsil etmektedir. Sigmoid katmanı ( $\sigma$ ), verilerin geçiş oranını kontrol edecek 0 ile 1 arasında bir sayı üretmek için kullanılır.  $\tanh$  katmanı, hücre durumuna eklenecek  $\tilde{C}_t$  vektörünü oluşturur (Denklem 3.18, Denklem 3.19). Daha sonra hücre durumu başka bir  $\tanh$  katmanından geçerek ve sigmoid katmanı ile çarpılarak sonucun çıktısı  $h_t$ 'yi vermektedir (Denklem 3.21).

### 3.2.5. Topluluk öğrenmesi algoritmaları

Çalışmada makine öğrenmesi yöntemlerinin birleştirilmesi ve tahmin doğruluğunun artırılması amacı ile topluluk öğrenmesi algoritmalarından torbalama ve istifleme algoritmaları kullanılmıştır.

#### 3.2.5.1. Torbalama yöntemi

Breiman tarafından ortaya atılan torbalama yaklaşımında eğitim setinin örneklerini ve tahmin metodlarının çıktısının ortalamasını ya da oylamasını kullanarak bağımsız tahmin ediciler oluşturulması amaçlanmaktadır. Önyükleme ile yani değiştirme ile yeniden örnekleme yapılarak birden fazla benzer veri kümesi oluşturulduğunda çıktı hatasının varyansı azalmakta ve aşırı uyum sorunu çözülmektedir (Breiman, 1996). Topluluk öğrenmesi algoritmaları birden fazla temel öğreneni birleştirerek daha iyi

genelleme yeteneğine ve tahmin performansına sahip modeller oluşturmaya olanak tanır (Dietterich, 2000). Breiman hem sınıflandırma hem de regresyon problemlerinde torbalama yönteminin tek ağaca oranla daha iyi sonuç verdiğini göstermiştir.

Torbalamada ilk adımda eğitim veri setinden yerine koyma ile homojen şekilde  $n$  boyutunda  $N$  adet yeni eğitim verisi oluşturulur. Ardından, topluluktaki her ağaç ayrı ayrı ilgili yeni eğitim verileri ile eğitilir. Son adımda, son bir tahmin yapmak için tüm tahminlerin ortalaması hesaplanır. Torbalama ağaçları modelinde tahmin şu şekilde tanımlanır:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N f_i(x) \quad (3.22)$$

Denklem 3.22’te  $f_i$  oluşturulan ağaçları,  $i$  ise önyükleme verilerini göstermekte ve oluşturulan her ağaç modeli önyükleme verileri üzerinde eğitilmektedir.

### 3.2.5.2. İstifleme yöntemi

Yığınlama olarak da bilinen istifleme yöntemi farklı tahmin modellerini tek bir modelde birleştiren farklı seviyelerden oluşan bir yöntemdir. Yaklaşım meta öğrenme kavramı ile optimal bir öğrenme sistemi ortaya koyarak genelleştiricilerin önyargısını azaltarak hataları minimize etmeyi amaçlar (Wolpert, 1992; Van der Laan ve ark., 2007).

İstifleme yaklaşımı Seviye-0 ve Seviye-1 olmak üzere iki seviyeden oluşmaktadır. Seviye-0’da farklı tahmin modelleri eğitilir ve çıktı değişkeni tahmin edilir. Seviye-0’da elde edilen tahminler Seviye-1 için girdi olarak kullanılır. Seviye-1 ‘deki model meta öğrenen olarak adlandırılır ve kendinden önceki seviyenin her modelinin en iyi tahminini veren önceki seviye modelleri ile öğrenir (Serbes ve ark., 2015; Shamaei ve Kaedi, 2016; Petropoulos ve ark., 2017). İstifleme yönteminde seviye sayısı 2 ile sınırlı değildir. Yöntemin çalışma prensibine göre  $n$  seviyeli model  $n-1$  seviyedeki modellerin tahminlerini kullanır. İstifleme yönteminde tahmin sonuçlarındaki değişim farklı seviyelerdeki modellerin çeşitlilik göstermesinden kaynaklanır. Bu durumun sebebi

farklı genelleme kurallarına sahip modellerin farklı sonuçlar üretme eğiliminde oluşudur (Mendes-Moreira ve ark., 2012).

### **3.3. Optimizasyon Algoritmaları**

Tahmin yöntemlerinin hiper parametre değerlerinin optimizasyonu ile tahmin hatasının büyük oranda azaldığı görülmüştür. Bu nedenle çalışmada önerilen ilk modelde tahmin yöntemlerinin en iyi hiper parametre değerlerinin seçilmesinde Bayes optimizasyonu algoritması kullanılmıştır. LSTM ağlarının öğrenmesinde önemli yere sahip olan ağ ağırlıklarının güncellenmesi için ise Adam optimizasyonu algoritmasından yararlanılmıştır.

#### **3.3.1. Bayes optimizasyon algoritması**

Makine öğrenmesi algoritmalarının başarısında hiper parametre seçimi önemli bir rol oynamaktadır. Izgara arama ve rastgele arama hiper parametreleri optimize etmek için kullanılan yaygın algoritmalarındandır. Izgara araması basit modellerde kapsamlı bir arama yaparken parametre uzayı boyutunun yüksek olması durumunda bu yaklaşım karmaşık bir hal almaktadır (Cornejo-Bueno ve ark., 2018; Alade ve ark.,2019). Rastgele arama algoritması ise arama alanını rastgele örnekleyerek çalışır. Bu algoritmanın dezavantajı ise bir sonraki ayarı belirlemek için önceki deneylerden elde edilen bilgileri kullanmamasıdır. Bu durum özellikle işlevin bilinmediği ve deney çalıştırma maliyetinin yüksek olduğu hiper parametre seçim problemlerinde sorun yaratmaktadır. Bayes optimizasyon algoritması bu noktada devreye girmektedir (Wang ve ark., 2012). Bayes optimizasyon algoritması, bayes teoremini kullanarak objektif fonksiyonun posterior dağılımını tahmin eder ve bu dağılıma göre bir sonraki örneğin hiper parametre kombinasyonunu belirler. Rastgele aramanın aksine önceki deneyden elde edilen tüm bilgilerden yararlanır ve sonucu global maksimuma ulaştıracak parametreyi bulmaya çalışır. Yerel optimadan kaçınmak için keşif (sonucun en belirsiz olduğu hiperparametreler) ve sömürünün (optimuma yakın beklenen hiperparametreler) değiştirilmesi gerekmektedir. Bu değiş tokuşu şifrelemek ve bir sonraki noktayı belirlemek için kazanç fonksiyonu tanımlanmaktadır. Kazanç



fonksiyonu,  $f(x)$  'in bir sonraki adımı için aday noktaların fayda tahminlerini döndürür ve maksimum faydayı üreten  $x_{(t+1)}$ 'i seçer (Brochu ve ark., 2010). Bu işlemler algoritma iterasyon sayısı ve işlem süresi gibi durdurma kriterleri sağlanana kadar devam eder ve sonrasında beklenen en düşük kayba sahip nokta seçilir.

Bu çalışmada beklenen iyileştirme (EI) kazanç fonksiyonu olarak kullanılmıştır. Verilen  $f(x)$  fonksiyonunun global minimumunu arayan  $x$  değeri Denklem 3.23'teki gibi elde edilir:

$$x = \operatorname{argmax} E(\max\{0, f_{t+1}(x) - f(x^+)\})|D_t \quad (3.23)$$

Burada  $D$ ,  $x$  bileşenlerinin sayısını temsil etmektedir. Bir Gauss süreci kullanıldığında, EI Denklem 3.24 ve Denklem 3.25 ile ifade edilir.

$$EI(x) = \begin{cases} (\mu(x) - f(x^+))\Phi(z) + \sigma(x)\varphi(z), & \text{eğer } \sigma(x) > 0 \\ 0, & \text{eğer } \sigma(x) = 0 \end{cases} \quad (3.24)$$

$$z = \frac{\mu(x) - f(x^+)}{\sigma(x)} \quad (3.25)$$

Denklem 3.24'te  $\varphi(z)$  olasılık yoğunluk fonksiyonunu ve  $\Phi(z)$  normal dağılımın kümülatif dağılım fonksiyonu ifade etmektedir.

### 3.3.2. Adam optimizasyonu algoritması

Adam algoritması derin sinir ağlarının eğitiminde parametrelerin her biri için öğrenme oranlarının keşfinde uyarlanabilir öğrenme oranı yaklaşımlarından yararlanır. Adam, momentum ve RMSprop ile stokastik gradyan inişinin (SGD) bir karışımı olarak kabul edilmektedir. Bu yöntem, öğrenme oranını RMSprop'a benzer şekilde ölçeklendirmek için kare gradyanları kullanır. Momentumlu SGD'ye benzer şekilde, gradyanın kendisi yerine gradyanın hareketli ortalamasını kullanarak momentumdan yararlanır. Adam algoritması daha düşük yineleme sayısına göre çok yüksek doğruluk elde etmede diğer optimizasyon yöntemlerine üstündür. Algoritma AdaDelta algoritmasından iki kat, AdaGrad algoritmasından ise yaklaşık dört kat

hızla tahmin sonuçlarına ulaşmaktadır. Daha az yinelemenin yanı sıra yüksek doğruluk avantajı Adam'ı çok sayıda önbelleğe alınmış veriye ve daha az veri noktasına sahip derin sinir ağlarını eğitmek için en güçlü araç yapmaktadır. Adam algoritması momentleri yaklaşık olarak hesaplamak için değerlendirilen gradyan üzerinde hesaplanan üssel hareketli ortalamaları kullanır (Kingma ve Ba, 2014).

$$m_{t+1} = \beta m_t - 1 + (1 - \beta_1) \nabla_w L_t \quad (3.26)$$

$$v_{t+1} = \beta_2 v_t - 1 + (1 - \beta_2) (\nabla_w L_t)^2 \quad (3.27)$$

Denklem 3.26 ve Denklem 3.27'da verilen m ve v hareketli ortalamalardır ve  $L_t$ , mevcut eğitim döngüsündeki kayıp fonksiyonunu gösterir. t, mevcut eğitim döngüsünü temsil eder. Ayrıca  $\beta_1$  ve  $\beta_2$ , sırasıyla gradyanlar ve gradyanların ikinci momentleri için unutma faktörüdür. Denklem 3.28-3.30, eğitim döngüsü boyunca tekrarlayarak daha düşük kayıp değerini elde etmek için derin sinir ağının ağırlıklarını Adam algoritması tarafından güncelleme sürecini göstermektedir (Kingma ve Ba, 2014).

$$m'_t = \frac{m_{t+1}}{1 - \beta_1^{t+1}} \quad (3.28)$$

$$v'_t = \frac{v_{t+1}}{1 - \beta_2^{t+1}} \quad (3.29)$$

$$w_{t+1} = w_t - n \frac{m'}{\sqrt{v' + \epsilon}} \quad (3.30)$$

Denklem 3.30'da  $\epsilon$ , 0'a bölmeyi önlemek için kullanılan sonsuz küçük bir ölçekleyicidir.

#### 3.4. Özellik Seçimi Algoritması

Özellik seçimi veri setindeki yararlı girdi değişkenlerinin bulunması sürecidir. Fazla sayıda özellik eğitim süresini arttırabileceği gibi aşırı öğrenme sebebiyle modelin

tahmin hatasını arttırabilmektedir. Buradan hareketle çalışmada önerilen ikinci modelde etkili girdilerin belirlenebilmesi için Komşuluk bileşen analizi algoritması uygulanmıştır.

### 3.4.1. Komşuluk bileşen analizi (NCA)

Bir özellik seçme yöntemi olarak NCA, eğitim verisi üzerinden sınıflandırma doğruluğunu değerlendiren amaç fonksiyonunu maksimize ederek en iyi özellik alt kümesini seçmek için bir özellik ağırlıklandırma yöntemidir (Yang ve ark., 2012). Algoritma, en yakın komşu öğrenme sınıflandırıcısını optimize ederek özellik vektörü  $x_i$ 'ye karşılık gelen ağırlık vektörünü ( $w$ ) bulmaya çalışır. NCA çerçevesinde, tüm örneklerden örnek  $x_i$  için bir referans örnek noktası  $x_j$  seçilir.  $x_j$  'nin  $x_i$  için referans noktası olarak seçilme olasılığı, iki örnek arasındaki mesafenin yakınlığına bağlı olarak daha yüksektir. Bu mesafe ağırlıklandırılmış  $D_w$  mesafesi ile ölçülebilir ve Denklem 3.31 kullanılarak hesaplanır (Yang ve ark., 2012):

$$D_w(x_i, x_j) = \sum_{m=1}^r w_m^2 |x_{im} - x_{jm}| \quad (3.31)$$

$w_m$ , m'inci özelliğin atanan ağırlığıdır.  $P_{ij}$  olasılığı ve  $D_w$  ağırlıklı mesafesi arasındaki ilişki, küçük  $D_w$  için büyük değerler döndüren bir çekirdek fonksiyonu ile kurulabilir.  $P_{ij}$  Denklem 3.32'deki gibi tanımlanır:

$$P_{ij} = \frac{k(D_w(x_i, x_j))}{\sum_{j=1, j \neq i}^n D_w(x_i, x_j)} \quad (3.32)$$

Ayrıca,  $i = j$  ise  $P_{ii} = 0$  değil ise 1 değerini almaktadır. Çekirdek işlevi  $k(z) = \exp\left(-\frac{z}{\sigma}\right)$  şeklinde tanımlanır.  $k$  ve  $\sigma$  parametresi çekirdek genişliğidir ve bu,  $x_j$  numunesinin referans noktası olarak seçilme olasılığını etkiler.  $x_i$  'nin doğru şekilde sınıflandırılma olasılığı Denklem 3.33'teki gibi yazılır:

$$P_{ij} = \sum_{j=1, j \neq i}^n P_{ij} Y_{ij} \quad (3.33)$$

NCA, sınıflandırma problemlerinde kullanıldığı gibi regresyon problemlerinde de özellik seçim problemlerinde kullanılmaktadır. Sınıflandırma problemlerinden tek

farkı  $y$  yanıt değerlerinin sürekli olmasıdır. Bu durumda amaç eğitim seti ( $S$ ) verilen yanıtı tahmin etmektir. Bir regresyon modelinde  $x$  için bir referans noktası olarak eğitim veri setinden rastgele bir şekilde  $Ref(x)$  noktası toplanır.  $Ref(x)$  referans noktasının yanıt değerine eşit  $x$ 'teki yanıt değerini ayarlar. Yine,  $x_j$  noktasının eğitim kümesinden  $x$  için referans noktası olarak seçilme olasılığı  $P(Ref(x)=x_j|S)$  Denklem 3.34 ile ifade edilir.

$$P(Ref(x) = x_{ij} | S) = \frac{k(D_w(x_i, x_j))}{\sum_{j=1}^n k(D_w(x_i, x_j))} \quad (3.34)$$

Regresyon modelinin çapraz doğrulama uygulamasında, yani  $S^{-i}$  deki verileri kullanarak ve eğitim setinde  $(x_i, y_i)$  noktası hariç tutarak  $x_i$  için yanıt tahmin edildiğinde  $x_j$  noktasının  $x_i$  için referans noktası olarak seçilme olasılığı Denklem 3.35 ile hesaplanır.

$$P_{ij} = P(Ref(x_i) = x_j | S^{-i}) = \frac{k(D_w(x_i, x_j))}{\sum_{j=1, j \neq i}^n k(D_w(x_i, x_j))} \quad (3.35)$$

$\tilde{y}_i$ , regresyon modelinin tahmin ettiği yanıt değeri ve  $y_i$  ise  $x_i$  için gerçek yanıt olduğunda iki değer arasındaki uyumsuzluğu ölçen kayıp fonksiyonu  $l(\tilde{y}_i, y_i)$  olarak Denklem 3.36' daki gibi tanımlanır.

$$l_i = E(l(\tilde{y}_i, y_i) | S^{-i}) = \sum_{j=1, j \neq i}^n P_{ij} l(\tilde{y}_i, y_i) \quad (3.36)$$

NCA modelinin aşırı uyumunu önlemek için nihai amaç fonksiyonuna düzeltme parametresi  $\lambda$  eklenmiştir.

Böylece, amaç işlevi şu şekilde ifade edilebilir:

$$A = \sum_{i=1}^n P_i - \lambda \sum_{m=1}^r w_m^2 \quad (3.37)$$

NCA'nın hedefi, amaç fonksiyonu  $A$ 'yı maksimize etmektir. Bunu gerçekleştirmek için  $A$ , eşlenik gradyan yaklaşımı kullanılarak çözülebilir.  $A$ , köşegen bir matrisle

sınırlı olduğunda köşegen değerleri her bir özelliğin ağırlığını vermektedir ve ağırlık sonucuna bağlı olarak, özelliklerin en iyi alt kümesi seçilir.

### 3.5. Eksik Veri İşleme Algoritması

Tahmin modellerinden daha doğru sonuçlar elde edilebilmesi için eksik verilerin işlenmesi önemlidir. Eksik verilerin işlenmesinde başarılı tekniklerden olan ve eksik verilerin veri kümesinin mevcut bilgilerinden hareketle tahmini değerlerle tamamlandığı veri atama yöntemlerinden olan K-NN algoritması kullanılmıştır.

#### 3.5.1. K-NN algoritması

K-NN veri atama algoritması, uzaydaki benzerliği veya yakınlığı kullanarak veri kümesindeki K örneği tanımlamayı sağlayan bir yöntemdir (Beretta ve Santaniello, 2016). Eksik değer ataması için kullanılan diğer yöntemler olan enterpolasyon, hareketli ortalama, Kalman yumuşatma tek değişkenli veri setlerinde atama yapabilirken çok değişkenli veri setlerinde daha iyi çözüm yapabilen yöntemler bulunmaktadır. K-NN veri atama algoritması hem nicel hem de nitel nitelikler için kullanılabilir ve bir dizi eksik değeri doğrudan işleyebilmektedir (Kim ve ark., 2019).

Denklem 3.38'deki  $A_x$  uzaklıkların artan sırasına göre yerleştirilen  $x$  girdi değişkeninin en yakın K komşusunu temsil etmektedir. K tane en yakın komşu seçildikten sonra, eksik veri  $A_x$ 'in  $j$ . özellik verilerinden elde edilen bir tahminle tamamlanmıştır.  $j$ . özellik sayısal bir değişken ise, tamamlanan veri, en yakın K komşusunun ortalama değeri kullanılarak elde edilir. K-NN algoritması ile tamamlanan veri  $\tilde{x}_j$ ,  $k$ 'inci komşunun ağırlığı  $w_k$  ve  $x$  girdi değişkeninin en yakın komşusu  $v_{kj}$  ile hesaplanmaktadır (Denklem 3.39). K en yakın komşu değerlerine ait ağırlık  $w_k$  Denklem 3.40 ile elde edilir (Sanjar ve ark., 2020).

$$A_x = \{v_k\}_{k=1}^K \quad (3.38)$$

$$\tilde{x}_j = \frac{1}{K} \sum_{k=1}^K w_k v_{kj} \quad (3.39)$$

$$w_k = \frac{1}{d(x, v_k)^2} \quad (3.40)$$

Her gözlemin  $x$  girdi değişkenine olan mesafesine göre, daha yakın komşulara daha büyük bir ağırlık atanır (Denklem 3.40).

### 3.6. Performans Kriterleri

Rüzgar enerjisi tahmin çalışmalarında; türbin özelliklerinin, RES'lerin yer aldığı arazinin fiziksel özelliklerini gibi birçok değişkenin etkisi sebebiyle karşılaştırma yapılamamaktadır. Rüzgar enerjisi tahmin modellerinin doğruluğunun yüzde cinsinden ölçülebilmesi ve rüzgar türbininin özelliklerinden etkilenmemesi amacıyla performans kriterleri kurulu güç miktarına oranlanmaktadır (Madsen ve ark., 2005). Çalışmada bu sebeple tahmin modellerinin performansının değerlendirilmesinde NMAE, NRMSE ve  $R^2$  kullanılmıştır.

$$NRMSE (\%) = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (t_i - g_i)^2}}{K} \times 100 \quad (3.41)$$

$$NMAE (\%) = \frac{\frac{1}{N} \sum_{i=1}^N |t_i - g_i|}{K} \times 100 \quad (3.42)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (t_i - g_i)^2}{\sum_{i=1}^N (t_i - \bar{t}_i)^2} \quad (3.43)$$

Denklem 3.41 ve Denklem 3.42 'de belirtilen  $t_i$  i'inci veriye ait tahmin değerini,  $g_i$  i'inci güne ait gerçekleşen değeri ve  $K$  türbinin kapasitesini ifade etmektedir. Denklem 3.43'te ise  $\bar{t}_i$  i'inci veriye ait ortalama tahmin değeridir.

## **BÖLÜM 4. ÖNERİLEN KISA DÖNEM RÜZGÂR GÜCÜ TAHMİN MODELLERİ**

Çalışmada kısa dönem rüzgâr gücünü tahmin etmek için makine öğrenmesi algoritmalarına dayanan farklı yapıda tahmin modellerinin geliştirilmesi amaçlanmış ve önerilen modeller Türkiye’de faaliyet gösteren bir rüzgâr türbini üzerinde uygulanmıştır. Kısa dönem rüzgâr gücü tahmini için biri topluluk öğrenmesi ve bayes optimizasyonu algoritmalarına dayanan statik yapıda ve diğeri LSTM ve NCA algoritmalarına dayanan dinamik yapıda iki farklı hibrit model geliştirilmiştir. Önerilen modellerin farklı tahmin yapısına sahip olması sebebi ile kullanılan tahmin algoritmaları farklılık göstermektedir. Önerilen tahmin modellerinin sistem verimliliğinin artırılmasının yanında santrallerin elektrik piyasasında doğru fiyat teklifini yapmasında da yardımcı olabileceği düşünülmektedir.

Çalışmada tahmin modellerinin performanslarının mevsimlere göre değişkenlik göstermesi sebebiyle test verileri her mevsimden belirli aylar seçilerek oluşturulmuştur. Bu sayede tahmin modellerinin mevsim bazlı performansları da değerlendirilmiştir.

### **4.1. Önerilen Model 1**

Şekil 4.1.’de akış şeması verilen statik yapıdaki Model 1, 6 adımdan oluşmaktadır.

Bu adımlar şu şekildedir:

-Verilerin toplanması: Çalışmada kısa dönem rüzgar güç tahmini için kullanılan parametreler olan rüzgar hızı, rüzgar yönü, sıcaklık, atmosferik basınç, bağıl nem ve rüzgar gücü parametreleri Kaggle açık veri seti platformundan ve NOAA (Ulusal

Okyanus ve Atmosfer Dairesi) kuruluşunun Ulusal Çevresel Bilgi Merkezlerinin arşivinden alınır.

-Verilerin düzenlenmesi: Veriler toplandıktan sonra eksik veriler tamamlanmalı ve veriler standart hale getirilmelidir. Oluşturulan veri seti test verilerinin etkilenmemesi ve aşırı öğrenmeye neden olmamak için ilk olarak eğitim ve test verisine ayrılır sonra veri tamamlama ve normalizasyon işlemleri gerçekleştirilir. Modelde sabit bir eğitim veri seti kullanılarak yılın her mevsiminden veri içerecek şekilde hazırlanır. Veri setindeki eksik veriler K-NN algoritması ile Denklem 3.38-3.40 kullanılarak tamamlanmıştır. Tamamlanan veriler literatürde sıklıkla kullanılan min-maks normalizasyonu ile normalize edilir (Denklem 4.1).

$$x_{nor,i} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (4.1)$$

- Makine öğrenmesi algoritmalarının uygulanması: Ele alınan rüzgar türbininin kısa dönem rüzgar gücü tahmini için literatürde sıklıkla kullanılan ve başarılı sonuçlar veren DT (Heinermann ve Kramer, 2016; Aksoy, 2018), GPR (Lee ve Baldick, 2014; Chen ve ark., 2014) ve SVR (Li ve ark., 2018; Fu ve ark., 2019; Li ve ark., 2020) kullanılır. Makine öğrenmesi algoritmalarının tahmin performansında etkili olan hiper parametre değerleri literatürdeki rüzgar tahmin çalışmalarından farklı olarak makine Bayes optimizasyonu ile elde edilir.

- Makine öğrenmesi algoritmalarının karşılaştırılması: Optimize edilen makine öğrenmesi algoritmalarının tahmin performansları karşılaştırılır ve en başarılı algoritma seçilir. Performans karşılaştırılması yapılırken rüzgar gücü tahmininde sıkça kullanılan MAE ve RMSE hata kriterlerinin türbin kapasitesine oranlanmasıyla normalize edilen NMAE ve NRMSE (Madsen ve ark., 2005) ve  $R^2$  (Tahir ve ark., 2018; Banik ve ark., 2020; Liu ve ark., 2021) kullanılır.

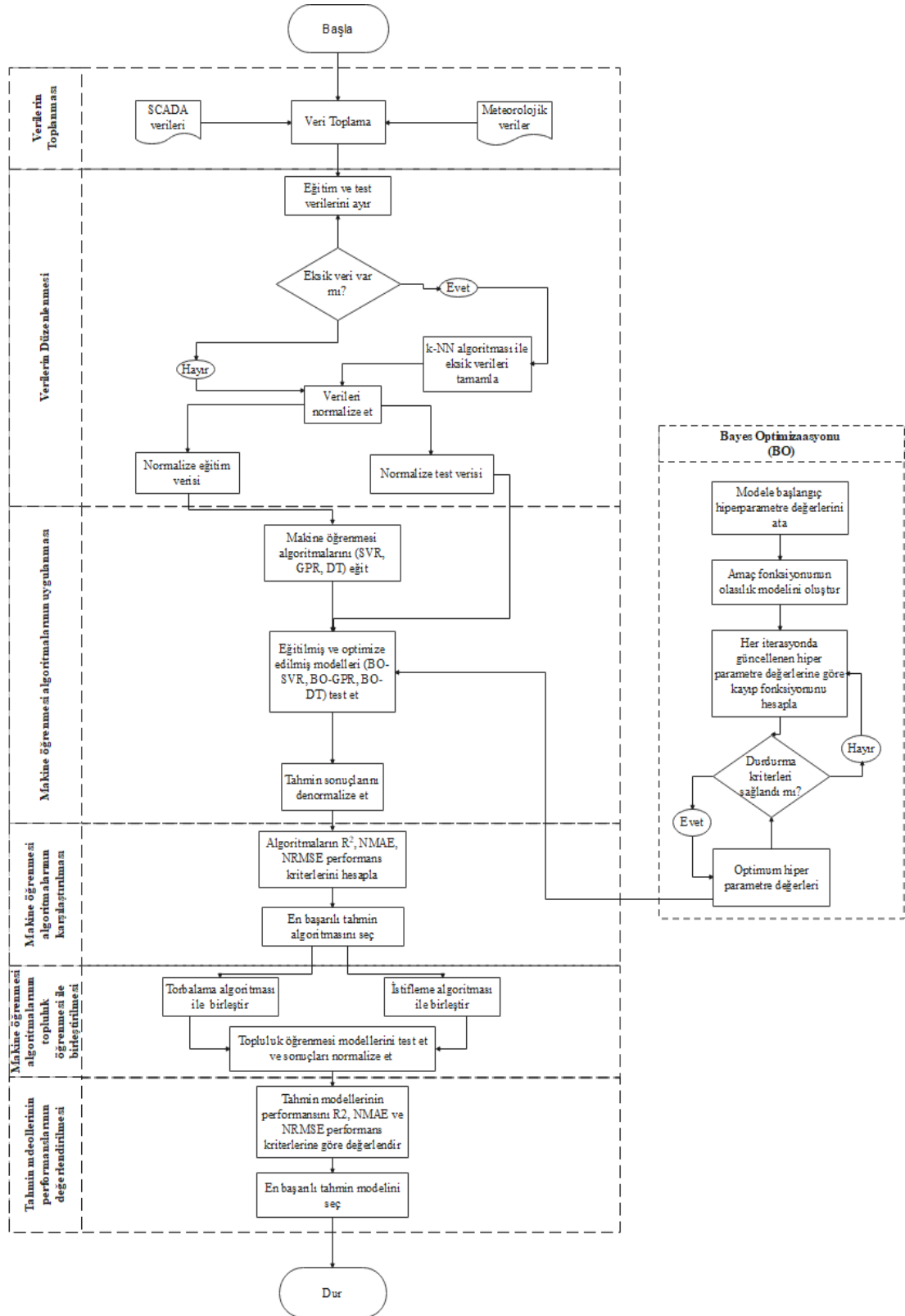
- Makine öğrenmesi algoritmalarının topluluk öğrenmesi ile birleştirilmesi: Modelde optimize edilen makine öğrenmesi algoritmalarını birleştirerek daha iyi tahmin performansına sahip modeller oluşturmak amacıyla topluluk öğrenmesi algoritmaları kullanılır. Çalışmada topluluk öğrenmesi algoritmaları olan torbalama ve istifleme



algoritmaları önceki adımda seçilen makine öğrenmesi algoritmaları ile birleştirilerek homojen ve heterojen yapıda 3 farklı topluluk öğrenmesi modeli oluşturulur.

-Tahmin modellerinin performanslarının değerlendirilmesi: Son adımda oluşturulan topluluk öğrenmesi modellerinin NRMSE, NMAE ve  $R^2$  performans kriterleri mevsimsel olarak hesaplanır ve en başarılı tahmin modeli seçilir.

Şekil 4.1. Model 1'e ait akış şemasını göstermektedir.



Şekil 4.1. Model 1 akış şeması

## 4.2. Önerilen Model 2

Şekil 4.2.'de akış şeması verilen dinamik yapıdaki Model 2, 4 adımdan oluşmaktadır.

Bu adımlar şu şekildedir:

-Verilerin Toplanması: Çalışmada her iki modelin performansını karşılaştırabilmek için ilk modelde kullanılan veri setinden yararlanılır. Veri seti rüzgar hızı, rüzgar yönü, sıcaklık, atmosferik basınç ve bağıl nem girdi parametreleri ve üretilen rüzgar gücü çıktı parametresinden oluşmaktadır.

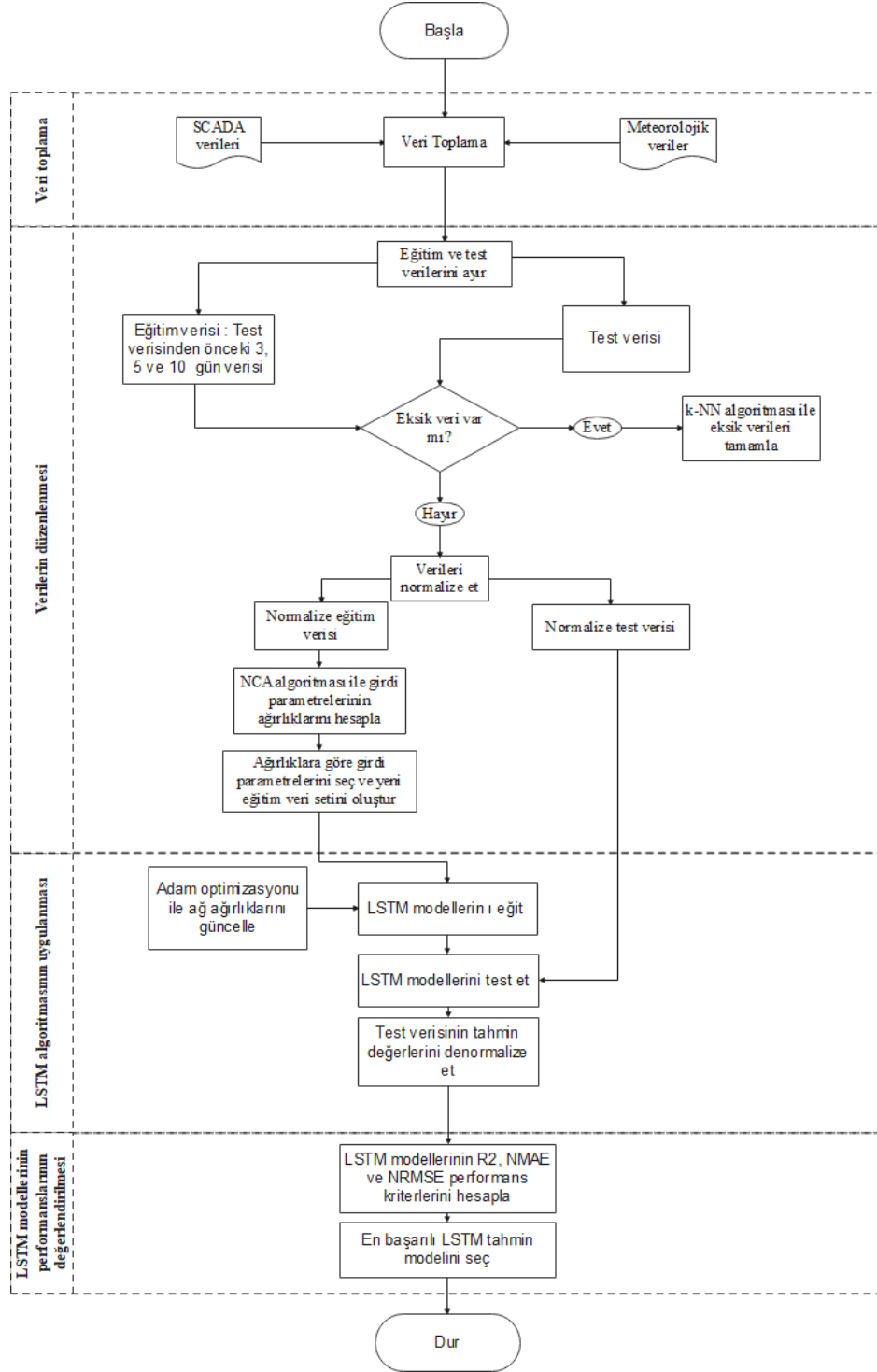
-Verilerin Düzenlenmesi: Dinamik tahmin modelinde eğitim veri seti test veri setindeki günlere bağlı olarak değişmektedir. Model test verisinden önceki 3 gün, önceki 5 gün ve önceki 10 gün olmak üzere 3 farklı eğitim verisi ile eğitilir. Eğitim ve test verisindeki eksik veriler K-NN algoritması ile tamamlanır. Tamamlanan veriler min-maks normalizasyonu ile normalize edilir. Geçmiş rüzgar gücüne ek olarak kullanılacak girdi parametresinin seçimi için Komşuluk Bileşen Analizi (NCA) kullanılır. NCA algoritması ile yararlı girdi değişkenlerinin saptanması ve aşırı öğrenmeyi önlemek amaçlanmıştır.

-LSTM algoritmasının uygulanması: Zamana bağlı dinamik yapıdaki hibrit model için tahmin algoritması olarak zaman serilerinin modellenmesinde sıklıkla kullanılan ve uzun dönem bağımlılıklarını içeren örneklerde başarılı olan LSTM algoritması kullanılır (Xiaoyun ve ark., 2016; Cali ve Sharma, 2019; Son ve ark., 2019; Görgel ve Kavlak, 2020). Ağın öğrenme sürecinde oldukça etkili olan ağırlıklarının güncellenmesi için hesaplama maliyeti düşük olan ve az hafıza gerektiren Adam optimizasyonu kullanılır (Liu ve ark., 2021). Modeli yeniden oluşturmak için önceki zamana ait gerçekleşen rüzgar gücü verileri eğitim veri setine eklenir.

-LSTM modellerinin performans kriterlerinin karşılaştırılması: Test verisinden önceki 3, 5 ve 10 güne ait NCA ile seçilmiş girdi parametrelerinden oluşan eğitim seti ile

eđitilen LSTM modellerinin tahmin sonuları NRMSE, NMAE ve  $R^2$  performans kriterleri ile karřılařtırılarak en iyi LSTM tahmin modeli seilir.

řekil 4.2. Model 2'e ait akıř řemasını gstermektedir.



Şekil 4.2. Model 2 akış şeması

## **BÖLÜM 5. UYGULAMA**

### **5.1. Problem Tanımı**

Elektrik sistemindeki belirsizliği azaltabilmek için işletilen elektrik piyasalarında katılımcılardan gelecek zaman için üretim, tüketim tahminleri ve fiyat teklifleri talep edilmektedir. Elektrik piyasası ile toplanan bu bilgiler neticesinde şebekedeki elektriğin fiyatı minimum olacak şekilde santral çalışma programları oluşturulmaktadır. Bu sayede üretim ve tüketim dengesi sağlanıp belirsizlik azaltılarak elektrik sistem verimliliğinde artış sağlanabilmektedir. Rüzgar gücü tahminleri elektrik sisteminin verimliliğinin artışının yanı sıra santral işletmecileri için de önem taşımaktadır. Santraller piyasaya verdikleri yanlış teklif ve tahmin hatalarından kaynaklanan cezalar ödemekte ve bu da gelirlerini etkilemektedir. Bu nedenle çalışmada hem daha fazla rüzgar enerjisinin sisteme entegre edilerek verimliliğin artması hem de santrallere daha doğru güç tahmini ve piyasa teklifi ile gelirlerinde artış sağlamak amacıyla makine öğrenmesi algoritmalarına dayanan iki hibrit kısa dönem rüzgar gücü tahmin modeli önerilmiştir. Önerilen modeller Türkiye’de faaliyet gösteren bir rüzgar türbini üzerinde uygulanmış ve modellerin tahmin performansları karşılaştırılmıştır.

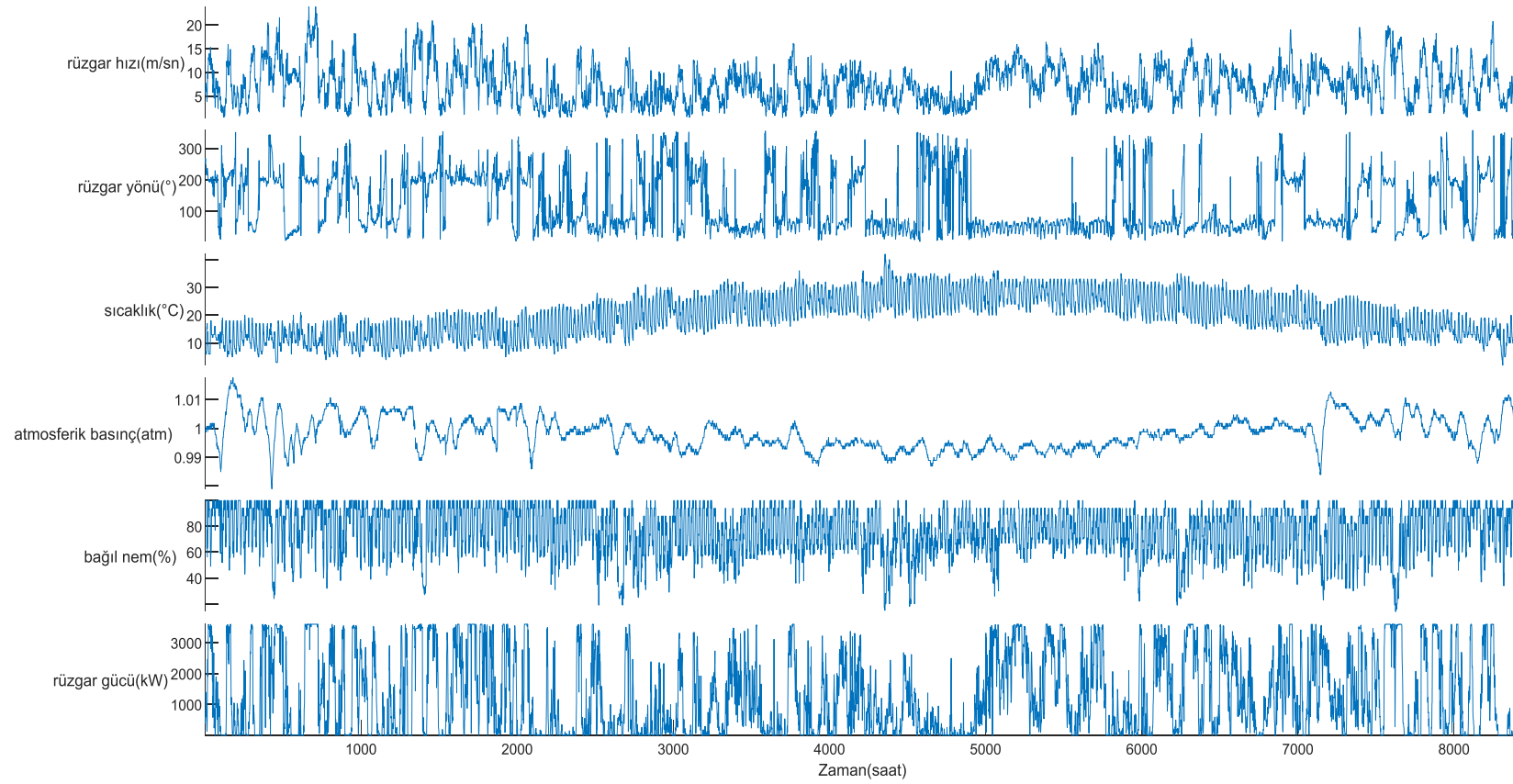
### **5.2. Önerilen Modellerin Uygulanması**

#### **5.2.1. Önerilen model 1’in uygulama adımları**

##### **5.2.1.1. Verilerin toplanması**

Çalışmada Türkiye’de faaliyette olan ve 9 türbinden oluşan bir rüzgar santralinin bir türbinine ait Merkezi Denetleme Kontrol ve Veri Toplama Sistemi (SCADA) sistem verileri kullanılmıştır (Erisen, 2019). Santralin toplam kurulu gücü 30,45

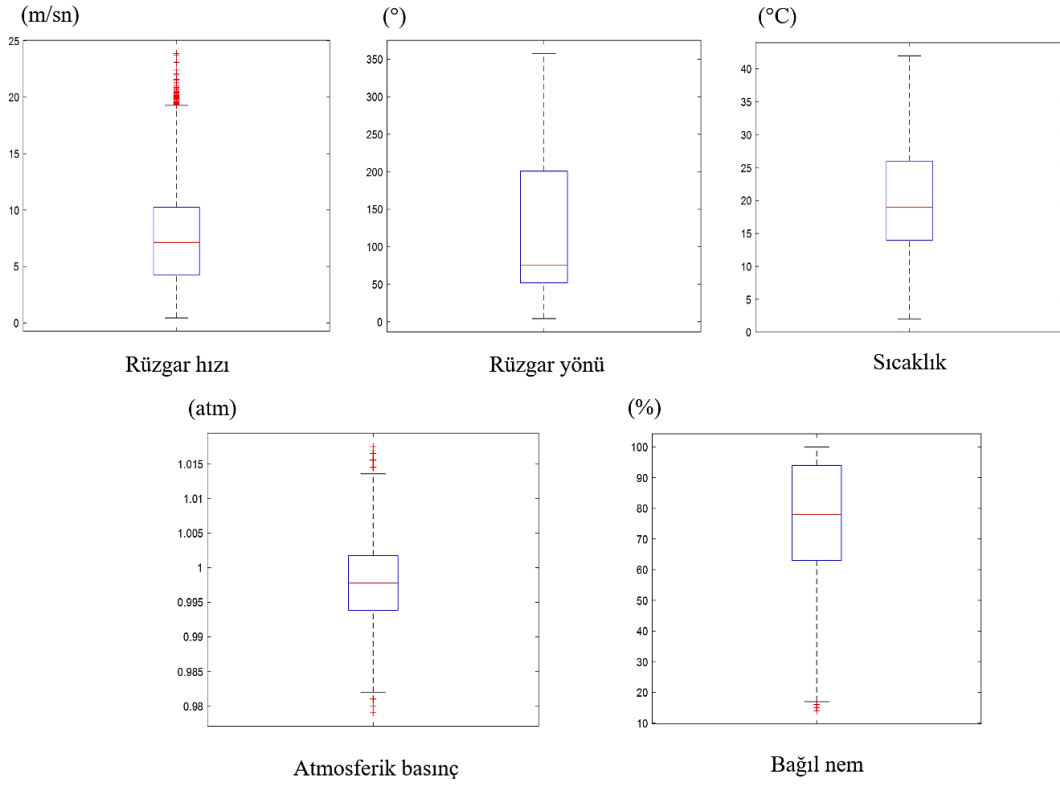
MW olup ele alınan türbinin kurulu gücü ise 3,6 MW'dir. Kullanılan veri seti 01/01/2018-16/12/2018 tarihlerine ait 10 dakikalık zaman adımında rüzgar hızı (m/sn), rüzgar yönü ( $^{\circ}$ ) ve üretilen rüzgar gücü (kW) verilerinden oluşmaktadır. Rüzgar hızı ve yönünün dışında rüzgar gücünü etkileyen ana faktörlerden biri de hava yoğunluğudur. Rüzgar türbinin bulunduğu ortamın sıcaklığı, basıncı ve nemi, türbinin hem hava yoğunluğunu hem de çalışma performansını etkileyebilmektedir (Zhang ve ark., 2020). Bu sebeple ilgili türbinin yer aldığı konumda bulunan ölçüm istasyonundan 2018 yılına ait 8399 saatlik sıcaklık ( $^{\circ}$ C), atmosferik basınç (atm) ve bağıl nemden (%) oluşan meteorolojik veriler ise 1 saatlik zaman adımında sağlanmıştır (<https://www.ncdc.noaa.gov/cdo-web> Erişim tarihi: 15.10.2020). Elde edilen veriler farklı zaman adımında olduklarından hepsi 1 saatlik zaman adımında düzenlenmiştir. Şekil 5.1.' de kısa dönem rüzgar gücü tahmini yapılan türbin için oluşturulan veri seti verilmiştir.



Şekil 5.1. Veri seti



Şekil 5.2.' de girdi parametreleri için oluşturulan kutu diyagramları gösterilmiştir. Kutu diyagramları ile rüzgar hızı, atmosferik basınç ve bağıl nem parametre verilerinde anormal değerlerin yer aldığı görülmektedir.



Şekil 5.2. Girdi parametrelerine ait kutu diyagramları

### 5.2.1.2. Verilerin düzenlenmesi

Veri düzenleme aşamasında sırasıyla eğitim ve test verisinin belirlenmesi, eksik verilerin tamamlanması ve normalizasyon uygulamaları yapılmıştır.

Rüzgar gücünün hava durumuna bağlı olarak mevsimlere göre değişkenlik göstermesi tahmin modellerinin doğruluğunun da değişkenlik göstermesine sebep olmaktadır (Jung ve Broadwater, 2014). Çalışmada bu nedenle önerilen modelin performansının daha kapsamlı değerlendirilmesi için her mevsime ait veri içeren bir veri seti üzerinde test edilmesi amaçlanmıştır. Bu nedenle test için her mevsimi temsilen aylar seçilmiştir. Bu seçim yapılırken ayların güç eğrileri ve rüzgar hız ve üretilen rüzgar gücü arasındaki korelasyon ilişkisi dikkate alınmıştır. Tablo 5.1.' de

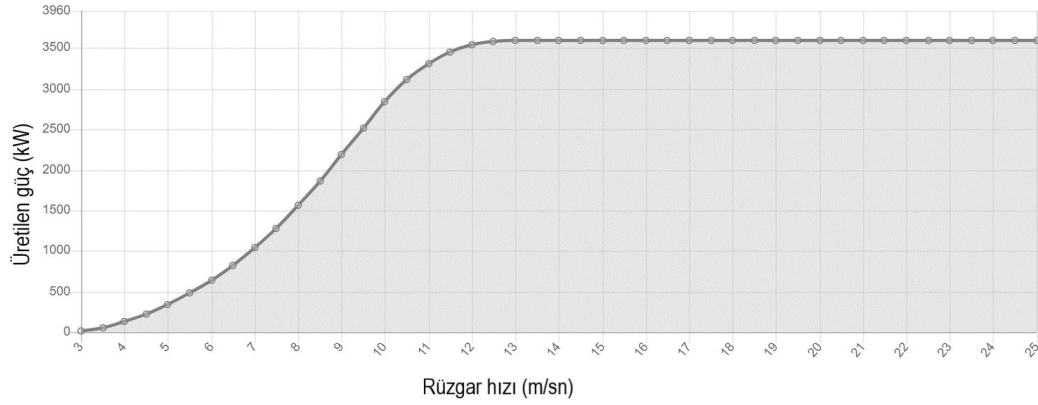
aylara ait rüzgar hızı-rüzgar gücü korelasyon değerlerine bakıldığında kış mevsiminde Ocak ayı, ilkbahar mevsiminde Nisan ayı, yaz mevsiminde Temmuz ayı ve sonbahar mevsiminde ise Ekim ayı en yüksek korelasyon değerlerine sahip aylar olduğu görülmektedir.

Tablo 5.1. Aylara ait rüzgar hızı- rüzgar gücü korelasyon değerleri

Ocak	Şubat	Mart	Nisan	Mayıs	Haziran
0,797	0,705	0,929	0,951	0,950	0,954
Temmuz	Ağustos	Eylül	Ekim	Kasım	Aralık
0,975	0,938	0,949	0,975	0,915	0,784

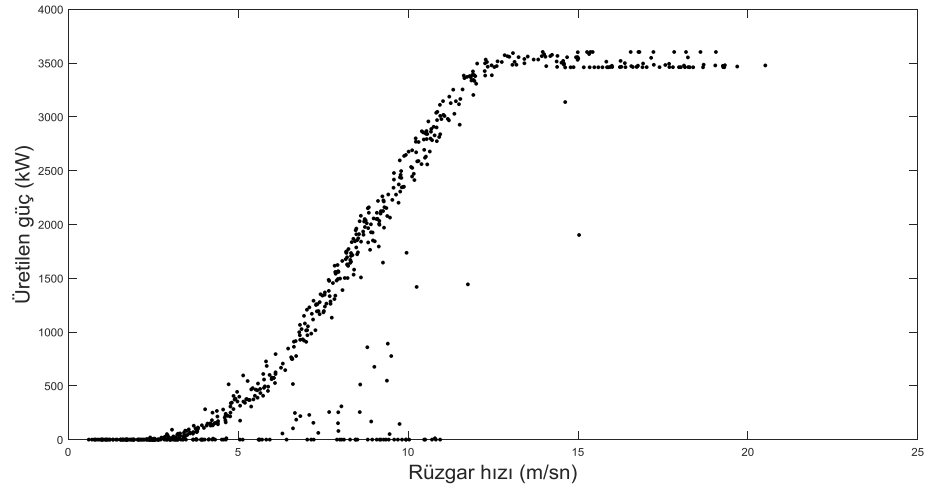
Rüzgar hızı üretilen rüzgar gücünü belirleyici faktörler arasında en önemlisi olduğu için korelasyon analizine ek olarak bu aylara ait güç eğrileri de incelenmiştir.

Güç eğrisi rüzgar hızına bağlı olarak rüzgar gücünün değişimini gösteren eğrilerdir. Bir rüzgar güç eğrisinde rüzgar hızı devreye girme hızından yüksek ise elektrik enerjisi üretilmeye başlanmakta ve rüzgar hızı arttıkça üretilen enerji de doğru orantılı olarak artmaktadır. Rüzgar hızı nominal hıza ulaştığında ise üretilen güç maksimum olmakta ve nominal hız aşırsa türbin hızı düşürülmeye çalışılmaktadır. Rüzgar hızı devreden çıkma hızına ulaştığında ise sistem kapatılmaktadır. Çalışmada kısa dönem rüzgar gücü tahmini yapılan türbinin devreye girme hızı 3 m/sn, nominal hız 13 m/sn, devreden çıkma hızı 25 m/sn ve kapasitesi 3600 kW'dır. Bu bilgiler doğrultusunda ele alınan türbinin teorik güç eğrisi Şekil 5.3.'deki gibi olmaktadır.



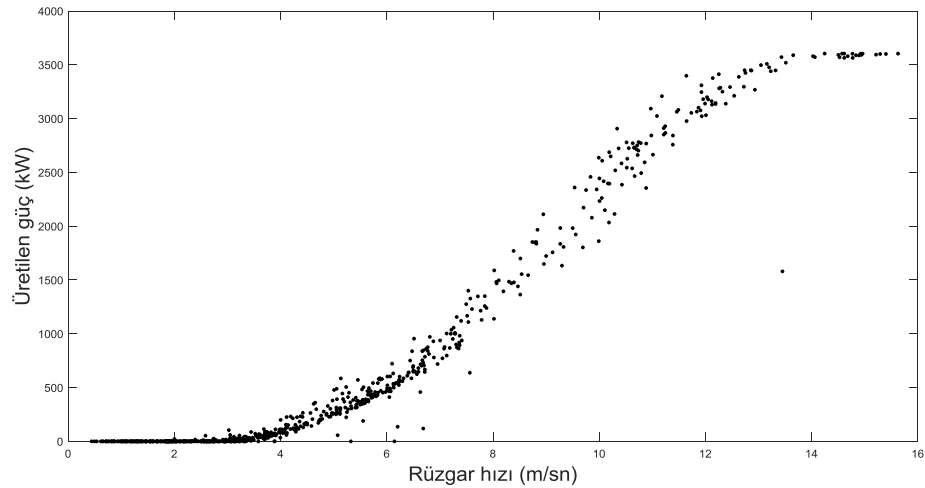
Şekil 5.3. Ele alınan türbinin teorik güç eğrisi

SCADA ölçümleri, genellikle bakım ve arızalardan kaynaklanan aykırı verileri içerebilmektedir. Aykırı veriler tahmin modelinin performansını olumsuz etkileyebilmektedir. Aylara göre güç eğrilerine bakıldığında arıza verileri, mantıksız veriler ve kesinti verileri olmak üzere aykırı veriler olduğu görülmüştür. Mantıksız veriler, rüzgar hızı düşük olduğunda rüzgar gücünün yüksek olduğu durumları ifade eder ve veri toplamadaki ve iletimdeki hatalardan kaynaklanıyor olabilir. Arıza verileri, rüzgar hızı yüksek olduğunda rüzgar gücünün düşük olduğu durumları belirtmektedir bu durum rüzgar türbinlerinin bağlantı kaybı ve veri iletim hatasından kaynaklanabilir. Kesinti verileri ise rüzgar hızının devreye giriş hızından büyük olmasına karşın üretilen gücün 0 olduğu durumları göstermektedir. Test için seçilecek ayların Şekil 5.4., Şekil 5.5., Şekil 5.6. ve Şekil 5.7. sırasıyla Ocak, Nisan, Temmuz ve Ekim aylarına ait güç eğrilerini göstermektedir.



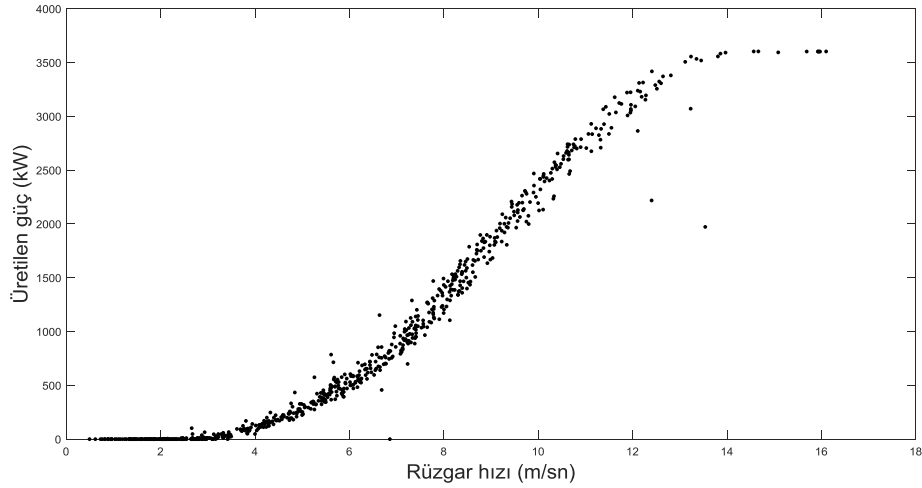
Şekil 5.4. Ocak ayı rüzgar güç eğrisi

Kış aylarında hava koşulları sebebiyle büyük oranda kesintiler gerçekleşmiş ve bu durum üretilen rüzgar gücünü büyük ölçüde etkilemiştir. Ocak ayı yıl içerisinde en yüksek rüzgar hızının gerçekleştiği ay olmasına rağmen en yüksek üretimin gerçekleştiği ay olamamıştır. Ay içerisinde yaşanan kesintiler nedeniyle rüzgar hızı devreye giriş hızından yüksek olmasına rağmen gerçekleşen üretim 0 olarak kaydedilmiştir. Kesintilerin dışında türbindeki arızalara bağlı olarak yüksek rüzgar hızına rağmen düşük güç eldesi sağlanmıştır.



Şekil 5.5. Nisan ayı rüzgar güç eğrisi

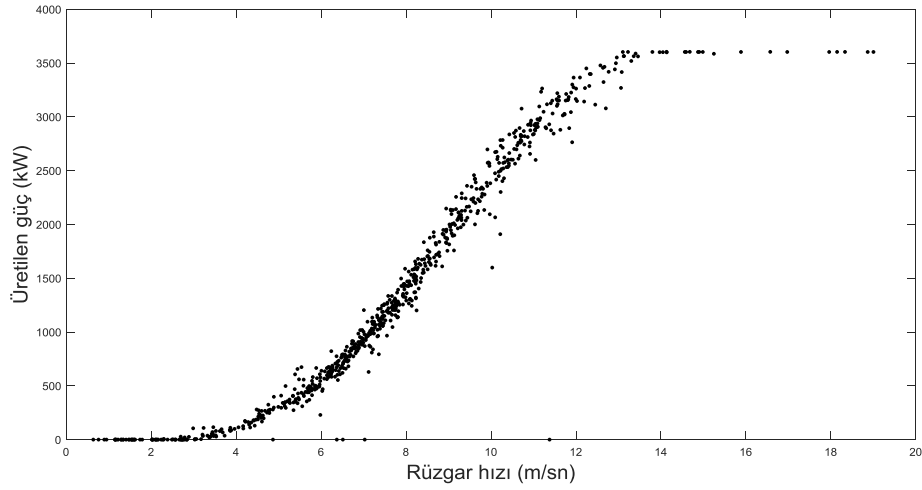
Nisan ayı güç eğrisine bakıldığında türbindeki arızalar nedeniyle yüksek rüzgar hızına rağmen düşük üretim miktarı elde edilen durumlar olduğu görülmektedir. Ayrıca bu ay içerisinde kısa süreli olmasına rağmen rüzgar hızının devreye girme hızından yüksek olmasına rağmen üretimin gerçekleşmediği zaman aralıkları mevcuttur.



Şekil 5.6. Temmuz ayı rüzgar güç eğrisi

Temmuz ayında diğer aylara oranla daha düşük rüzgar hızı görülmüş ve buna bağlı olarak daha az üretim yapılmıştır. Bu ay içerisinde düşük rüzgar hızına rağmen yüksek güç eldesinin sağlandığı hatalı veriler görülmektedir. Ayrıca arıza kaynaklı yüksek rüzgar hızına rağmen düşük güç elde edilen zaman aralıkları olduğu da görülmektedir.

Ekim ayında ise Temmuz ayına oranla daha yüksek rüzgar hızı gerçekleşmiş ve daha fazla üretim yapılmıştır. Bu ayda bağlantı kaybı ve iletim hatası nedeniyle rüzgar hızının yüksek olmasına rağmen beklenenden daha düşük rüzgar gücü üretiminin gerçekleştiği zaman aralıkları gözlenmiştir.



Şekil 5.7. Ekim ayı rüzgar güç eğrisi

Özetlenecek olursa test için seçilen ayların güç eğrilerinin aykırı verilerin varlığına rağmen teorik güç eğrisine yakın ve uyumlu olduğu görülmüştür ve korelasyon verileri de bu durumu destekler niteliktedir. Bu nedenle kış, ilkbahar, sonbahar ve yaz mevsimlerini temsilen seçilen Ocak, Nisan, Temmuz ve Ekim aylarından rastgele seçilen 15' er günden oluşan toplamda 60 güne (1440 saat) ait veri test verisi olarak belirlenmiş geri kalanı (6981 saat) ise eğitim verisi olarak atanmıştır. Eğitim ve test verisi belirlendikten sonra veri tamamlama ve normalizasyon işlemleri gerçekleştirilmiştir.

Çalışmada kullanılan veri setinde meteorolojik veriler olan sıcaklık, atmosferik basınç ve bağıl nem parametrelerinde toplamda 228 adet eksik veri olduğu görülmüştür. Eksik verilerin tamamlanması için makine öğrenmesi algoritmalarının ön işleminde sıklıkla kullanılan K-NN algoritmasından yararlanılmıştır. Eksik veriler k en yakın satır komşu değerinin ağırlıklı ortalaması ile doldurulmuştur. Uygulamada varsayılan parametreler olan  $k=5$  ve Öklid uzaklık metriği kullanılmıştır.

Kullanılan parametrelerin birbirinden çok farklı değer aralıklarına sahip olması, tahmin doğruluğunu etkilemekte ve belirsizliğe neden olmaktadır. Bu durumu önlemek için K-NN algoritması ile tamamlanan eğitim ve test verileri Denklem 4.3

kullanılarak normalize edilmiştir. Rüzgar yönü verisi normalize edilirken önce dereceden radyana çevrilmiş daha sonra sinüs ve kosinüs değerleri hesaplanmıştır.

### 5.2.1.3. Makine öğrenmesi algoritmalarının uygulanması

Çalışmada kısa dönem rüzgar gücü tahmini için makine öğrenmesi algoritmalarından DT, GPR ve SVR uygulanmıştır. Makine öğrenmesi algoritmalarının eğitim davranışlarını kontrol ederek model performansı üzerinde doğrudan etkiye sahip olan hiper parametreler optimize edilmesi önemlidir. Bu nedenle çalışmada literatürdeki rüzgar tahmin çalışmalarından farklı olarak makine öğrenmesi algoritmaları Bayes optimizasyonu ile optimize edilmiştir. Bayes optimizasyonu bir sonraki veri noktasını bulmak için önceki iterasyonlara ait tüm bilgilerden yararlanarak daha az sayıda iterasyon ile optimum çözümü sağladığı için çalışmada bu yöntem tercih edilmiştir (Wang ve ark., 2012).

Seçilen tahmin algoritmaları olan GPR, DT ve SVR ilk olarak varsayılan parametrelerle eğitilmiş ve tahmin sonuçları edilmiştir. Daha sonra Bayes optimizasyonu algoritması ile 5 kat çapraz doğrulama kaybını minimuma indiren hiperparametre değerleri araştırılmıştır. Çalışmada, Bayes optimizasyonu sırasında parametre uzayının nasıl araştırılması gerektiğini ifade eden edinim fonksiyonu olarak beklenen iyileştirme kullanılmıştır.

Optimizasyonun ilk adımında, başlangıç hiperparametre değerleri modele atanmıştır. Model, verilerin 4/5 ile eğitilmiş ve 1/5'i üzerinde test edilmiştir. Her yinelemede, yeni bir hiper parametre vektörü oluşturulmuş ve kayıp fonksiyonu değeri ( $\log(1 + \text{kayıp})$ ) hesaplanmıştır. Yinelemeler tamamlandıktan sonra, kayıp fonksiyonunu minimum yapan en iyi hiper parametre vektörü seçilmiştir. Seçilen hiperparametre vektörü ile nihai model tüm veri seti üzerinde eğitilmiş ve tahmin sonuçları elde edilmiştir. GPR algoritması için sinyal standard sapması olan sigma parametresinin, DT algoritması için minimum yaprak örnek sayısı ve maksimum dal düğüm sayısı parametrelerinin, SVR algoritması için epsilon, düzenleme sabiti ve çekirdek fonksiyonu parametrelerinin optimum değerleri araştırılmıştır. Çalışmada

Bayes optimizasyonu ile optimize edilen hiper parametreler ve değer aralıkları Tablo 5.2.'de verilmiştir.

Tablo 5.2. Çalışmada kullanılan algoritma ve algoritmalara ait hiper parametre değer aralıkları

Algoritma	Hiper parametre	Değer aralığı
DT	Minimum yaprak örnek sayısı	[1, 3491]
	Maksimum dal düğüm sayısı	[1, 6980]
GPR	Sigma	[1e-4, 1,2900e+04]
SVR	Epsilon ( $\epsilon$ )	[4.8000e-05, 4,800]
	Düzenleme sabiti (C)	[1e-3,1e3]
	Çekirdek fonksiyonu	Gaussian, doğrusal, polinamiyal

DT algoritması için yaprak düğümün yanıtını hesaplamak için kullanılan minimum eğitim örneği sayısını ifade eden yaprak örnek sayısı ve karar bölünmelerinin maksimum sayısını ifade eden bölme sayısı parametrelerinin optimum değerleri araştırılmıştır. DT algoritması için optimizasyon iterasyonlarının hızlı tamamlanması sebebi ile durdurma kriteri varsayılan değer 30 yerine 50 iterasyon olarak belirlenmiştir. Optimizasyon iterasyonlarında olurlu çözüm veren hiperparametre vektörüne karşılık değerlendirme sonucu kabul, gelinen iterasyona kadar en düşük kayıp değerini veren hiperparametre vektörü için ise en iyi şekilde ifade edilmiştir. Tablo 5.3., DT algoritması için Bayes optimizasyonu iterasyonlarını göstermektedir.

Tablo 5.3. DT algoritması için Bayes optimizasyonu iterasyonları

İterasyon	Değerlendirme sonucu	Amaç fonksiyonu: $\log(1+kayıp)$	İşlem süresi	Minimum yaprak örnek sayısı	Maksimum dal düğüm sayısı
1	En iyi	0,011502	0,051844	467	15
2	Kabul	0,023687	0,036668	1748	922
3	En iyi	0,0092397	0,063725	20	152
4	Kabul	0,022633	0,041748	3	2
5	Kabul	0,0095915	0,066234	36	96
6	Kabul	0,0092688	0,057798	7	82



Tablo 5.3. (Devamı)

İterasyon	Değerlendirme sonucu	Amaç fonksiyonu: $\log(1+kayıp)$	İşlem süresi	Minimum yaprak örnek sayısı	Maksimum dal düğüm sayısı
7	Kabul	0,009631	0,055658	1	37
8	Kabul	0,040354	0,033621	2635	126
9	Kabul	0,0097459	0,049667	33	20
10	Kabul	0,01297	0,041502	329	4
11	Kabul	0,010356	0,079886	1	221
12	Kabul	0,0093031	0,12765	6	1257
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
<b>36</b>	<b>En iyi</b>	<b>0,0088459</b>	<b>0,070322</b>	<b>3</b>	<b>220</b>
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
43	Kabul	0,032119	0,040765	1	1
44	Kabul	0,010378	0,047445	271	6915
45	Kabul	0,010376	0,047113	269	13
46	Kabul	0,0097837	0,052952	102	275
47	Kabul	0,0095295	0,052477	5	19
48	Kabul	0,0095316	0,056205	79	37
49	Kabul	0,0093518	0,063937	28	504
50	Kabul	0,010085	0,054857	2	17

Optimizasyon sonuçlarına göre model, 36'ncı iterasyonda minimum yaprak örnek sayısı 3 ve maksimum dal düğüm sayısı 220 değerini aldığı anda minimum kayıp fonksiyonu değerine ulaşmıştır. Tablo 5.4. DT algoritması için varsayılan ve optimize edilmiş hiper parametre değerlerini göstermektedir.

Tablo 5.4. DT algoritması varsayılan ve optimize edilmiş hiperparametre değerleri

Hiperparametreler	Minimum yaprak örnek sayısı	Maksimum dal düğüm sayısı
Varsayılan	1	6980
Optimize edilmiş	3	220

GPR algoritması için Gauss süreç modelinin sinyal standart sapmasını ifade eden sigma parametresinin optimum değeri araştırılmıştır. Durdurma kriteri varsayılan değer 30 iterasyon olarak belirlenmiş ve 30 iterasyon sonunda optimum parametreler bulunmuştur. Tablo 5.5., GPR algoritması için Bayes optimizasyonu iterasyonlarını göstermektedir.

Tablo 5.5. GPR algoritması için Bayes optimizasyonu iterasyonları

İterasyon	Değerlendirme sonucu	Amaç fonksiyonu: log(1+kayıp)	İşlem süresi	Sigma
1	En iyi	0,0089561	60,936	0,030826
2	Kabul	0,084427	73,087	0,00016228
3	Kabul	0,022513	66,347	0,0017231
4	Kabul	0,016987	36,595	15,475
5	Kabul	0,010362	54,121	0,026294
6	En iyi	0,0089371	32,269	0,18133
7	En iyi	0,0072951	30,06	0,075497
8	En iyi	0,0071559	28,128	0,074311
9	Kabul	0,011164	32,353	0,4794
10	En iyi	0,0069087	31,476	0,054988
<b>11</b>	<b>En iyi</b>	<b>0,0068647</b>	<b>67,936</b>	<b>0,11505</b>
12	Kabul	0,0068656	74,729	0,055745
13	Kabul	0,007034	81,313	0,055743
14	Kabul	0,024263	105,37	35,789
15	Kabul	0,060864	147,6	0,0051597
16	Kabul	0,0094054	70,877	0,28679
17	Kabul	0,022137	115,52	0,00087553
18	Kabul	0,0083773	85,794	0,11097
19	Kabul	0,006899	76,786	0,062104
20	Kabul	0,0070965	78,572	0,046684

Tablo 5.5. (Devamı)

İterasyon	Değerlendirme sonucu	Amaç fonksiyonu: log(1+kayıp)	İşlem süresi	Sigma
21	Kabul	0,006994	75,026	0,060714
22	Kabul	0,043991	186,7	0,00010003
23	Kabul	0,01681	40,84	0,83821
24	Kabul	0,0069309	34,393	0,052376
25	Kabul	0,007028	29,293	0,063059
26	Kabul	0,0069396	78,824	0,052951
27	Kabul	0,062596	98,879	0,00040933
28	Kabul	0,060073	65,317	0,011038
29	Kabul	0,0069398	30,578	0,052232
30	Kabul	0,022756	57,746	0,0029126

Elde edilen sonuçlara göre, model 11'inci iterasyonda sigma 0,11505 değerini alarak optimum sonuca ve minimum çapraz doğrulama kaybına ulaşmıştır. Tablo 5.6. GPR algoritması için varsayılan ve optimize edilmiş hiper parametre değerlerini göstermektedir.

Tablo 5.6. GPR algoritması varsayılan ve optimize edilmiş hiperparametre değerleri

Hiperparametre	Sigma
Varsayılan	373
Optimize edilmiş	0,11505

SVR algoritması için eğitimden önce verilere uygulanan dönüşümü belirleyen çekirdek fonksiyonu, gözlemlere uygulanan cezayı kontrol eden düzenleme sabiti ve epsilon parametrelerinin optimum değerleri araştırılmıştır. Durdurma kriteri varsayılan değer 30 iterasyon olarak belirlenmiş ve Tablo 5.7.'de SVR algoritması için Bayes optimizasyonu iterasyonları gösterilmektedir.

Tablo 5.7. SVR algoritması için Bayes optimizasyonu iterasyonları

İterasyon	Değerlendirme sonucu	Amaç fonksiyonu: $\log(1+kayıp)$	İşlem süresi	C	E	Çekirdek fonksiyonu
1	En iyi	0,12055	0,98747	0,082336	17,448	Doğrusal
2	En iyi	0,018887	39,963	0,0039777	0,0013106	Polinomiyal
3	Kabul	0,12055	0,24973	61,741	0,94565	Polinomiyal
4	En iyi	0,010093	1473,7	391,74	0,02217	Gaussian
5	Kabul	0,03471	10,867	0,0014124	0,04543	Gaussian
6	Kabul	0,010687	144,67	10,186	0,0070008	Gaussian
7	Kabul	0,010414	1408,2	962,16	0,0004895	Gaussian
8	Kabul	0,035561	1118,4	873,34	0,00048878	Polinomiyal
9	Kabul	0,12055	0,22883	948,74	47,391	Gaussian
10	Kabul	0,010136	1108,9	992,49	0,0034751	Gaussian
11	Kabul	0,048992	51,711	0,0010264	0,00048909	Gaussian
12	Kabul	0,018291	37,885	0,0010086	0,00048806	Polinomiyal
13	Kabul	0,01958	982,71	908,03	0,00049402	Doğrusal
14	Kabul	0,053285	36,752	0,0010336	0,00049127	Doğrusal
15	Kabul	0,020945	989,63	999,79	0,0069272	Doğrusal
16	Kabul	0,010271	981,97	300,96	0,010709	Gaussian
17	Kabul	0,01029	1199,8	212,86	0,0014242	Gaussian
18	Kabul	0,01057	65,954	25,634	0,023347	Gaussian
19	Kabul	0,010614	2331,9	907,33	0,0010843	Gaussian
20	Kabul	0,010373	1648,1	131,12	0,0048452	Gaussian
21	Kabul	0,010434	753,29	92,637	0,014739	Gaussian
22	En iyi	0,0096581	1445,6	945,77	0,049521	Gaussian
23	Kabul	0,009843	157,19	134,11	0,062683	Gaussian
24	Kabul	0,0098601	587,89	300,97	0,044785	Gaussian
<b>25</b>	<b>En iyi</b>	<b>0,0096151</b>	<b>1036,7</b>	<b>399,1</b>	<b>0,01481</b>	<b>Gaussian</b>
26	Kabul	0,0096653	1382,2	913,86	0,04975	Gaussian
27	Kabul	0,009628	520,36	811,22	0,080983	Gaussian
28	Kabul	0,0096409	826,89	841,24	0,064184	Gaussian
29	Kabul	0,12055	0,67599	0,0010709	47,876	Polinomiyal
30	Kabul	0,12055	0,67059	0,0010016	45,804	Gaussian

SVR algoritmasının optimum hiper parametre değerlerine 25'inci iterasyonda ulaşılmıştır. Algoritmanın optimum çekirdek fonksiyonu gaussian, optimum düzenleme sabiti değeri 399,1 ve optimum epsilon değeri 0,01481 olarak bulunmuştur (Tablo 5.8.).

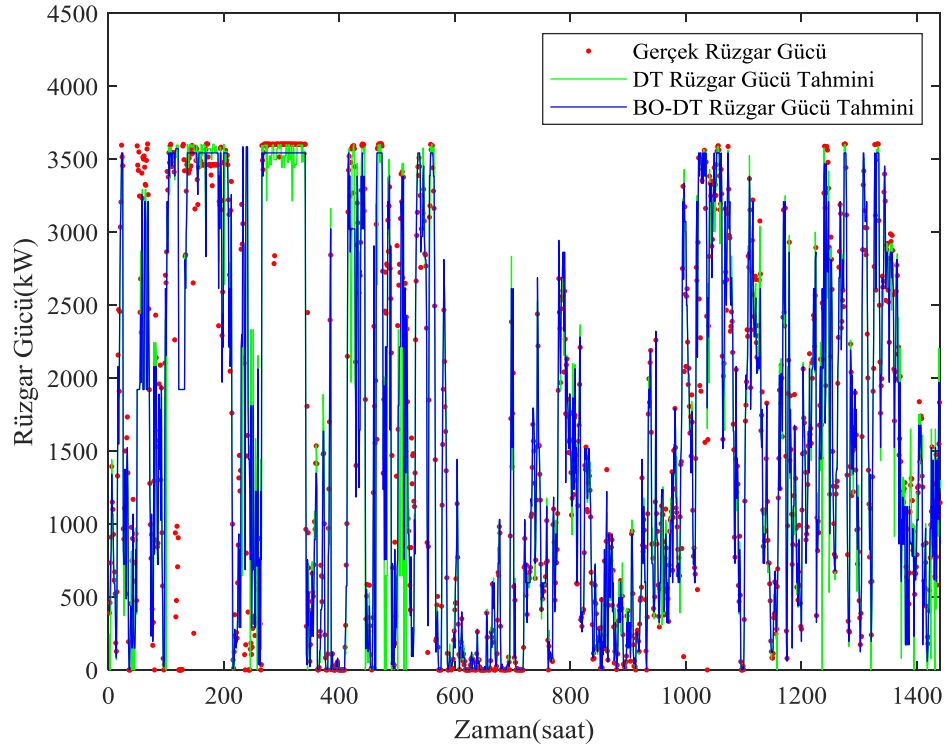
Tablo 5.8. SVR algoritması varsayılan ve optimize edilmiş hiperparametre değerleri

Hiperparametreler	C	Epsilon	Çekirdek Fonksiyonu
Varsayılan	1	0,048	Doğrusal
Optimize edilmiş	399,1	0,01481	Gaussian

Optimum hiper parametre değerleriyle eğitilen BO-DT, BO-GPR ve BO-SVR algoritmaları ile test verisi için tahmin sonuçları elde edilmiştir.

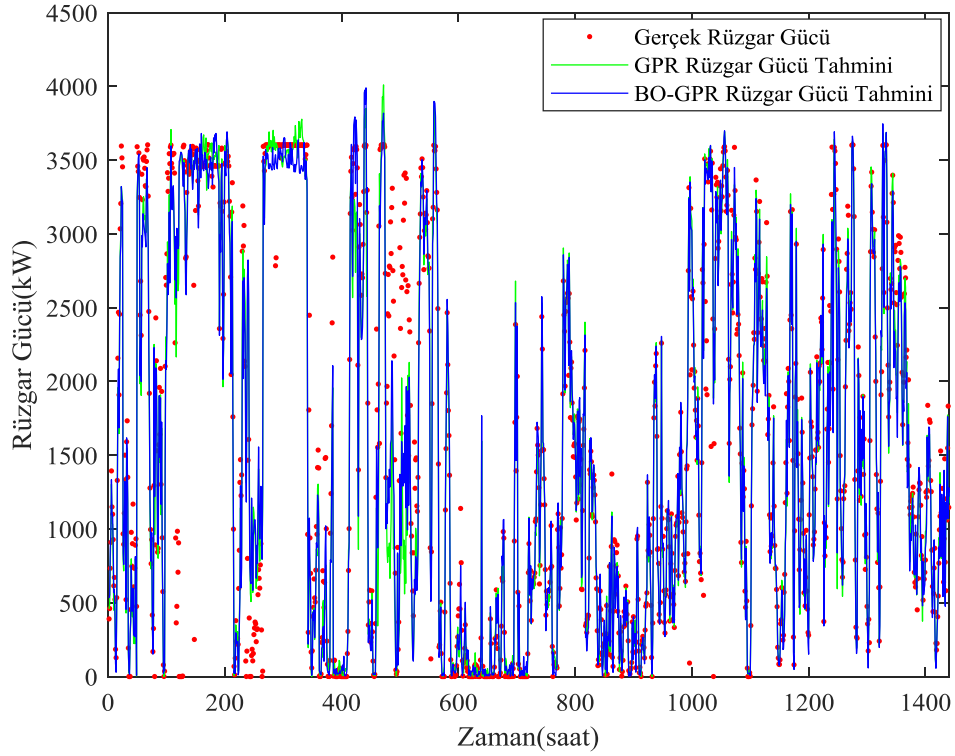
#### 5.2.1.4. Makine öğrenmesi algoritmalarının karşılaştırılması

Bayes optimizasyonu ile optimize edilen parametrelerle eğitilen BO-DT, BO-GPR ve BO-SVR algoritmalarının tahmin sonuçları ile varsayılan parametrelerle eğitilen DT, GPR ve SVR algoritmalarının tahmin sonuçlarının karşılaştırması Şekil 5.8., Şekil 5.9. ve Şekil 5.10. 'da verilmiştir.



Şekil 5.8. Test verisine ait üretilen güç, DT ve BO-DT rüzgar gücü tahmin verisi

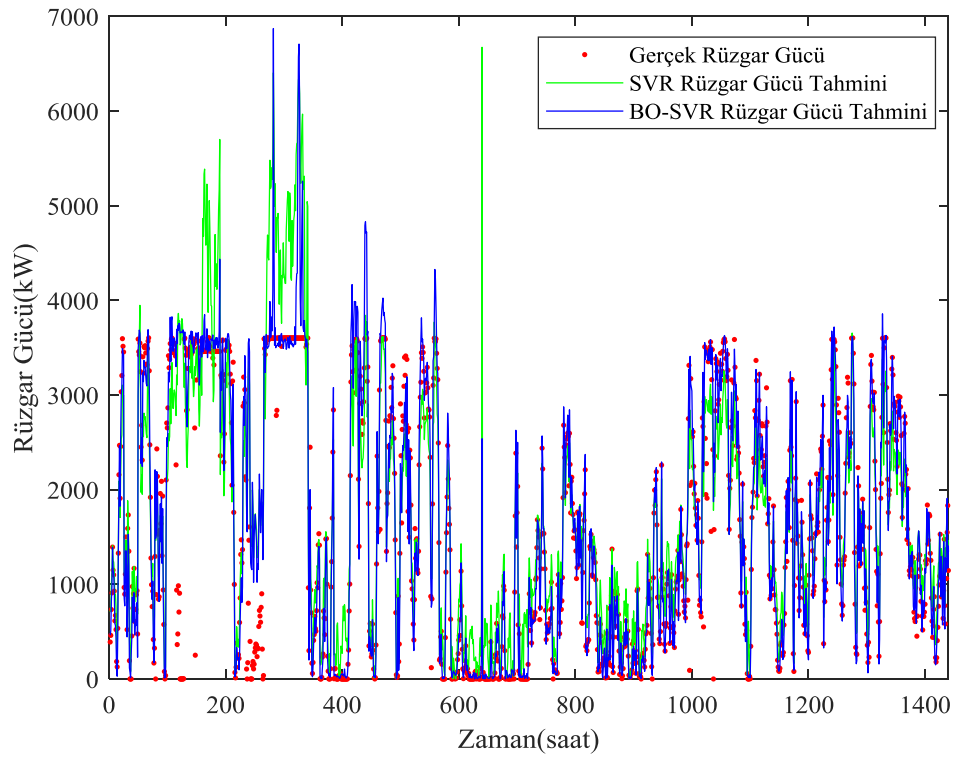
Şekil 5.8.' e bakıldığında BO-DT algoritmasının tahmin sonuçlarının DT algoritmasına oranla gerçek rüzgar gücü değerlerine daha yakın olduğu görülmektedir. Ayrıca test verisinin ilk yarısını oluşturan kış ve ilkbahar mevsimine ait Ocak ve Nisan ayında üretilen rüzgar gücündeki ani değişimin fazla olması sebebiyle algoritmaların bu verileri tahmin etmede zorlandığı görülmektedir.



Şekil 5.9. Test verisine ait üretilen güç, GPR ve BO-GPR rüzgâr gücü tahmin verisi

Şekil 5.9.' a bakıldığında GPR ve BO-GPR algoritmasının tahmin sonuçlarının gerçek değerlere yakın olduğu görülmüştür. Kış ve ilkbahar mevsimine ait Ocak ve Nisan ayında üretilen rüzgar gücündeki ani değişimin fazla olması sebebiyle algoritmaların bu verileri tahmin etmede zorlandığı görülmektedir. Ayrıca BO-GPR algoritmasının Ocak ayı tahmin değerleri BO-DT algoritmasına göre gerçek rüzgar gücü verilerine daha yakın bir sonuç vermiştir.

Şekil 5.10.'a bakıldığında BO-SVR algoritmasının SVR algoritmasından daha yakın sonuçlar vermesine rağmen BO-GPR ve BO-DT algoritmasına kıyasla daha başarısız olmuştur. Özellikle Ocak ayına ait gerçekleşen ve tahmin edilen güç değerlerine bakıldığında BO-SVR algoritmasının gerçek rüzgar gücü değerlerinden çok daha büyük tahmin değerleri ürettiği görülmüştür.



Şekil 5.10. Test verisine ait üretilen güç, SVR ve BO-SVR rüzgar gücü tahmin verisi

Çalışmada uygulanan GPR, DT, SVR ve optimize edilmiş BO-DT, BO-GPR, BO-SVR algoritmaları NRMSE, NMAE ve  $R^2$  performans kriterleri açısından değerlendirilmiştir. Tablo 5.9. kullanılan makine öğrenmesi algoritmalarına ait performans kriteri değerlerini göstermektedir.

Tablo 5.9. Makine öğrenmesi algoritmalarının performans kriterleri

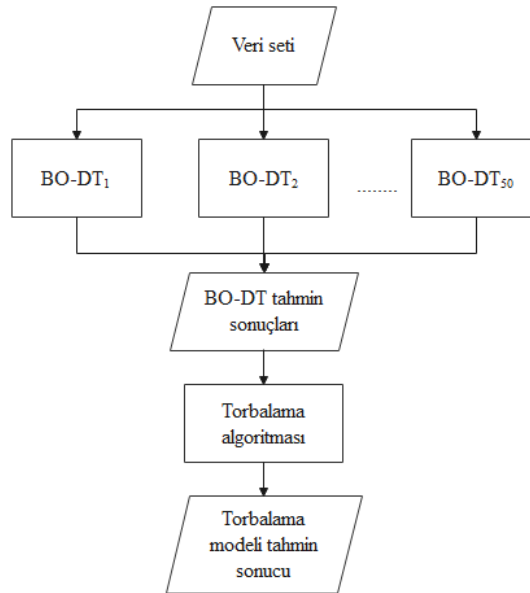
Makine öğrenmesi algoritmaları	Performans kriterleri		
	NRMSE (%)	NMAE (%)	$R^2$
GPR	13,210	5,968	0,846
BO-GPR	12,863	5,862	0,873
SVR	18,141	11,270	0,733
BO-SVR	13,821	5,700	0,845
DT	14,937	6,661	0,819
BO-DT	12,512	5,694	0,866



Varsayılan parametrelerle eğitilen algoritmalarla oranla Bayes optimizasyonu ile elde edilen parametrelerle eğitilen algoritmaların daha başarılı olduğu görülmüştür. Bayes optimizasyonu makine öğrenmesi algoritmalarının tahmin performansını iyileştirerek NRMSE ve NMAE değerlerinin düşmesini ve  $R^2$  değerinin artmasını sağlamıştır. BO-GPR, BO-DT ve BO-SVR algoritmalarının performans kriter değerlerine bakıldığında en az başarılı yöntemin BO-SVR olduğu görülmüştür. BO-DT algoritması en başarılı algoritma olmasına rağmen BO-GPR algoritması da yakın sonuçlar verdiği için her iki algoritma arasında seçim yapılamamış ve BO-DT ve BO-GPR temel öğrenen olarak seçilmiştir.

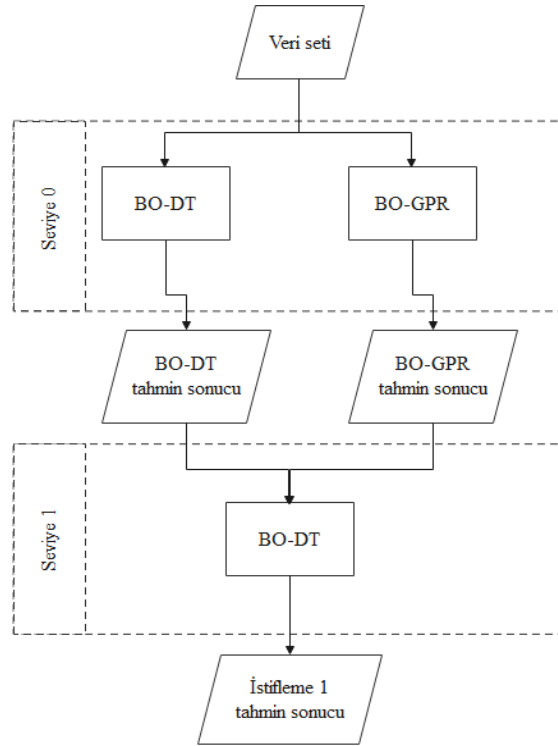
### 5.2.1.5. Makine öğrenmesi algoritmalarının topluluk öğrenmesi ile birleştirilmesi

Torbalama modelinde, BO-GPR ve BO-SVR'e kıyasla daha başarılı tahmin sonuçları veren BO-DT temel öğrenen olarak seçilmiştir. Torbalama modelinin tahmin sonucu eğitim veri setinin yarısına eşit ve ve homojen olarak seçilmiş veri ile eğitilen 50 tane BO-DT tahmin sonuçlarının ortalaması alınarak elde edilmiştir (Şekil 5.11.).  $BO - DT_1, BO - DT_2 \dots \dots BO - DT_{50}$  her biri veri setinden homojen bir şekilde seçilen farklı eğitim verisi ile eğitilmiş karar ağaçlarını temsil etmektedir.



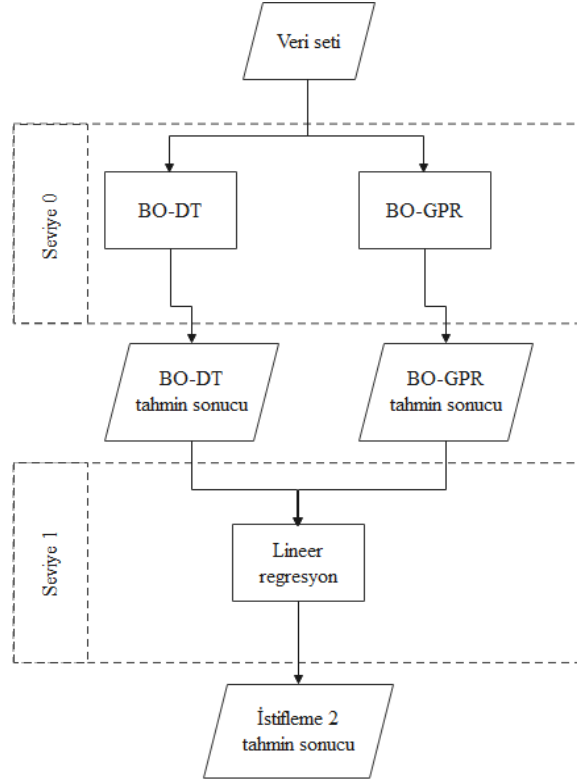
Şekil 5.11. Torbalama modeli

İstifleme modellerinde temel öğrenenler olan BO-GPR ve BO-DT algoritmaları meta öğrenen ile birleştirilerek her iki yöntemden daha iyi performansa sahip bir model oluşturulması amaçlanmıştır. Bu amaçla 2 seviyeli olmak üzere 3 farklı model oluşturulmuştur. Birinci istifleme modelinde BO-DT ve BO-GPR algoritmalarının tahmin sonuçları meta öğrenen BO-DT'nin girdisini oluşturmuş ve nihai tahmin sonucu elde edilmiştir (Şekil 5.12.).



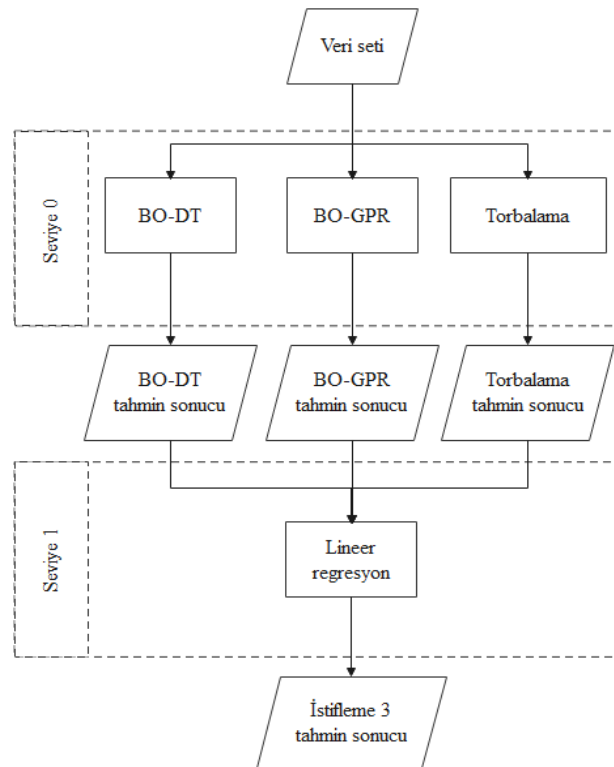
Şekil 5.12. İstifleme 1 modeli

İkinci istifleme modelinde yine BO-DT ve BO-GPR algoritmalarının tahmin sonuçlarını birleştirmede birinci modelden farklı olarak lineer regresyon kullanılmıştır (Şekil 5.13.).



Şekil 5.13. İstifleme 2 modeli

Üçüncü istifleme modelinde ise BO-DT ve BO-GPR temel öğrenenlerine ek olarak tahmin performansını arttıracacağı düşünülerek torbalama algoritması da eklenmiştir (Şekil 5.14.).



Şekil 5.14. İstifleme 3 modeli

Lineer regresyon algoritması, istifleme performansını artıran hızlı ve daha az karmaşık bir yöntem olduğu için meta öğrenen olarak kullanılmıştır. Tablo 5.10.'da topluluk öğrenmesi modelleri ve parametreleri gösterilmiştir.

Tablo 5.10. Topluluk öğrenmesi modellerinin parametre değerleri

Model	Parametreler
Torbalama	Temel öğrenen: BO-DT (Minimum yaprak örnek sayısı: 3, Maksimum dal düğüm sayısı: 220), Ağaç sayısı:50, Torbalama boyut yüzdesi: 50%
İstifleme1	Temel öğrenen: BO-DT (Minimum yaprak örnek sayısı: 3, Maksimum dal düğüm sayısı: 220), BO-GPR (Sigma: 0.11505) Meta öğrenen: BO-DT
İstifleme2	Temel öğrenen: BO-DT (Minimum yaprak örnek sayısı: 3, Maksimum dal düğüm sayısı: 220), BO-GPR (Sigma: 0.11505) Meta öğrenen: Lineer regresyon (Lamda: 1.8584e-05, Model türü: En küçük kareler)
İstifleme3	Temel öğrenen: BO-DT (Minimum yaprak örnek sayısı: 3, Maksimum dal düğüm sayısı: 220), BO-GPR, Torbalama Meta öğrenen: Lineer regresyon (Lambda: 1.8584e-05, Model türü: En küçük kareler)

Oluřturulan topluluk öğrenmesi algoritmaları ile test veri seti için rüzgar gücü tahmin deęerleri elde edilmiřtir.

#### **5.2.1.6. Tahmin modellerinin performanslarının deęerlendirilmesi**

Çalıřmada kullanılan tahmin modellerinin performansları  $R^2$ , NRMSE ve NMAE performans kriterleri ile ölçölmüřtür. Performans kriterleri her model için 24 saat zaman ufkunda hesaplanmıřtır. Ayrıca modellerin mevsime baęlı performans deęiřimlerini gösterebilmek performans kriterleri her mevsimden seçilen aylar için ayrı ayrı hesaplanmıřtır. Tablo 5.11. oluřturulan tahmin modellerinin performans kriterleri deęerlerini göstermektedir.

Tablo 5.11. Tahmin modellerinin performans kriterlerinin karşılaştırılması

Mevsimler	Performans kriterleri	Tahmin modelleri									
		GPR	DT	SVR	BO-GPR	BO-DT	BO-SVR	Toralama	İstifleme1	İstifleme2	İstifleme3
Kış (Ocak)	R <sup>2</sup>	0,868	0,707	0,502	0,869	0,809	0,723	<b>0,870</b>	0,801	0,782	0,823
	NRMSE (%)	13,563	21,902	27,518	13,460	15,940	21,459	<b>13,356</b>	17,666	17,039	15,875
	NMAE (%)	8,499	9,773	18,913	8,299	9,065	10,820	<b>8,259</b>	9,041	9,359	8,889
İlkbahar (Nisan)	R <sup>2</sup>	0,730	0,776	0,770	0,803	0,824	0,842	<b>0,888</b>	0,882	0,874	0,887
	NRMSE (%)	19,872	17,832	17,745	19,174	15,592	13,856	<b>13,830</b>	13,953	13,954	13,894
	NMAE (%)	8,709	7,916	10,354	8,512	6,320	5,407	<b>5,388</b>	5,636	5,408	5,980
Yaz (Temmuz)	R <sup>2</sup>	0,881	0,919	0,801	0,915	0,925	0,910	<b>0,926</b>	0,910	0,922	0,918
	NRMSE (%)	10,203	8,124	14,931	9,694	8,051	10,578	<b>8,001</b>	9,067	8,089	8,231
	NMAE (%)	3,876	3,992	8,950	3,864	3,481	3,513	<b>3,320</b>	3,802	3,321	3,689
Sonbahar (Ekim)	R <sup>2</sup>	0,907	0,876	0,861	0,908	0,908	0,905	<b>0,912</b>	0,900	0,909	0,906
	NRMSE (%)	9,202	11,890	12,373	9,124	10,464	9,390	<b>8,992</b>	9,342	9,034	10,034
	NMAE (%)	2,790	4,963	6,861	2,771	3,912	3,059	<b>2,650</b>	3,564	4,139	3,366
Ortalama	R <sup>2</sup>	0,846	0,819	0,733	0,873	0,866	0,845	<b>0,899</b>	0,873	0,872	0,883
	NRMSE (%)	13,210	14,937	18,141	12,863	12,512	13,821	<b>11,045</b>	12,507	12,029	12,009
	NMAE (%)	5,968	6,661	11,270	5,862	5,694	5,700	<b>4,880</b>	5,511	5,531	5,482

Tahmin modelleri sonuçlarına bakıldığında torbalama ve istifleme algoritmalarının optimize edilmiş GPR, SVR, DT'nin, torbalama ve istifleme modellerinin tahmin performansını artırdığı görülmüştür. BO-DT algoritması torbalama algoritması ile birleştirildiğinde NRMSE değeri %1,467 ve NMAE değeri % 0,814 azalmış  $R^2$  değeri ise 0,866'dan 0,899' a yükselmiştir. Çalışmada oluşturulan topluluk öğrenmesi modelleri kısa dönem rüzgar enerjisi tahmininde tekli algoritmalarından daha başarılı olmuştur. Mevsimsel performansa bakıldığında ise kış ve ilkbahar mevsimlerinde tahmin hatalarının daha yüksek olduğu görülmektedir. Bunun nedeni, bu mevsimlerde yaşanan kesinti ve arızalar nedeniyle rüzgar gücündeki ani değişikliklerdir. Yaz ve sonbahar mevsimlerinde aykırı veriler buna bağlı ani değişimler daha az olduğu için modellerin performansı bu mevsimlerde daha yüksek olmuştur. Mevsimsel ve ortalama performans kriter değerleri incelendiğinde, önerilen yöntemler arasında en iyi performansa sahip torbalama modelinin  $R^2$ , NRMSE ve NMAE değerleri sırasıyla 0,899, %11,045 ve %4,880 olarak ölçülmüştür.

## **5.2.2. Önerilen model 2'nin uygulama adımları**

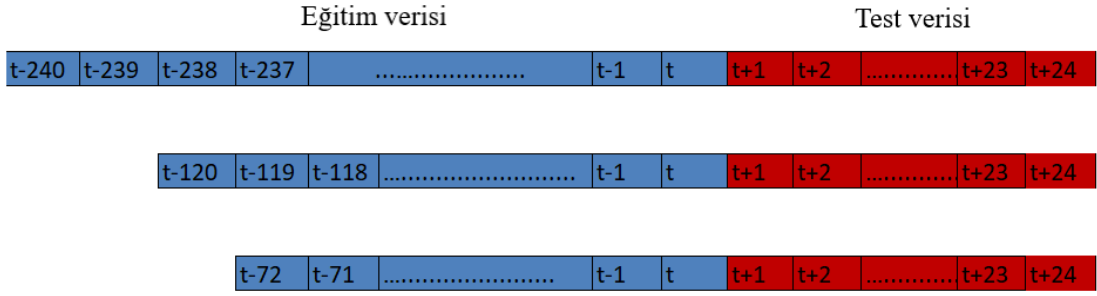
### **5.2.2.1. Verilerin toplanması**

Model saatlik rüzgar hızı, rüzgar yönü, sıcaklık, atmosferik basınç, bağıl nem ve rüzgar gücü parametrelerinden oluşan veri seti üzerinde uygulanmıştır. Veri setinin özellikleri 5.2.1.1 'de açıklanmıştır.

### **5.2.2.2. Verilerin düzenlenmesi**

Veriler ilk olarak eğitim ve test verisine ayrılmıştır. Ancak statik modeldeki gibi test verilerini tahmin etmek için sabit bir eğitim verisi kullanmak yerine zamana bağlı değişen eğitim verileri kullanılmıştır. Dinamik modelde statik modelde de kullanılan farklı mevsimlere ait günlerden oluşan test verilerini tahmin etmek için test verisindeki günlerin önceki zamanına ait veriler kullanılmış ve model her test gününü tahmin etmek için yeniden eğitilmiştir. Dinamik modelde en az veri ile ağın

eğitilmesi amaçlandığından ilk olarak önceki 1 gün (24 saat) ve 2 güne (48 saat) ait verilerle ağ eğitilmiş ancak tahmin performansının oldukça düşük olduğu görülmüştür. Bu nedenle eğitim verileri test verisi için seçilen günlerin önceki 3 (72 saat), 5 (120 saat) ve 10 (240 saat) gününe ait rüzgar hızı, rüzgar yönü, sıcaklık, atmosferik basınç ve rüzgar gücü değerlerinden oluşturulmuştur (Şekil 5.15.).



Şekil 5.15. LSTM ağlarının eğitim ve test verisi

Farklı uzunluktaki eğitim verisi ile LSTM ağını eğiterek eğitim verisinin uzunluğunun model performansına etkisinin araştırılması amaçlanmıştır. Test verisi ise her mevsimi temsilen belirlenen aylara ait rastgele seçilen toplam 60 günlük veriyi kapsamaktadır ve test verisi özellikleri bölüm 5.2.1.2' de açıklanmıştır. Eğitim ve test verisindeki eksik veriler K-NN algoritması ile tamamlanmış ve min-maks normalizasyonu ile normalize edilmiştir. Rüzgar yönü verileri normalize edilirken önce dereceden radyana çevrilmiş daha sonra sinüs ve kosinüs değerleri hesaplanmıştır.

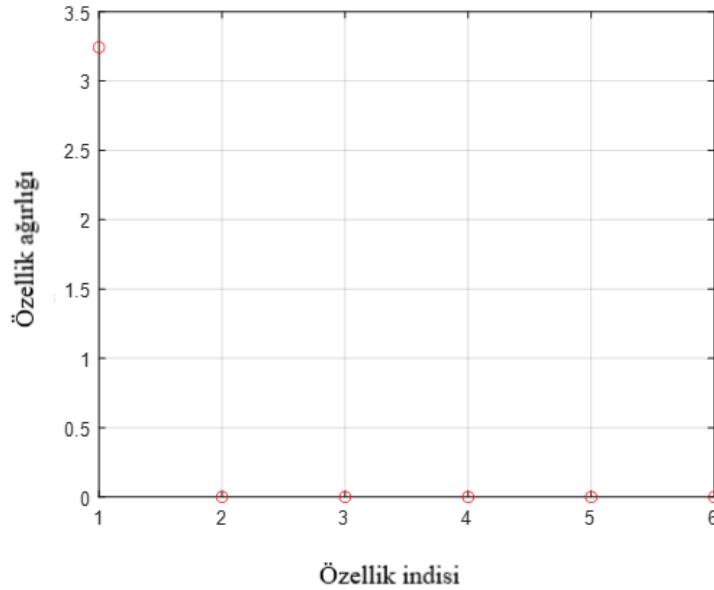
Geçmiş rüzgar gücü verileri ile beraber girdi olarak kullanılacak özellik seçimi yapılmıştır. PCA ve Sıralı Özellik Seçimi (SFS), özellik seçimi için klasik ve popüler yöntemlerdir. Bununla birlikte PCA, daha düşük boyutlara indirgenğinde bilgi kaybına neden olabilmekte ve SFS, diğer özellikler eklendikten sonra işe yaramaz hale gelen özellikleri kaldıramayabilir (Jin ve Deng, 2018). Bunun aksine NCA veri koşullarından etkilenmeden boyut indirgeme işlemi sırasında hiçbir bilginin kaybolmaması avantajına sahip olduğu için özellik seçiminde kullanılmıştır.



NCA algoritması eğitim verileri test verisine bağlı olarak değiştiği için her eğitim verisine ayrı ayrı uygulanmıştır. Aşırı öğrenmeyi engellemek için kullanılan düzeltme parametresi örnek sayısına bağlı olarak her eğitim verisi için ayrı ayrı hesaplanmıştır. Özellik ağırlıklarını tahmin etmek için kullanılan çözücü eğitim veri seti büyüklüğü 1000 örnekten az olduğu için Sınırlı bellekli BFGS (IBFGS) olarak ayarlanmıştır. Algoritmaya göre ağırlığı sıfırdan büyük olan özellikler eğitim veri setinde girdi parametresi olarak seçilmektedir. Tahmin edilen ağırlıkların test verisi üzerindeki doğruluğu ölçen kayıp fonksiyonu olarak ise Ortalama Mutlak Sapma (MAD) kullanılmıştır (Denklem 5.1).

$$l(y_i, y_j) = |y_i - y_j| \quad (5.1)$$

Şekil 5.16. 'da bir eğitim verisine uygulanan NCA algoritması ile hesaplanan özellik ağırlıkları yer almaktadır.



Şekil 5.16. NCA algoritması ile belirlenen özellik ağırlıkları

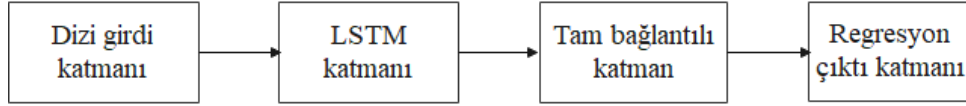
Şekil 5.16'da özellik numaraları 1,2,3,4,5,6 sırasıyla rüzgar hızı, rüzgar yönü sinüs değerini, rüzgar yönü kosinüs değerini, atmosferik basınç, sıcaklık ve bağıl nem parametrelerini ifade etmektedir.

NCA algoritması ile hesaplanan özellik ağırlıklarına bakıldığında birinci özellik olan rüzgar hızı 3,242 değerini diğerleri ise 0 değerini almıştır. Bu durumda hesaplanan kayıp fonksiyonu değeri 0,0012 olmuştur. Tüm eğitim veri setleri için NCA algoritması ile ağırlıklar hesaplanmış ve rüzgar hızının ağırlığının 1,882 ile 3,810 arasında değiştiği diğerlerinin ise 0 çıktığı görülmüştür. Bu nedenle özellik olarak rüzgar hızı seçilmiş ve LSTM ağırları önceki 72 saat, 120 saat ve 240 saatin rüzgar hızı ve rüzgar gücü verileriyle eğitilmiştir.

### 5.2.2.3. LSTM algoritmasının uygulanması

Gelecekte üretilecek rüzgar gücü meteorolojik değişkenlere bağlı olduğu kadar üretilen rüzgar gücünün önceki değerine de bağlıdır. Tekrarlayan sinir ağırları verilerin zamansal bilgilerini göz önünde bulunduran ileri ve geri besleme bağlantılarına sahip olduğu için rüzgar enerjisi tahmininde sıklıkla kullanılmaktadır. Tekrarlayan sinir ağırları geri yayılım ve gradyan azalması algoritmalarını kullanırken gradyan kaybolma problemleriyle karşılaşmasını önlemek için LSTM ağırları geliştirilmiş ve birçok rüzgar tahmin çalışmasında uygulanmıştır. Girdi, geri besleme ve özel bir hafıza birimine sahip olan LSTM bilginin ağıdaki durumu belirli kapılar ile seçici olarak etkilemesine izin verdiği ve önemli veri bilgilerini uzun süre saklayabildiği için çok değişkenli zaman serilerini tahmin etmede oldukça başarılı bir algoritmadır. Diğer makine öğrenimi algoritmalarından farklı olarak LSTM dizi verileri ve çok değişkenli veriler ile çalışmaya uygundur ve çok adımlı tahminlerde kullanılabilen değişken uzunlukta diziler üretebilir.

Bu nedenle dinamik modelde doğrusal olmayan zaman serilerinin analizinde sıklıkla kullanılan LSTM ağı uygulanmıştır. Dinamik LSTM modelinde test kümesindeki verilerin tahmininde modeli yeniden oluşturmak için önceki zaman adımının gerçekleşen çıktı değeri eğitim veri kümesine eklenmiştir (Nguyen ve ark., 2019). Bu sayede LSTM ağı gelecek zaman adımındaki rüzgar gücünü tahmin etmek için yeni rüzgar gücü verileri ile sürekli olarak güncellenebilmektedir.



Şekil 5.17. Çalışmada kullanılan LSTM ağ mimarisi

Şekil 5.17.' de görüldüğü gibi LSTM modeli dizi girdi katmanı (sequence input layer), lstm katmanı (lstm layer), tam bağlantılı katman (full connected layer) ve regresyon çıktı katmanından (regression output layer) oluşmaktadır. Dizi girdi katmanında girdi boyutu LSTM modelleri için 7 olup önceki 72 saat, 120 saat ve 240 saate ait rüzgar hızı, rüzgar yönü (sinus ve kosinüs değerleri), sıcaklık, atmosferik basınç, bağıl nem, rüzgar gücü verilerinden oluşmaktadır. NCA algoritması ile indirgemiş LSTM ağı (NCA-LSTM) modelleri için ise girdi boyutu 2 olup önceki 72 saat, 120 saat ve 240 saate ait rüzgar hızı ve rüzgar gücü verilerinden oluşmaktadır.

Zaman serisindeki zaman adımları arasındaki uzun vadeli bağımlılıkları öğrenen LSTM katmanında ise gizli birim sayısı, aktivasyon fonksiyonu, başlangıç parametre değerleri ve çıktı özellikleri belirlenmiştir. LSTM katmanında dizi uzunluğundan bağımsız olarak önceki zamana ait bilgi içeren gizli birim sayısı 200, aktivasyon fonksiyonu Tanjant Hiperbolik (tanh) olarak tanımlanmıştır (Xiaoyun ve ark., 2016; Son ve ark, 2019; Liu ve ark., 2021). Başlangıç ağırlıklarının belirlenmesinde ortalaması 0, varyansı  $2 / (\text{girdi boyutu} + (4 \times \text{gizli birim sayısı}))$ ' na eşit düzgün bir dağılımdan rastgele örnekler alarak çalışan glorot fonksiyonu kullanılmıştır (Liu ve ark., 2021). 24 saat zaman ufkundaki rüzgar gücü tahmin verilerini içeren çıktı özelliği ise dizi olarak tanımlanmıştır.

Tam bağlantılı katmanda girdi bir ağırlık matrisi ile çarpılır ve ardından bir sapma vektörü eklenir. Regresyon çıktı katmanında ise tahmin edilen çıktı değerleri için RMSE hata değerleri hesaplanmıştır.

Optimizasyon yöntemi olarak stokastik gradyan azaltma yönteminden daha verimli olan ve her bir parametre için öğrenme oranını sürekli olarak güncelleyen 'Adam' kullanılmıştır (Son ve ark., 2019; Liu ve ark., 2021). Döngü sayısı (epoch) parametresi 150-500 aralığında ve 50 birimde bir artan şekilde 8 kez denenmiştir.

Doğrulama ve eğitim kaybı epoch değeri 250 olana kadar azaldığı ve 250'den sonra artmaya başladığı için en uygun epoch değeri 250 olarak belirlenmiş ve 250 epoch boyunca ağ eğitilmiştir. Gradyan kaybolmasını önlemek için, gradyan eşiği 1 olarak ayarlanmıştır. Başlangıç öğrenme oranı 0,005 olarak belirlenmiş ve 125 epoch sonra öğrenme hızı 0,2 değerindeki seyreltme oranı ile çarpılarak düşürülmüştür (Kumar ve ark., 2020; Son ve ark., 2019). Seyreltme oranı ile tam bağlantılı katmanlarda eşik değerinin altındaki düğümlerin azaltılması ile zayıf bilgilerin unutulması sağlanmıştır.

Birden çok zaman adımının değerlerini tahmin etmek için, zaman adımları birer birer tahmin edilmiş ve her tahminde ağ durumu tahmin edilen rüzgar gücünün gerçek değerinin girdi olarak eklenmesiyle güncellenmiştir.

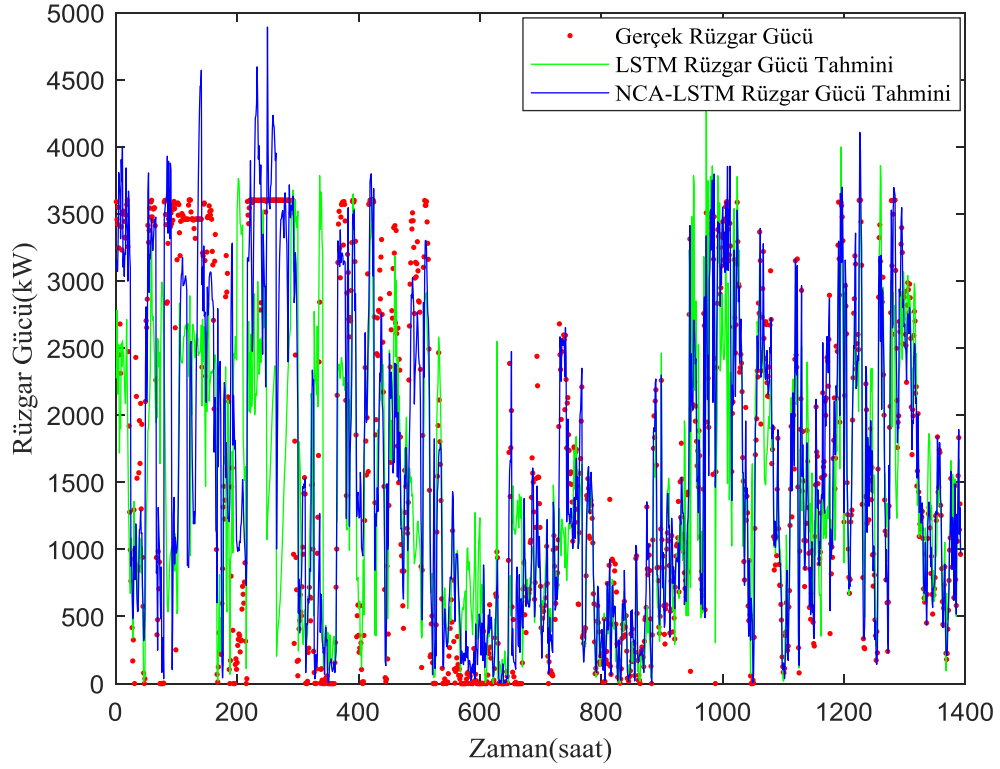
#### **5.2.2.4. LSTM modellerinin performanslarının değerlendirilmesi**

3 farklı uzunluktaki eğitim verisi kullanılarak eğitilen LSTM ağ modelleri ile NCA-LSTM tahmin sonuçları karşılaştırılmıştır. Modellerin kullandığı eğitim verisi LSTM ağ modelinin yanında LSTM-72 saat, LSTM-120 saat, LSTM-240, NCA-LSTM-72 saat, NCA-LSTM-120 saat, NCA-LSTM-240 saat şeklinde belirtilmiştir.

Şekil 5.12. LSTM-72 saat ağına ait tahmin sonuçları ve NCA-LSTM-72 saat ağına ait tahmin sonuçlarının gerçek rüzgar gücü değerler ile karşılaştırması görülmektedir.

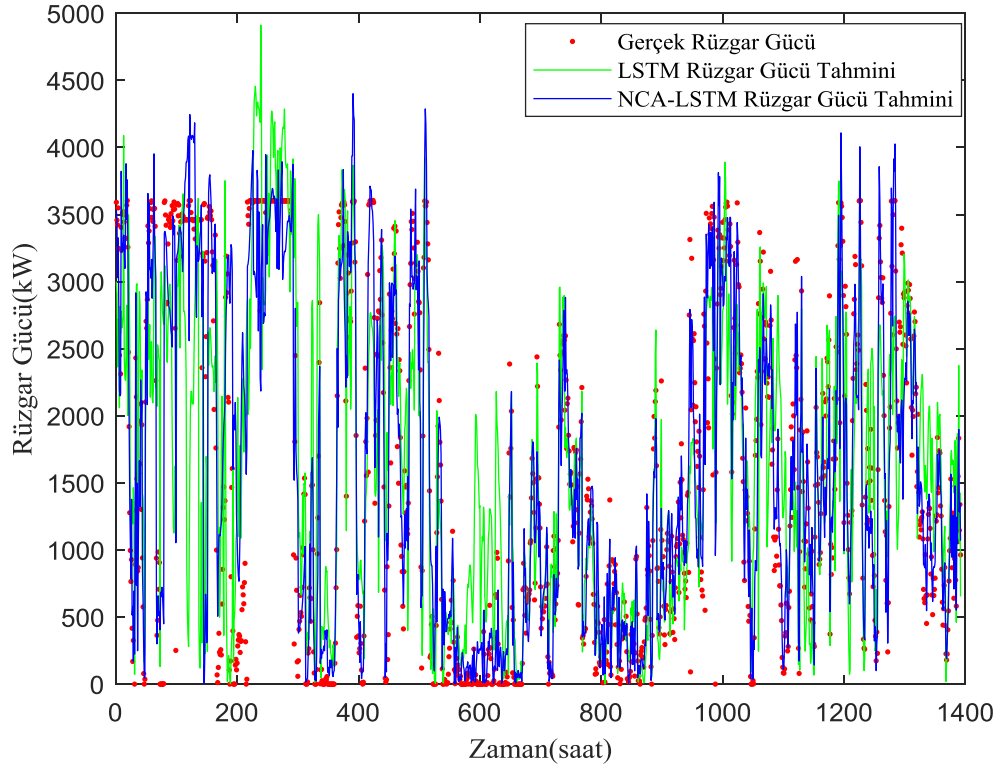
Şekil 5.18.'e bakıldığında test verisinin bazı günlerinde LSTM-72 saat ağının gerçek değere çok yakın tahmin değerleri ürettiği bazılarında ise gerçek değerden oldukça düşük veya yüksek sonuçlar verdiği görülmüştür. LSTM-72 saat modeli test verisinin ilk yarısında gerçek değerden düşük tahmin değerleri üretirken NCA-LSTM-72 saat modeli gerçek değerden daha yüksek değerler üretmiştir. Ayrıca NCA-LSTM-72 saat modelinin tahmin sonuçlarının LSTM-72 saat modeline göre gerçek rüzgar gücü değerlerine daha yakın olduğu görülmüştür. Her iki model de

rüzgar gücü değerlerinde dalgalanmalar olduğunda gerçek rüzgar gücü değerinden uzak tahmin sonuçları üretmiştir.



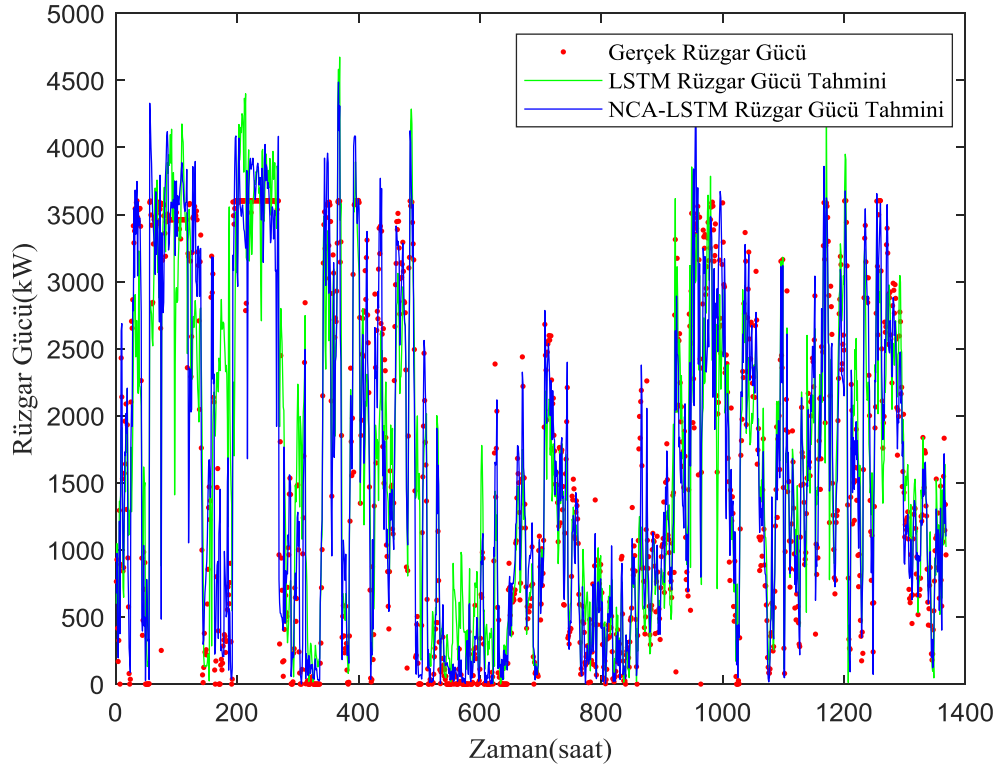
Şekil 5.18. Test verisine ait üretilen güç, LSTM-72 saat ve NCA-LSTM-72 saat rüzgar gücü tahmin verisi

Şekil 5.19. LSTM-120 saat ve NCA-LSTM-120 saat modellerinin tahmin sonuçlarının gerçek değerler ile karşılaştırılması verilmiştir. LSTM-120 saat modelinin tahmin sonuçları bazı durumlarda gerçek rüzgar gücünden oldukça büyük olurken bazı durumlarda da daha düşük olmuştur. NCA-LSTM-120 saat modeli ise gerçek değerlere daha yakın sonuçlar vermiştir. Her iki model de özellikle test verisinin ilk döneminde görülen rüzgar gücündeki ani düşüş ve artışın yaşandığı durumlarda gerçek değerden uzak tahmin sonuçları üretmiştir.



Şekil 5.19. Test verisine ait üretilen güç, LSTM-120 saat ve NCA-LSTM-120 saat rüzgar gücü tahmin verisi

Şekil 5.20. LSTM-240 saat ve NCA-LSTM-240 saat modellerinin tahmin sonuçlarını göstermektedir. Şekil 5.18.'e bakıldığında LSTM-240 saat modeli genel olarak gerçek rüzgar gücü değerlerinden daha büyük tahmin değerleri üretirken NCA-LSTM-240 saat modeli daha yakın sonuçlar üretmiştir. Özellikle rüzgar gücünün sürekli değiştiği ve 0 olarak ölçüldüğü zaman dilimlerinde NCA-LSTM-240 saat tahmin modeli doğru sonuçlar vermiştir.



Şekil 5.20. Test verisine ait üretilen güç, LSTM-240 saat ve NCA-LSTM-240 saat rüzgar gücü tahmin verisi

Ayrıca Şekil 5.18.-5.20. incelendiğinde LSTM-240 saat ve NCA-LSTM-240 saat modellerinin önceki 72 saat ve 120 saatlik veri ile eğitilen modellere oranla gerçek rüzgar gücü değerlerine daha yakın sonuçlar verdiği görülmektedir.

Tablo 5.12. LSTM ağları ile oluşturulan tahmin modellerinin 24 saatlik zaman ufkunda  $R^2$ , NRMSE ve NMAE performans kriterleri değerlerini göstermektedir. LSTM ve NCA-LSTM modellerinde test verisi için seçilen Ocak ayına ait önceki 3, 5 ve 10 güne ait yeterli veri olmadığı için modellerin performans değerlendirmesi sırasıyla 59, 58 ve 57 gün üzerinden yapılmıştır. LSTM ağlarının tahmin performanslarının mevsime bağlı değişimini göstermek için modellerin performans kriterleri her mevsimi temsilen seçilen aylar için ayrı ayrı hesaplanmıştır.

Tablo 5.12. LSTM tahmin modellerinin performans kriterlerinin karşılaştırılması

Mevsimler	Performans kriterleri	Tahmin modelleri					
		LSTM-72 saat	LSTM-120 saat	LSTM-240 saat	NCA-LSTM-72 saat	NCA-LSTM-120 saat	NCA-LSTM-240 saat
Kış (Ocak)	R <sup>2</sup>	0,448	0,599	0,593	0,601	0,599	<b>0,627</b>
	NRMSE (%)	34,930	28,746	28,778	30,853	28,285	<b>28,206</b>
	NMAE (%)	22,050	21,637	24,205	22,234	19,713	<b>18,836</b>
İlkbahar (Nisan)	R <sup>2</sup>	0,619	0,655	0,699	0,682	0,693	<b>0,784</b>
	NRMSE (%)	24,722	19,777	18,573	19,540	18,229	<b>17,031</b>
	NMAE (%)	14,592	13,842	11,559	10,423	9,845	<b>9,235</b>
Yaz (Temmuz)	R <sup>2</sup>	0,740	0,781	0,798	0,850	0,851	<b>0,892</b>
	NRMSE (%)	12,528	11,552	12,695	10,517	9,361	<b>9,196</b>
	NMAE (%)	8,389	8,848	8,273	6,324	6,877	<b>4,510</b>
Sonbahar (Ekim)	R <sup>2</sup>	0,712	0,735	0,711	0,754	0,851	<b>0,882</b>
	NRMSE (%)	14,356	18,067	16,100	12,971	13,285	<b>10,446</b>
	NMAE (%)	9,714	9,090	9,695	7,918	6,767	<b>5,615</b>
Ortalama	R <sup>2</sup>	0,630	0,692	0,700	0,722	0,748	<b>0,796</b>
	NRMSE (%)	21,634	19,535	19,037	18,470	17,290	<b>16,220</b>
	NMAE (%)	13,686	13,354	13,433	11,725	10,800	<b>9,549</b>

Tablo 5.12. incelendiğinde eğitim verisinin uzunluğu arttıkça LSTM ağlarının tahmin performansının arttığı görülmektedir. Test verisinin geçmiş 72 saatine ait rüzgar hızı, rüzgar yönü, sıcaklık, bağıl nem, atmosferik basınç ve rüzgar gücü verileri ile eğitilen LSTM-72 saat ağının R<sup>2</sup>, NRMSE, NMAE performans kriterleri sırasıyla 0,630, % 21,634, % 13,686 değerlerini alırken eğitim verisi 240 saate çıkarıldığında LSTM ağının R<sup>2</sup> değeri 0,07 artarken NRMSE % 2,597 ve NMAE % 0,253 azalmıştır. Bu durum eğitim verisinin uzunluğunun artması ile ağın daha fazla ve farklı sayıda örnek görmesi ile açıklanabilir. Ayrıca LSTM-120 saat ve LSTM-240 saat modellerinin birbirine yakın sonuçlar verdiği görülmüştür.



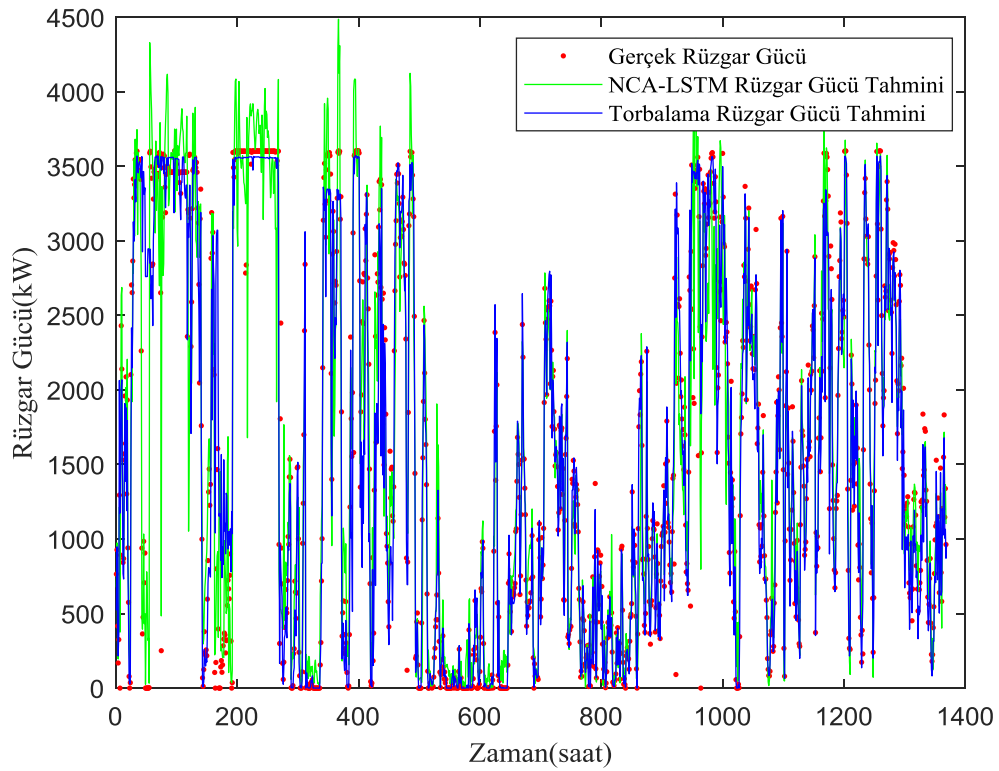
Tahmin modellerinde eğitim için gereğinden fazla sayıda veri kullanılması modelin performansını düşürebilmektedir. LSTM modelleri ile NCA-LSTM modellerinin sonuçları karşılaştırıldığında NCA-LSTM modellerinin daha başarılı olduğu görülmüştür. NCA algoritması veri boyutunu indirgeme işleminde bilgi kaybına izin vermemesi ve ağların girdi parametreleri içerisinde rüzgar gücünü en iyi açıklayan girdi ile eğitilmesine imkan sağladığı için LSTM modellerinin tahmin performansını arttırmıştır.

Ayrıca LSTM modellerinin tahmin performanslarının mevsimsel açıdan oldukça değişken olduğu görülmüştür. LSTM modelleri kış ve ilkbahar mevsimine ait rüzgar gücü verilerini tahmin etmekte zorlanmış, yaz ve sonbahar mevsimine ait verileri tahmin etmekte daha başarılı olmuştur. LSTM modelleri en iyi tahmin performansını yaz mevsimi verilerinde gösterirken en kötü performansı kış mevsiminde göstermiştir.

Oluşturulan LSTM ve NCA-LSTM modelleri ortalama performans kriterleri açısından karşılaştırıldığında en iyi tahmini test verisinin önceki 240 saate ait rüzgar hızı ve rüzgar gücü verileriyle eğitilen NCA-LSTM-240 saat modeli vermiştir. Modelin ortalama  $R^2$ , NRMSE ve NMAE değerleri sırasıyla 0,796, % 16,220 ve % 9,549 olarak hesaplanmıştır.

### **5.2.3. Önerilen modellerin karşılaştırılması**

Çalışmada Türkiye’de faaliyet gösteren bir rüzgar türbinin kısa dönem rüzgar gücünün tahmininde iki farklı tahmin modelinin geliştirilmesi ve uygulanması amaçlanmıştır. Modeller kullandığı eğitim verisi ve tahmin yapısı ile statik ve dinamik yapıda olmak üzere oluşturulmuştur. Statik yapıdaki modellerde en iyi sonucu veren torbalama modeli ve dinamik yapıdaki modellerde en iyi sonucu veren NCA-LSTM-240 saat modellerinin gerçek rüzgar gücü değerleri ile karşılaştırılması Şekil 5.21. ‘de verilmiştir.



Şekil 5.21. Test verisine ait üretilen güç, torbalama ve NCA-LSTM-240 saat rüzgar gücü tahmin verisi

Şekil 5.21.' e bakıldığında statik yapıdaki torbalama modelinin dinamik yapıdaki NCA-LSTM-240 saat modeline oranla gerçek rüzgar gücü değerlerine daha yakın sonuçlar verdiği görülmüştür. NCA-LSTM-240 saat modeli genel olarak gerçek rüzgar gücü değerlerinden daha yüksek tahmin sonuçları vermesine rağmen torbalama modeline oranla rüzgar gücündeki ani değişimlerin ve dalgalanmaların yaşandığı zaman dilimlerinde gerçek değere daha yakın tahmin sonuçları vermiştir. Buna karşılık durağan test verilerinin görüldüğü zaman dilimlerinde ise statik yapıdaki torbalama modeli NCA-LSTM-240 saat modeline kıyasla daha doğru tahmin sonuçları vermiştir. Dalgalanma görülen zaman diliminde dinamik NCA-LSTM-240 saat modelinin daha başarılı olması modelin yakın zamana ait ve dinamik bir eğitim veri setine sahip olması ve gerçek rüzgar gücü değerleri ile ağırlıklı olarak sürekli olarak güncellenmesi ile açıklanabilir. Statik yapıdaki torbalama modeli ise dinamik modele oranla her mevsimden veri içerdiği ve sayıca çok daha fazla veri ile eğitildiği için genel olarak gerçek güç değerlerine daha yakın sonuçlar vermiştir.

Tablo 5.13. Torbalama ve NCA-LSTM tahmin modellerinin karşılaştırılması

Mevsimler	Performans kriterleri	Tahmin modelleri	
		Torbalama	NCA-LSTM-240 saat
Kış (Ocak)	R <sup>2</sup>	0,870	0,627
	NRMSE (%)	13,356	28,206
	NMAE (%)	8,259	18,836
İlkbahar (Nisan)	R <sup>2</sup>	0,888	0,784
	NRMSE (%)	13,830	17,031
	NMAE (%)	5,388	9,235
Yaz (Temmuz)	R <sup>2</sup>	0,926	0,892
	NRMSE (%)	8,001	9,196
	NMAE (%)	3,320	4,510
Sonbahar (Ekim)	R <sup>2</sup>	0,912	0,882
	NRMSE (%)	8,992	10,446
	NMAE (%)	2,650	5,615
Ortalama	R <sup>2</sup>	0,899	0,796
	NRMSE (%)	11,045	16,220
	NMAE (%)	4,880	9,549

Tablo 5.13.' te de görüldüğü her iki tahmin modellerinin performansları mevsimsel olarak benzerlik göstermektedir. Her iki tahmin modeli de en yüksek hata değerlerine kış mevsiminde en düşük hata değerlerine ise yaz mevsiminde ulaşmıştır. Yaz mevsimine ait rüzgar gücü verilerinin tahminin de her iki yöntem de başarılı sonuçlar vermiştir. Bu durum kış ve ilkbahar mevsiminde yaz ve sonbahar mevsimine oranla daha fazla aykırı verinin görülmesi ile açıklanabilir. Arıza, kesinti ve mantıksız verilerden oluşan aykırı verilerle eğitilen modeller öğrendiği hatalı veriler sebebiyle kötü tahmin sonuçları üretebilmektedir.

## BÖLÜM 6. SONUÇ VE ÖNERİLER

Son yıllarda giderek önemini arttıran rüzgâr gücünün durağan olmayan ve değişken yapısı elektrik piyasalarında belirsizliğe yol açarak sistem verimsizliğine sebep olmaktadır. Enerji piyasasındaki belirsizlik santrallerin piyasaya sundukları üretim tahminleri ve ürettikleri rüzgâr gücü arasındaki farktan kaynaklanmaktadır. Piyasadaki belirsizliğin azaltılması, santrallerin gelirlerinin artırılması ve sisteme daha fazla rüzgâr gücünün entegre edilebilmesi rüzgâr tahmin modellerinin doğruluklarının artırılması ile gerçekleştirilebilir.

Çalışmada kısa dönem rüzgâr gücü tahmini için iki farklı yapıda tahmin modeli önerilmiş ve önerilen modeller Türkiye’de faaliyet gösteren bir rüzgâr türbinin 24 saatlik zaman ufkunda rüzgâr gücünü tahmininde uygulanmıştır. Önerilen modellerin tahmin performansını mevsimsel olarak ölçebilmek için test verileri her mevsimden veri içerecek şekilde oluşturulmuştur.

6 adımdan oluşan statik yapıdaki birinci modelde kısa dönem rüzgâr gücü tahmini için ilk adımda rüzgâr hızı, rüzgâr yönü, sıcaklık, atmosferik basınç ve bağıl nem girdilerinden oluşan veri seti oluşturulmuştur. Sonraki adımda her mevsimden veri içeren sabit bir eğitim verisi oluşturulmuş eksik adımlar K-NN algoritması ile tamamlanarak normalize edilmiştir. Üçüncü adımda düzenlenen verilerin tahmininde rüzgâr gücü tahmininde sıklıkla kullanılan DT, SVR, GPR algoritmalarından yararlanılmış ve algoritmaların optimum hiper parametre değerleri literatürdeki rüzgâr tahmin çalışmalarından farklı olarak Bayes optimizasyonu algoritması ile elde edilmiştir. Ardından optimize edilen algoritmalar olan BO-DT, BO-SVR, BO-GPR arasında en iyi performansı gösteren algoritma olarak BO-DT ve BO-GPR seçilmiştir. Sonraki adımda seçilen algoritmalarından BO-DT topluluk öğrenmesi yöntemlerinden olan torbalama ile birleştirilmiştir. BO-DT algoritmasından sonra ikinci en iyi algoritma olan BO-GPR ile BO-DT ise istifleme algoritması ile 3 farklı şekilde

birleştirilmiş ve çalışmanın son adımında oluşturulan tüm tahmin modelleri karşılaştırılmıştır.

Dinamik yapıdaki ikinci model 4 adımdan oluşturulmuş ve birinci modeldeki veri seti kullanılmıştır. İkinci adımda üretilen rüzgar gücünün önceki üretim değerlerine ve meteorolojik verilere bağlı değişimi verilerin dikkate alınarak dinamik bir eğitim verisi oluşturulmuştur. Dinamik modelde en az veri ile tahmin yapılması amaçlandığı için tahmin verileri olarak test verisi için seçilen günün önceki 72, 120, 240 saatine ait geçmiş rüzgar gücü ve meteorolojik veriler olan rüzgar hızı, rüzgar yönü, sıcaklık, atmosferik basınç ve nem verileri belirlenmiş ve eksik veriler K-NN algoritması ile tamamlanarak normalize edilmiştir. Faydalı girdi parametrelerinin belirlenmesi ve aşırı öğrenmeyi engellemek için ise diğer rüzgar gücü tahmin çalışmalarından farklı olarak NCA algoritmasından yararlanılmış ve meteorolojik parametreler olan rüzgar hızı, rüzgar yönü, sıcaklık, atmosferik basınç ve nem parametreleri içerisinde faydalı olanlar seçilmiştir. Sonraki adımda zaman serisi modellerinde başarılı sonuçlar veren LSTM algoritması kullanılarak test verileri tahmin edilmiştir. Dinamik LSTM modelinin gelecek rüzgar gücünün tahmini için önceki adımın gerçekleşen rüzgar gücü verisi ile beslenerek modelin her adımda güncellenmesi sağlanmıştır. Son adımda ise LSTM modelleri ve NCA-LSTM modellerinin tahmin performansları karşılaştırılmıştır.

Statik tahmin modelleri içerisinde %11,416 NRMSE, %4,880 NMAE ve 0,899 R<sup>2</sup> ile en başarılı model torbalama olurken dinamik modeller içerisinde en başarılı model 0,796 R<sup>2</sup>, %16,220 NRMSE ve %9,549 NMAE ile önceki 240 saat verisi ile eğitilen NCA-LSTM-240 saat modeli olmuştur.

Çalışmanın sonuçları şu şekilde özetlenebilir:

- Statik modellerde hiper parametrelerin aldığı değerlerin tahmin performansına etkisi Bayes optimizasyonu ile gösterilmiş ve optimize edilen hiper parametrelere sahip algoritmaların tahmin performansı artmıştır.
- Topluluk öğrenmesi algoritmaları ile oluşturulan modeller test verisinin tahmininde tekli algoritmalarından daha başarılı olmuştur.

- Dinamik modellerde geçmişe dayalı eğitim verisinin uzunluğunun artması ile tahmin performansında artış görülmüştür.
- NCA algoritmasının rüzgar gücünü en iyi açıklayan faydalı girdi parametrelerini bilgi kaybı olmadan seçme özelliği sayesinde NCA-LSTM modelleri LSTM modellerine oranla daha yüksek tahmin performansına sahip olmuştur.
- Statik yapıdaki torbalama modeli dinamik yapıdaki NCA-LSTM-240 saat modeline oranla hem mevsimsel hem de ortalama performans kriterleri açısından daha başarılı olmuştur. Statik yapıdaki modelin her mevsimden çok sayıda veri ile eğitilmesi modelin tahmin hatasının azalmasını sağlamıştır.
- Statik modele oranla çok daha az sayıda veri ile eğitilmesine rağmen dinamik model bazı zaman dilimlerinde gerçek rüzgar gücü değerlerine oldukça yakın ve kabul edilebilir sonuçlar vermiştir.
- Statik yapıdaki torbalama modeli durağan verileri tahmin etmede daha başarılı olurken dinamik yapıdaki NCA-LSTM-240 saat modeli ise dalgalanmaların olduğu verilerin tahmininde daha başarılı olmuştur.
- İki modelin de performansı mevsimsel olarak çok fazla değişkenlik göstermiş ve en iyi tahmin performansını yaz mevsiminde en kötü performansını ise kış mevsiminde göstermişlerdir.
- Aykırı veriler modellerin hatalı veriler ile eğitilmesine ve kötü tahmin sonuçları üretmesine neden olabilmektedir. Bu sebeple her mevsim için seçilen aydaki aykırı verilerin yoğunluğu modellerin mevsimsel tahmin performansının değişken olmasında etkili olmuştur.

Gelecek çalışması olarak,

- Önerilen birinci modelde kullanılan topluluk öğrenmesi algoritmaları olan torbalama ve istifleme yöntemlerinin optimum hiper parametreleri Bayes optimizasyonu algoritması ile araştırılabilir.
- Önerilen ikinci modelde eğitim verilerinin seçiminde modelin geçmiş ne kadar veriye bakması gerektiği otokolerasyon fonksiyonu gibi yöntemlerle belirlenebilir.

- Önerilen ikinci modelde LSTM algoritması hiper parametre değerleri olan öğrenme oranı, katman sayısı, gizli birim sayısının sezgisel olarak belirlenmesi yerine Bayes optimizasyonu ile belirlenebilir.
- Her iki modelde de rüzgar güç eğrileri ile gösterilen aykırı verilerin saptanması ve işlenmesi gerçekleştirilebilir. Bu sayede önerilen modellerin tahmin performansında artış sağlanabilir.

## KAYNAKLAR

- Aksoy, F. 2018. Short-Term Forecasting of Wind Power Production Using Machine Learning and Deep Learning Methods. İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik Mühendisliği, Yüksek Lisans Tezi.
- Alade, I.O., Abd Rahman, M.A., Saleh, T.A. 2019. Predicting the specific heat capacity of alumina/ethylene glycol nanofluids using support vector regression model optimized with Bayesian algorithm. *Solar Energy*, 183: 74–82.
- Arthur, C. K., Temeng, V. A., Ziggah, Y. Y. 2020. Novel approach to predicting blast-induced ground vibration using Gaussian process regression. *Engineering with Computers*, 36(1): 29-42.
- Bağcı, E. 2019. Türkiye’de Yenilenebilir Enerji Potansiyeli, Üretimi, Tüketimi ve Cari İşlemler Dengesi İlişkisi. *R&S- Research Studies Anatolia Journal*, 2(4): 101-117.
- Banik, R., Das, P., Ray, S., Biswas, A. 2020. Wind power generation probabilistic modeling using ensemble learning techniques. *Materials Today: Proceedings*, 26: 2157–2162.
- Bayraç, H. N. 2011. Küresel rüzgâr enerjisi politikaları ve uygulamaları. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 30(1): 37-57.
- Beretta, L., Santaniello, A. 2016. Nearest neighbor imputation algorithms: a critical evaluation. *BMC medical informatics and decision making*, 16(3): 197-208.
- BP, 2021. <https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html>, Erişim tarihi: 08.01.2021.
- Breiman, L., Friedman, J.H., Olshen, R.A. 1984. CART: Classification and Regression Trees. *Biometrics*, 40: 358–380.
- Breiman, L., 1996. Bagging predictors. *Machine learning*, 24: 123–140.
- Brochu, E., Cora, V. M., De Freitas, N. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv preprint arXiv:1012.2599



- Cali, U., Sharma, V. 2019. Short-term wind power forecasting using long-short term memory based recurrent neural network model and variable selection. *International Journal of Smart Grid and Clean Energy*, 8(2): 103-110.
- Cantürk, S. 2018. Bir Rüzgar Çiftliğinden Yapay Sinir Ağlarıyla Kısa Süreli Elektrik Üretim Tahmini. Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Temiz Tükenmez Enerjiler, Yüksek Lisans Tezi.
- Cao, Q., Ewing, B. T., Thompson, M. A. 2012. Forecasting wind speed with recurrent neural networks. *European Journal of Operational Research*, 221(1): 148-154.
- Chen, N., Qian, Z., Nabney, I., Meng, X. 2014. Wind power forecasts using Gaussian processes and numerical weather prediction. *IEEE Trans. Power Syst.*, 29(2): 656-665.
- Cornejo-Bueno, L., Garrido-Merchán, E.C., Hernández-Lobato, D., Salcedo-Sanz, S., 2018. Bayesian optimization of a hybrid system for robust ocean wave features prediction. *Neurocomputing*, 275: 818–828.
- Çevik, H.H., 2019. Yapay Zeka Teknikleri Kullanılarak Kısa Dönem Rüzgar Gücünün Çok Katmanlı Tahmini. Fen Bilimleri Enstitüsü, Elektrik-Elektronik Mühendisliği, Doktora Tezi.
- De'ath, G., Fabricius, K. E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, 81(11): 3178-3192.
- Deng, H., Fannon. D., Eckelman. M.J., 2018. Predictive modeling for US commercial building energy use: a comparison of existing statistical and machine learning algorithms using CBECS microdata. *Energy Build*, 163: 34-43.
- Dietterich, T. G. 2000. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, Springer, Berlin, Heidelberg, 1-15.
- Dong, B., Cao, C., & Lee, S. E. 2005. Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings*, 37(5): 545-553.
- Dong, D., Sheng, Z., Yang, T. 2018. Wind power prediction based on recurrent neural network with long short-term memory units. *International Conference on Renewable Energy and Power Engineering*, 34-38.
- Dongmei, Z., Yuchen, Z., Xu, Z. 2011. Research on wind power forecasting in wind farms. In *2011 IEEE Power Engineering and Automation Conference*, 1: 175-178.
- 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.

- Erisen, B. 2019. Wind turbine scada dataset, <https://www.kaggle.com/berkerisen/wind-turbine-scada-dataset>, Erişim tarihi: 10.08.2020.
- ETKB, 2019. <https://enerji.gov.tr/bilgi-merkezi-enerji-elektrik>, Erişim tarihi: 04.01.2021
- EWEA, 2011. Wind in our Sails: The Coming of Europe's Offshore Wind Energy Industry. [http://www.ewea.org/fileadmin/files/library/publications/reports/Offshore\\_Report.pdf](http://www.ewea.org/fileadmin/files/library/publications/reports/Offshore_Report.pdf), Erişim tarihi: 12.02.2021.
- Fu, C., Li, G. Q., Lin, K. P., Zhang, H. J. 2019. Short-term wind power prediction based on improved chicken algorithm optimization support vector machine. *Sustainability*, 11(2): 512.
- Gao, Z., Shi, J., Li, H., Chen, C., Tan, J., Liu, L. 2020. Substation Load Characteristics and Forecasting Model for Large-scale Distributed Generation Integration. In *IOP Conference Series: Materials Science and Engineering*, 782(3).
- Garcia, A., De-La-Torre-Vega, E. 2009. A Statistical wind power forecasting system—A Mexican wind-farm case study. In *European Wind Energy Conference & Exhibition—EWEC Parc Chanot*.
- Giebel, G., Draxl, C., Brownsword, R., Kariniotakis, G., Denhard, M. 2011. The state-of-the-art in short-term prediction of wind power. A literature overview.
- Glorot, X., Bengio, Y. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 249-256.
- Görgel, P., Kavlak, E. 2020. Uzun Kısa Süreli Hafıza ve Evrimsel Sinir Ağları ile Rüzgar Enerjisi Üretim Tahmini. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 11(1): 69-80.
- GWEC, 2017. Global Wind Report, <https://gwec.net/publications/global-windreport-2/>. Erişim tarihi: 25.02.2021
- GWEC, 2021. Global Wind Report <https://gwec.net/global-wind-report-2021/>. Erişim tarihi: 28.02.2021
- Heinermann, J., Kramer, O. 2016. Machine learning ensembles for wind power prediction. *Renewable Energy*, 89: 671-679.
- Heo, Y., Zavala, V.M., 2012. Gaussian process modeling for measurement and verification of building energy savings. *Energy Build*, 53: 7-18.
- Hochreiter, S., Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, 9(8): 1735-1780.

<https://ourworldindata.org/renewable-energy>, Erişim tarihi: 20.02.2021

<https://irena.org/wind>, Erişim tarihi: 30.01.2021

IRENA (2019), Future of wind: Deployment, investment, technology, grid integration and socio-economic aspects (A Global Energy Transformation paper), International Renewable Energy Agency, Abu Dhabi.

Ito, K., Nakano, R. 2003. Optimizing support vector regression hyperparameters based on cross-validation. In Proceedings of the International Joint Conference on Neural Networks, 3, 2077-2082.

Jin, M., Deng, W. 2018. Predication of different stages of Alzheimer's disease using neighborhood component analysis and ensemble decision tree. Journal of neuroscience methods, 302: 35-41.

Jung, J., Broadwater, R. P. 2014. Current status and future advances for wind speed and power forecasting. Renewable and Sustainable Energy Reviews, 31: 762-777.

Karık, F., Sözen, A., İzgeç, M.M. 2017. Rüzgâr gücü tahminlerinin önemi: Türkiye elektrik piyasasında bir uygulama. Politeknik Dergisi. 20(4): 851-861.

Kerem, A. 2018. Rüzgar Parametrelerinin Değişiminin İzlenmesi ve Yapay Zeka Algoritmaları Kullanarak Tahmini. Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik Elektronik Mühendisliği, Doktora Tezi.

Kim, T., Ko, W., Kim, J. 2019. Analysis and impact evaluation of missing data imputation in day-ahead PV generation forecasting. Applied Sciences, 9(1): 204.

Kingma, D. P., Ba, J. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Köksoy, C.E. 2015. Improved Wind Power Forecasting Using Combination Methods. Orta Doğu Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği, Yüksek Lisans Tezi.

Kumar, D., Abhinav, R., Pindoriya, N. 2020. An ensemble model for short-term wind power forecasting using deep learning and gradient boosting algorithms. In 2020 21st National Power Systems Conference (NPSC), 1-6.

Lange, M., Focken, U. 2008. New developments in wind energy forecasting. In 2008 IEEE power and energy society general meeting-conversion and delivery of electrical energy in the 21st century, 1-8.

Lee, D., Baldick, R. 2014. Short-term wind power ensemble prediction based on Gaussian processes and neural networks. IEEE Trans Smart Grid, 5(1): 501-510.

- Lewis, R. J. 2000. An introduction to classification and regression tree (CART) analysis. İçinde: Annual meeting of the society for academic emergency medicine in San Francisco, California, 14.
- Li, C., Lin, S., Xu, F., Liu, D., Liu, J., 2018. Short-term wind power prediction based on data mining technology and improved support vector machine method: a case study in Northwest China. *J. Cleaner Prod.*, 205(4): 909-922.
- Li, L. L., Zhao, X., Tseng, M. L., Tan, R. R. 2020. Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. *Journal of Cleaner Production*, 242: 118447.
- Li, Q., Meng, Q., Cai, J., Yoshino, H., Mochida, A. 2009. Applying support vector machine to predict hourly cooling load in the building. *Appl Energy*, 86(10): 2249-2256.
- Liu, H., Chen, D., Lin, F., Wan, Z. 2021. Wind Power Short-Term Forecasting Based on LSTM Neural Network with Dragonfly Algorithm. In *Journal of Physics: Conference Series*, 1748 (3).
- Liu, X., Zhou, J., & Qian, H. 2021. Short-term wind power forecasting by stacked recurrent neural networks with parametric sine activation function. *Electric Power Systems Research*, 192, 107011.
- Ma, Y.-J., Zhai, M.-Y. 2019. A dual-step integrated machine learning model for 24h-ahead wind energy generation prediction based on actual measurement data and environmental factors. *Appl Sci.*, 9: 21-25.
- Madsen, H., Pinson, P., Kariniotakis, G., Nielsen, H. A., Nielsen, T. S. 2005. Standardizing the performance evaluation of short-term wind power prediction models. *Wind engineering*, 29(6): 475-489.
- Mendes-Moreira, J.A., Soares, C., Jorge, A.M., Sousa, J.F.D., 2012. Ensemble approaches for regression: A survey. *ACM Comput. Surv.*, 45(1): 1-10.
- MGM, 2002. Türkiye Rüzgar Atlası, Meteoroloji Genel Müdürlüğü, <https://www.mgm.gov.tr/genel/ruzgar-atlasi.aspx>. Erişim tarihi: 03.02.2021.
- Nguyen, D. H. D., Tran, L. P., Nguyen, V. 2019. Predicting stock prices using dynamic LSTM models. In *International Conference on Applied Informatics*, 199-212.
- NOAA, 2018. Climate Data Online. <https://www.ncdc.noaa.gov/cdo-web/>, Erişim tarihi: 15.10.2020.
- Olah, C. 2015. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>, Erişim Tarihi: 16.02.2021.

- Our World in Data, 2021. Electricity production by source, World. <https://ourworldindata.org/grapher/electricity-production-by-source>, Erişim tarihi: 20.02.2021.
- Ozturk, M., Yuksel, Y. E. 2016. Energy structure of Turkey for sustainable development. *Renewable and Sustainable Energy Reviews*, 53: 1259-1272.
- Öz, S., Alyürük, M. 2020. Energy Sector Overview and Future Prediction for Turkey. *Journal of Industrial Policy and Technology Management*. 3(1): 59-69.
- Petropoulos, A., Chatzis, S.P., Siakoulis, V., Vlachogiannakis, N., 2017. A stacked generalization system for automated forex portfolio trading. *Expert Syst. Appl.*, 90: 290–302.
- Qi, M., Luo, H., Wei, P., Fu, Z. 2019. Estimation of low calorific value of blended coals based on support vector regression and sensitivity analysis in coal-fired power plants. *Fuel*, 236: 1400-1407.
- Qu, X.Y., Kang, X.N., Zhang, C., Jing, S., Ma, X.D. 2016. Short-Term Prediction of Wind Power Based on Deep Long Short-Term Memory. In *Proceedings of the 2016 IEEE PES Asia-Pacific Power and Energy Conference*, Xi'an, China.
- 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.
- Sanjar, K., Bekhzod, O., Kim, J., Paul, A., Kim, J. 2020. Missing Data Imputation for Geolocation-based Price Prediction Using KNN–MCF Method. *ISPRS International Journal of Geo-Information*, 9(4): 227.
- Serbes, G., Sakar, B.E, Gulcur, H.O., Aydin, N. 2015. An emboli detection system based on dual tree complex wavelet transform and ensemble learning. *Appl. Soft Comput.*, 37: 87–94.
- Shamaei, E., Kaedi, M., 2016. Suspended sediment concentration estimation by stacking the genetic programming and neuro-fuzzy predictions. *Appl. Soft Comput.*, 45: 187–196.
- Soman, S. S., Zareipour, H., Malik, O., Mandal, P. 2010. A review of wind power and wind speed forecasting methods with different time horizons. In *North American Power Symposium*, Arlington, TX.
- Son, N., Yang, S., Na, J. 2019. Hybrid Forecasting Model for Short-Term Wind Power Prediction Using Modified Long Short-Term Memory. *Energies*, 12: 3901.
- 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.

- Şenel, M. C., Koç, E. 2015. Dünyada ve Türkiye’de rüzgar enerjisi durumu-Genel değerlendirme. *Mühendis ve Makina*, 56(663): 46-56.
- Tahir, M., El-Shatshat, R., Salama, M.M.A. 2018. Improved stacked ensemble based model for very short-term wind power forecasting. In *Proceedings – 53rd International universities power engineering conference (UPEC)*, Glasgow, 1-6.
- Tiryaki, B. 2008. Predicting intact rock strength for mechanical excavation using multivariate statistics, artificial neural networks, and regression trees. *Engineering Geology*, 99(1-2): 51-60.
- TSKB, 2020. Enerji görünümü raporu. <http://www.tskb.com.tr/i/assets/document/pdf/enerji-sektor-gorunumu-2020.pdf>, Erişim tarihi: 02.02.2021.
- TÜREB, 2021a. Türkiye Rüzgar Enerji İstatistik Raporu.
- TÜREB, 2021b. Türkiye Rüzgar Enerjisi Birliği RES Veritabanı, <https://www.tureb.com.tr/>, Erişim tarihi: 31.12.2020.
- Smil, V. (2017). *Energy transitions: global and national perspectives*. ABC-CLIO.
- Van der Laan, M. J., Polley, E. C., Hubbard, A. E. 2007. Super learner. *Statistical applications in genetics and molecular biology*, 6(1).
- Vapnik, V. 2013. *The nature of statistical learning theory*. Springer science & business media.
- Vega, F. A., Matías, J. M., Andrade, M. L., Reigosa, M. J., Covelo, E. F. 2009. Classification and regression trees (CARTs) for modelling the sorption and retention of heavy metals by soil. *Journal of Hazardous Materials*, 167(1-3): 615-624.
- Wan, Z.Y., Sapsis, T.P. 2017. Reduced-space Gaussian Process Regression for data-driven probabilistic forecast of chaotic dynamical systems. *PhysicaD*, 345: 40–55.
- Wang, C., Li, S., Zhu, M. 2012. Bayesian network learning algorithm based on unconstrained optimization and ant colony optimization. *Syst Eng Electron*, 5: 784–90.
- Wang, X., Guo, P., Huang, X. 2011. A review of wind power forecasting models. *Energy Procedia*, 12: 770–778.
- Williams, C. K., Rasmussen, C. E. 2006. *Gaussian processes for machine learning*. Cambridge, MA: MIT press.
- Wolpert, D. H. 1992. Stacked generalization. *Neural networks*, 5(2): 241-259.

- Wu, Y.K., Hon, J.S. 2007. A Literature Review of Wind Forecasting Technology in the World. Proceedings of the IEEE Conference on Power Tech, Lausanne, 504-509.
- Xiaoyun, Q., Xiaoning, K., Chao, Z., Shuai, J., Xiuda, M. 2016. Short-term prediction of wind power based on deep long short-term memory. In 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), 1148-1152.
- Yang, W., Wang, K., Zuo, W. 2012. Neighborhood component feature selection for high-dimensional data. *J. Comput.*, 7(1): 161-168.
- Yaniktepe B., Savrun M., Koroglu T. 2013. Current status of wind energy and wind energy policy in Turkey, *Energy Conversion and Management*, 72: 103-110.
- YEKDEM, 2020. Yenilenebilir Enerji Kaynakları Destekleme Mekanizması [https://www.enerjiportali.com/wpcontent/uploads/2020/09/KSD\\_YEKDEM\\_2.0\\_9.2020-1-1.pdf](https://www.enerjiportali.com/wpcontent/uploads/2020/09/KSD_YEKDEM_2.0_9.2020-1-1.pdf), Erişim tarihi: 08.01.2021.
- Yu, P. S., Chen, S. T., Chang, I. F. 2006. Support vector regression for real-time flood stage forecasting. *Journal of hydrology*, 328(3-4): 704-716.
- Yu, R., Gao, J., Yu, M., Lu, W., Xu, T., Zhao, M., Zhang, Z. 2019. LSTM-EFG for wind power forecasting based on sequential correlation features. *Future Generation Computer Systems*, 93: 33-42.
- Zhang, P., Wang, Y., Liang, L., Li, X., Duan, Q. 2020. Short-Term Wind Power Prediction Using GA-BP Neural Network Based on DBSCAN Algorithm Outlier Identification. *Processes*, 8(2): 157.
- Zhong, H., Wang, J., Jia, H., Mu, Y., Lv, S. 2019. Vector field-based support vector regression for building energy consumption prediction. *Applied Energy*, 242, 403-414.

## ÖZGEÇMİŞ

**Adı Soyadı** : Kübra YAZICI

### ÖĞRENİM DURUMU

Derece	Eğitim Birimi	Mezuniyet Yılı
Yüksek Lisans	Sakarya Üniversitesi / Fen Bilimleri Enstitüsü / Endüstri Mühendisliği	Devam ediyor
Lisans	Sakarya Üniversitesi / Mühendislik Fakültesi / Endüstri Mühendisliği	2018
Lise	Gebze Anadolu Öğretmen Lisesi	2014

### İŞ DENEYİMİ

Yıl	Yer	Görev
2019-Halen	Türk-Alman Üniversitesi	Araştırma Görevlisi

### YABANCI DİL

İngilizce, Almanca

### ESERLER (makale, bildiri, proje vb.)

1. Yazıcı, K., Gökler, S.H., Boran, S. 2020. An integrated SMED-fuzzy FMEA model for reducing setup time. J Intell Manuf, Journal of Intelligent Manufacturing, 32(6): 1547-1561.
2. Yazıcı, K., Boran,S., Gökler, S.H. 2019. Bir Lojistik Firmasında 6 Sigma Uygulaması, Alphanumeric Journal, 7 (): 87-98.
3. Aylak, B. L., Oral, O., Yazıcı, K. 2021. Yapay Zeka ve Makine Öğrenmesi Tekniklerinin Lojistik Sektöründe Kullanımı. El-Cezeri Journal of Science and Engineering, 8(1): 74-93.



4. Yazıcı, K., Boran, S., Short-Term Wind Power Prediction Approach Based On Bayesian Optimization and Ensemble Learning, Journal of Intelligent Systems: Theory and Applications 4(2) 2021: 142-155. DOI: 10.38016/jista.889991
5. Yazıcı, K., Gökler, S.H., Boran, S. 2019. A Novel Smed Model Using Fuzzy Fmea Method, IMSS'19 Sakarya University - Sakarya/Turkey, 9-11 September, 530-539.
6. Boran, S., Demircioğlu, D., Yazıcı, K. 2019. Pattern Recognition in Process Control Charts With Ensemble Machine Learning Techniques, IMSS'19 Sakarya University - Sakarya/Turkey, 9-11 September, 937-945.
7. Yazıcı, K., Dörtköşe, S., Say, B., Kökçam, A.H., Turna, F. 2019. Classification of Heart Attack Risk Using Fuzzy Inference System, IMSS'19 Sakarya University - Sakarya/Turkey, 9-11 September, 540-548.
8. Yazıcı, K., Boran, S., Gökler, S.H. 2018. Bir Lojistik Firmasında 6 Sigma Uygulaması, 2.Ulaştırma ve Lojistik Ulusal Kongresi, Sakarya.