

AKÜ FEMÜBİD 22 (2022) 055103 (1028-1040)

AKU J. Sci. Eng. 22 (2022) 055103 (1028-1040)

DOI: 10.35414/akufemubid.1117779

Araştırma Makalesi / Research Article

Rüzgar Gücü Tahmininde Genetik Algoritma ile Öznitelik Seçimi

Ece YAĞMUR¹, Sercan YAĞMUR²¹ Konya Teknik Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Endüstri Mühendisliği Bölümü, Konya.² Konya Teknik Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Makine Mühendisliği Bölümü, Konya.e-posta: *Corresponding author: ecyagmur@ktun.edu.tr. ORCID ID: <http://orcid.org/0000-0001-5865-3483>esyagmur@ktun.edu.tr. ORCID ID: <http://orcid.org/0000-0002-5478-5451>

Geliş Tarihi: 17.05.2022

Kabul Tarihi: 05.10.2022

Öz

Sürdürülebilir gelişim için yenilenebilir enerji kaynaklarına olan ihtiyaç her geçen gün artmaktadır. Bu kaynaklardan birisi de rüzgar enerjisidir. Rüzgarın stokastik yapısı nedeniyle rüzgar hızı ve rüzgar gücünün tahmini son yıllarda araştırmacılar tarafından oldukça ilgi çeken bir konu haline gelmiştir. Yapılan çalışmada Türkiye’de yer alan bir rüzgar türbini için 2018 yılı boyunca SCADA sistemi ile elde edilen veri seti ile aynı konum için NASA tarafından paylaşılan meteorolojik veri seti kullanılarak rüzgar gücü tahmini gerçekleştirilmiştir. Girdi değişkenleri olarak SCADA sisteminden çekilen rüzgar hızı, rüzgar yönü ve teorik güç eğrisi; NASA sisteminden çekilen meteorolojik parametreler ve rüzgar gücüne ait geçmiş veriler kullanılmıştır. Modelde yer alan ve hesaplama karmaşıklığına neden olan gereksiz öznitelikler model performansını artırmak amacıyla sarmal seçim yöntemi ile modelden çıkarılmıştır. Sarmal seçim yöntemi olarak Genetik Algoritma (GA) kullanılmıştır. Yapılan çalışmada hem farklı makine öğrenme algoritmalarının tahmin gücü, farklı performans ölçütlerine göre karşılaştırılmış hem de öznitelik seçiminin modele etkisi değerlendirilmiştir. GA ile önerilen nihai modelde değişken sayısı 47’den 9’a indirgenerek gereksiz değişkenler modelden uzaklaştırılmış ve en az sayıda değişken ile R² değeri 0,98 olan güçlü bir tahmin modeli elde edilmiştir.

Anahtar kelimeler

Makine öğrenmesi;
Rüzgar gücü;
Yenilenebilir enerji;
Öznitelik seçimi;
Genetik algoritma

Feature Selection by Genetic Algorithm for Wind Power Prediction

Abstract

The need for renewable energy sources for sustainable development has been increasing every day. One of these sources is wind energy. Due to the stochastic nature of the wind, the estimation of wind speed and wind power has been a subject of great interest to researchers in recent years. In this study, wind power estimation was carried out for a wind turbine in Turkey, using the data set obtained by the SCADA system during 2018 and the meteorological data set shared by NASA for the same location. Wind speed, wind direction, and theoretical power curve were taken from the SCADA system as input variables; Meteorological parameters were taken from the NASA system and historical data of wind power were used. Unnecessary features in the model that cause computational complexity are removed from the model with the wrapper selection method to increase model performance. Genetic Algorithm (GA) was used as the wrapper selection method. In the study, the predictive power of different machine learning algorithms was compared according to different performance criteria and the effect of feature selection on the model was evaluated. In the final model proposed by GA, the number of variables was reduced from 47 to 9, unnecessary variables were removed from the model, and a strong prediction model with R² value of 0.98 was obtained with the least number of variables.

Keywords

Machine learning;
Wind power;
Renewable energy;
Feature selection;
Genetic algorithm

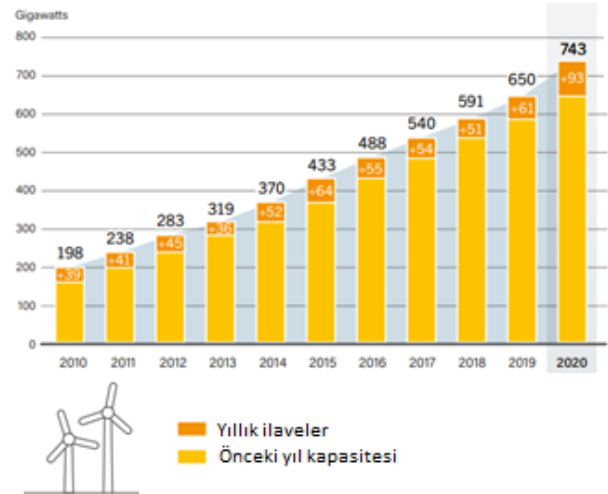
1. Giriş

Sürdürülebilirlik, gelecek nesillerin kendi ihtiyaçlarını karşılama yeteneğinden ödün vermeden bugünün ihtiyaçlarını karşılayarak kalkınma süreci olarak tanımlanmaktadır (WCED 1987). Özellikle doğal kaynakların azaldığı, çevresel problemlerin küresel boyuta ulaştığı günümüzde, sürdürülebilir gelişim, sadece bireysel sorumluluklarla değil, kolektif farkındalıklarla ulaşılabilen ve devamlılığı hem bizler hem de gelecek nesiller için oldukça önemli olan bir süreç haline gelmiştir. Ülkelerin sürdürülebilirlik hedefleri doğrultusunda tüm dünyaya karşı üstlenmiş oldukları sorumlulukları düzenlemek ve takip etmek adına Montreal Protokolü (1987), Kyoto Protokolü (1997), Paris İklim Anlaşması (2015) gibi pek çok yasal düzenleme ve kurallar ortaya koyulmuştur. Bunun yanı sıra Birleşmiş Milletler (BM) tarafından, “Sürdürülebilir Kalkınma Amaçları” adı altında 2030 yılına kadar ulaşılması planlanan on yedi küresel amacın ortaya konulduğu bir eylem planı gerçekleştirilmiştir (Int Kyn. 1). Sürdürülebilirliğin üç temel boyutu olan ekonomik, sosyal ve çevresel sürdürülebilirliğin ele alındığı bu plan dahilinde her yıl ülke bazında sürdürülebilir kalkınma raporları paylaşmakta ve ulaşılan hedefler değerlendirilmektedir. Sürdürülebilir Kalkınma Amaçları’nda tanımlanan “Erişilebilir ve Temiz Enerji” (Amaç 7) için Şekil 1’de görüldüğü gibi farklı hedefler geliştirilmiştir. Buna göre yenilenebilir enerjinin küresel payının artırılması hedefi (Hedef 7.2) doğrultusunda güneş, rüzgar, jeotermal gibi doğal kaynaklardan elde edilen yenilenebilir enerji kaynaklarına olan yatırım faaliyetlerinin artırılması planlanmaktadır.

Sürdürülebilir enerji kaynaklarından biri olan rüzgar enerjisi için yapılan çalışmalar incelendiğinde 2010 - 2020 yılları arasında küresel rüzgar gücü kapasitesindeki artış Şekil 2’deki gibi raporlanmıştır. Aynı raporda 2020 yılı için Danimarka ‘da rüzgar enerjisinin elektrik üretimindeki payının %58’in üzerinde olduğu ve kurulu güç ilavesi bakımından da Türkiye’nin dünya sıralamasında ilk onda yer aldığı belirtilmiştir.



Şekil 1. Erişilebilir ve Temiz Enerji Hedefleri (BM) (Int Kyn. 2)



Şekil 2. Rüzgar enerjisi için küresel kapasite (2010-2020) (Int Kyn. 3)

Rüzgar enerjisine olan talebin artmasıyla birlikte literatürde farklı çalışma alanları ortaya çıkmıştır. Etkin enerji üretimi açısından, rüzgarın stokastik yapısı göz önünde bulundurulduğunda rüzgar gücü ve rüzgar hızının doğru tahmini önem arz etmektedir. Makine öğrenmesi ile önerilen tahmin modelleri, kendi kendine organizasyon mekanizması ve uyarlanabilir öğrenmenin avantajları nedeniyle

genellikle geleneksel istatistiksel yöntemlerden daha iyi performans göstermektedir (Liu and Chen 2019).

Literatürde rüzgar hızı ve gücü tahmininde makine öğrenme yöntemlerinin kullanıldığı pek çok çalışma bulunmaktadır. Heinermann ve Kramer (2016), rüzgar gücü tahmini için karar ağaçları ve destek vektör regresyonunu kullanarak entegre bir model önermişlerdir. Demolli *et al.* (2019), Türkiye’de dört farklı bölgede yer alan rüzgar türbinleri için geçmiş dönemli rüzgar hızı verileri ile uzun dönemli rüzgar gücü tahmininde beş farklı makine öğrenme algoritmasının performansını karşılaştırmışlardır. Deng *et al.* (2021)’nin yapmış olduğu kaynak araştırmasında, 2012-2019 yılları arasında rüzgar gücü ile ilişkili olan ve Web of Science (WoS) Core Collection’da yer alan çalışmalar makine öğrenme teknikleriyle incelenerek oldukça fazla sayıdaki makale analiz edilmiş, araştırmacıların eğilimleri ve bölgesel dağılımları gösterilmiştir. Foley *et al.* (2012), rüzgar gücü için literatürde yer alan mevcut tahmin yöntemlerini araştırarak bazı rüzgar enerjisi yazılım modellerini listelemişler ve bu sektördeki yenilikleri analiz etmişlerdir.

Liu ve Chen (2019), yapmış oldukları çalışmada, rüzgar gücü tahmini ile ilgili yapılan çalışmaların büyük çoğunluğunun sadece geçmiş rüzgar hızı veya güç serilerini kullanmakta olduğunu ve meteorolojik koşullar gibi dış değişkenlerin tahmin modellerine eklenmesiyle modelin daha da güçlenebileceğini belirtmişlerdir. Ancak bireysel olarak iyi performans gösteren özelliklerin kombinasyonlarının modele nasıl etki edeceği belirsizdir. Bu nedenle bu dezavantajı gidermek ve gereksiz bilginin neden olduğu tahmin hatası ve hesaplama karmaşıklığını azaltmak amacıyla boyut indirgeme yöntemleri kullanılmaktadır. Makine öğrenmesinde boyut indirgeme amacıyla kullanılan iki yaklaşım bulunmaktadır. Bunlar öznitelik seçimi ve öznitelik çıkarımıdır. Öznitelik seçiminde belirli kriterlere göre mevcut özelliklerin bir alt kümesi seçilirken, öznitelik çıkarımında mevcut özelliklerden yeni öznitelikler üretilebilir. Literatürde rüzgar hızı ve gücü tahmini için öznitelik çıkarımına başvurulmuş pek çok çalışma bulunmaktadır (Higashiyama *et al.* 2018, Shi *et al.* 2019, Lu *et al.* 2021). Bu çalışmada ise boyut indirgeme yöntemi olarak öznitelik seçimine

başvurulmuştur. Literatürde öznitelik seçimi için farklı metasezgisel algoritmaların kullanıldığı görülmüştür. Salcedo-Sanz *et al.* (2014), kısa dönemli rüzgar hızı tahmini için öznitelik seçiminde Mercan Resifleri Optimizasyon (Coral Reef Optimization) algoritmasını kullanmışlar ve başlangıçta yer alan 6534 özelliği sadece 9 değişkene indirgeyerek hesaplama karmaşasını oldukça azaltmış ve tahmin gücü yüksek bir model geliştirmişlerdir. Azimi *et al.* (2016) modelin doğruluğunu artırmak amacıyla düzensiz kalıptaki yapıları gidermek için K-Ortalama Kümeleme (K-Means Clustering) yöntemini kullanmışlardır. Wang *et al.* (2017), kısa dönemli rüzgar hızı tahmininde optimal girdi değişkenlerini belirlemek amacıyla Rassal Orman (Random Forests) algoritmasını kullanmışlardır. Çalışma sonuçlarına göre algoritma ilişkisiz değişkenleri kaldırarak aday girdi değişkenlerden en güçlü şekilde ilişkili olan değişkenleri belirlemiştir. Sun *et al.* (2017), rüzgar hızı tahmininde girdi değişkenlerin tahmin modelinde dinamik olarak seçildiği ve parametre optimizasyonunun yapıldığı iki aşamalı bir yöntem önermişlerdir. Wang *et al.* (2018), K-Ortalama Kümeleme yaklaşımı ile girdi verilerinin belirsizliğini hesaba katarak, tahmin doğruluğu üzerinde en büyük etkiye sahip olan ve önerilen modelin verimliliğini artıran en büyük geçmiş örnekleri tespit etmişlerdir. Liu *et al.* (2020), ilişkisiz ve gereksiz faktörleri modele dahil etmenin, önemli faktörlerin rolünü maskeleyebileceğini ifade ederek model oluşturulmadan önce karşılıklı bilgi (mutual information) yaklaşımı ile öznitelik seçimini gerçekleştirmiştir. Görüldüğü gibi öznitelik seçim süreci modelin performansını artıran bir süreç olup literatürdeki pek çok çalışmada konu edinmiştir. Yapılan çalışmada Türkiye’de yer alan bir rüzgar türbini için 2018 yılı boyunca SCADA sistemi ile elde edilen veri seti ile aynı konum için NASA tarafından paylaşılan meteorolojik veri seti kullanılarak rüzgar gücü tahmini gerçekleştirilmiştir. Modelde yer alan ancak model doğruluğunun azalmasına ve hesaplama karmaşıklığına neden olan özellikleri elimine etmek amacıyla öznitelik seçim sürecine başvurulmuştur. Öznitelik seçim sürecinde Genetik Algoritma (GA) kullanılmıştır.

Rüzgar gücü tahmininde Çoklu Doğrusal Regresyon, Ridge Regresyon, Lasso Regresyon, Elastik Net, Karar Ağaçları ve Rassel Orman Algoritması kullanılmıştır. Modeller geliştirilirken verilerin %75'i modeli eğitmek, %25'i ise geliştirilen modeli test etmek amacıyla kullanılmıştır. Bu oran literatürde de sıklıkla kullanılmaktadır (Olson and Moore 2016, Theofilatos *et al.* 2019, Le *et al.* 2020). Geliştirilen modeller determinasyon katsayısı (R^2), ortalama mutlak hata (Mean Absolute Error, MAE) ve kök ortalama hata karesi (Root Mean Squared Error, RMSE) performans ölçütlerine göre karşılaştırılmıştır. Ayrıca öznitelik seçim sürecinin modelin performansına etkisi de araştırılmıştır.

Çalışmanın bundan sonraki kısmı şu şekildedir: Materyal ve Metot bölümünde Makine Öğrenme Algoritmalarının genel işleyişi sırasıyla verilerin toplanması, veri ön işleme, modelin seçilmesi ve öznitelik seçimi alt başlıklarıyla incelenmiştir. Bulgular bölümünde kullanılan performans ölçütleri açıklanarak geliştirilen modellere ilişkin karşılaştırmalı sonuçlar verilmiştir. Tartışma ve Sonuç bölümünde ise çalışmayla ilgili genel bir değerlendirme yapılarak gelecekte yapılması planlanan çalışmalara değinilmiştir.

2. Materyal ve Metot

Bu bölümde rüzgar gücü tahmini için kullanılan makine öğrenme algoritmalarının uygulama adımları olan verilerin toplanması, veri ön işleme, model seçimi ve öznitelik seçimi başlıkları sırasıyla açıklanmıştır.

2.1 Verilerin toplanması

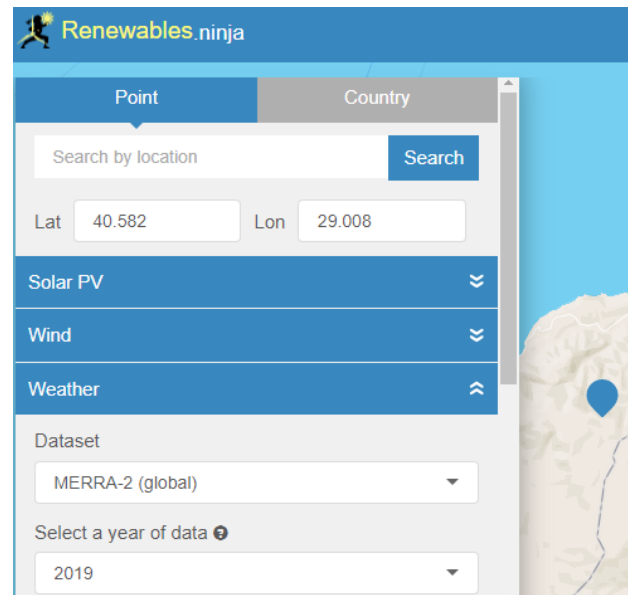
Yapılan çalışmada Türkiye'de elektrik üreten bir rüzgar türbininin SCADA sistemi aracılığıyla elde edilen CSV formatındaki rüzgar verileri kullanılmıştır (Int Kyn. 4). SCADA sistemi ile rüzgar hızı (m/sn), rüzgar yönü ($^{\circ}$), aktif güç (kW) ve teorik güç (kW-s) bilgileri bir yıllık dönem boyunca (01.01.2018-31.12.2018) on dakikalık periyotlarla sistemden çekilmiş ve raporlanmıştır. Rüzgar türbinine ait koordinat bilgileri "X:668478 Y:4494833 UTM ED 50" olarak verilmiş ve dönüşüm yapılarak türbinin enlem ve boylam bilgileri sırasıyla 40,58 ve 28,99 olarak elde edilmiştir. Şekil 3'te lokasyon bilgileri verilen NORDEX N117 model türbin, TUREB (Türkiye

Rüzgar Enerjisi Birliği) tarafından paylaşılan RES veri tabanına göre Yalova Esenköy' de bulunmakta olup türbin gücü 3,6 MW olarak raporlanmıştır.



Şekil 3. Rüzgar türbini konum bilgileri (Int Kyn. 5)

Rüzgar gücüne meteorolojik verilerin etkisini değerlendirmek amacıyla NASA tarafından paylaşılan MERRA-2 Global (Modern Era Retrospective analysis for Research and Applications) veri seti kullanılmıştır. Şekil 4'te görüldüğü gibi sisteme rüzgar türbininin bulunduğu lokasyonun enlem ve boylam değerleri girilerek uydu gözlemleri ile mevsimsel parametrelere ilişkin veriler çekilmiştir. Kullanılan veri setinde, bir yıllık dönem boyunca (01.01.2018-31.12.2018) saatlik periyotlarla hava sıcaklığı ($^{\circ}$ C), yağış (mm/s), kar yağışı (mm/s), kar kütlesi (kg/m^2), hava yoğunluğu (kg/m^3), bulutluluk oranı ve radyasyon oranlarına (W/m^2) ait veriler paylaşılmaktadır.



Şekil 4. Meteorolojik verilerin elde edilmesi (Int Kyn. 6)

Çizelge 1. SCADA değişkenlerine ait tanıttıcı istatistikler

	Aktif Güç	Rüzgar Hızı	Teorik Güç Eğrisi	Rüzgar Yönü
Maksimum	3604,41	23,75	3600,00	357,27
Minimum	-0,18	0	0	0
Ortalama	1305,54	7,55	1490,50	123,67
Standart Sapma	1291,27	4,17	1345,04	88,21
Gelen Veri	8440	8440	8440	8440
Veri Boyutu	8760	8760	8760	8760
Veri Yüzdesi	%96,35	%96,35	%96,35	%96,35

Meteorolojik veriler bir saatlik periyotlar halinde orijinaline sadık kalınarak alınırken, rüzgar türbinine ait veriler on dakikalık altı adet verinin ortalamaları alınarak saatlik periyota dönüştürülmüştür. Her iki veri setine ilişkin tanıttıcı istatistiklere ait özet bilgiler Çizelge 1 ve Çizelge 2’de verilmiştir. Her iki veri setinin saatlik bazda birleştirilmesiyle nihai veri seti oluşturulmuştur.

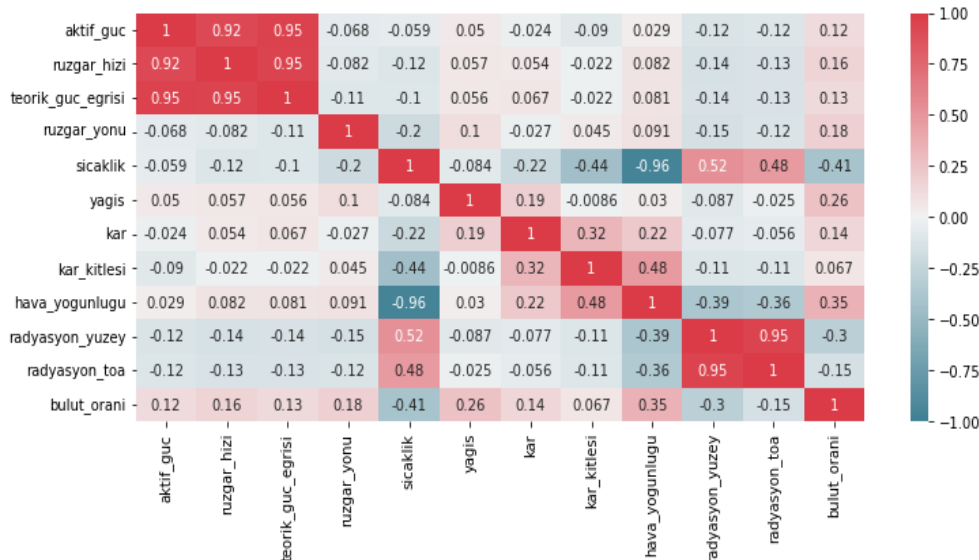
Çizelge 2. NASA değişkenlerine ait tanıttıcı istatistikler

	Sıcaklık	Yağış	Kar	Kar Kitlesi	Hava Yoğ.	Radyasyon Yüzey	Radyasyon Atmosfer	Bulutluluk Oranı
Maksimum	31,95	5,93	0,17	1,25	1,29	1005,19	1251,01	0,99
Minimum	-2,05	0	0	0	1,12	0	0	0
Ortalama	15,74	0,09	0,00	0,04	1,19	189,42	326,02	0,43
Standart Sapma	7,44	0,29	0,01	0,15	0,04	271,94	407,54	0,36
Gelen Veri	8760	8760	8760	8760	8760	8760	8760	8760
Veri Boyutu	8760	8760	8760	8760	8760	8760	8760	8760
Veri Yüzdesi	%100	%100	%100	%100	%100	%100	%100	%100

2.2. Veri önışleme

Çizelge 1’de görüldüğü gibi mevsimsel değişkenlerde veri kaybı olmazken, rüzgar türbininden elde edilen veri setinde %3,65 oranında eksik veri bulunmaktadır. Eksik verilerin doldurulması için uygulama kolaylığı açısından K-En Yakın Komşu (K-Nearest Neighbors, kNN) Algoritması kullanılmıştır. Algoritma seçilen K adet en yakın komşuya ait uzaklıkları kullanarak kayıp gözlem değerlerinin tahmin edilmesi prensibine dayanmaktadır (Zhang 2012). İncelenen problem için eksik verilerin doldurulmasında “sklearn” kütüphanesinde yer alan “KNNImputer” sınıfından yararlanılmıştır. Eksik verilerin doldurulması aşamasından sonra, değişkenler arasındaki ilişkiyi

gösteren korelasyon matrisi “seaborn” kütüphanesi ile oluşturulmuştur. Korelasyon matrisinde ilişkiye ait renk koyulaştıkça ilişkinin gücünün arttığı söylenebilir. Buna göre Şekil 5’te verilen korelasyon matrisi incelendiğinde aktif güç üzerinde en fazla etki teorik güç eğrisi ve rüzgar hızına aittir. Aktif güç üzerinde negatif etkiye sahip olan değişkenler ise rüzgar yönü, sıcaklık, kar ve kar kitlesi ile radyasyon oranlarıdır. Sıcaklık artışı hava yoğunluğunda düşüşe sebep olduğundan rüzgar gücü üzerinde negatif bir etkiye sahiptir. Kar ve kar kitlesi, kanatların havayı verimli bir şekilde yakalama yeteneğini azaltarak yine rüzgar gücünde düşüşe sebep olmaktadır. Solar radyasyon artışı ise sıcaklığı artırdığından, dolaylı olarak rüzgar gücünde azalmaya sebep olmaktadır.

**Şekil 5.** Korelasyon matrisi

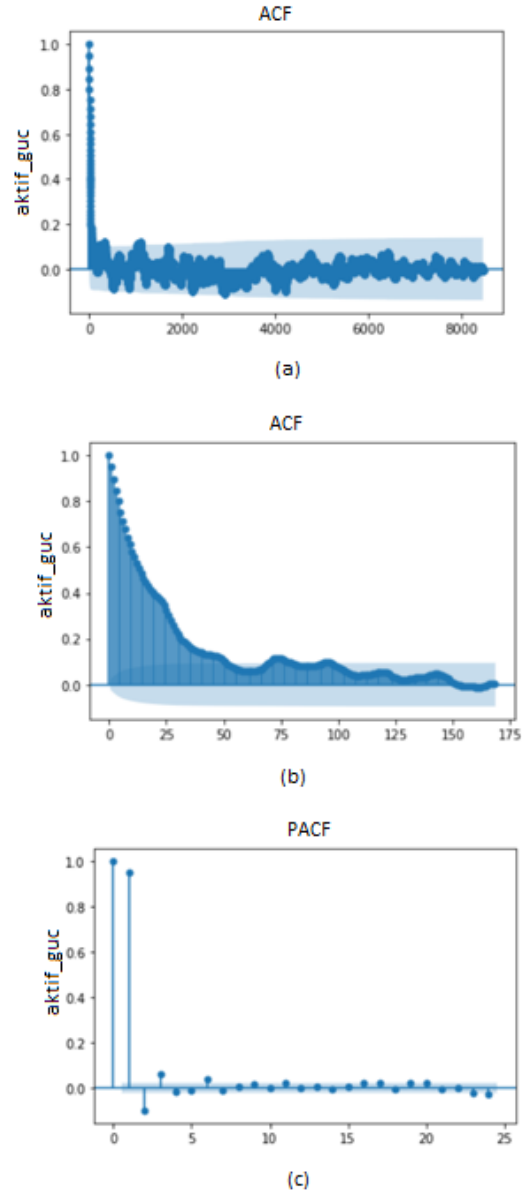
Makine öğrenmesinde kullanılan algoritmalar değişkenlerin ölçeklerine duyarlı olabilir. Bu nedenle veriler "sklearn" kütüphanesinde yer alan "MinMax Scaler" ile normalize edilmiştir. Veriler normalize edildikten sonra, eğitim ve test olmak üzere iki gruba bölünür. Eğitim setinde bulunan veriler modelin eğitilmesinde kullanılırken, test setinde yer alan veriler geliştirilen modelin performansını değerlendirmek için kullanılır. Yapılan çalışmada verilerin %75'i modelin eğitilmesi, %25'i ise modeli test etmek için kullanılmıştır.

2.3. Modelin seçilmesi

Bir rüzgar enerjisi tahmin sürecinde hedef değişken olan rüzgar gücü; sıcaklık, rüzgar hızı, rüzgar yönü, gibi meteorolojik değişkenlerin bir fonksiyonudur. Ancak herhangi bir t anında üretilen güç, sadece t anındaki meteorolojik değişkenlere değil, aynı zamanda bu değişkenlerin ve hatta hedef değişken olan rüzgar gücünün geçmiş değerlerine de bağlıdır (Renani *et al.* 2016). t anından önceki verileri tutan bu değişkenler "gecikme değişkeni (lag variable)" olarak adlandırılmaktadır. Bu ifadeye bağlı olarak gecikme değişkenleri açıklayıcı değişken olarak regresyon modeline dahil edilebilir. Bu amaçla öncelikle veri setinde $t-1$ anındaki gözlemlerin t anındaki gözlemleri ne kadar etkilediğini araştırmak amacıyla oto korelasyon fonksiyonu (ACF) kullanılmıştır. Kısmi oto korelasyon fonksiyonu (PACF) ise diğer gecikmeli serilerin etkileri ihmal edildiğinde t ile $t-k$ serileri arasındaki ilişki miktarını göstermektedir.

Şekil 6.c'de görüldüğü gibi 4 gecikmeden sonraki kısmi oto korelasyon katsayıları güven aralığı içerisinde kalmaktadır. Buna göre tahmin modeli; Y_t , hedef değişken olan aktif rüzgar gücünü, X_i , i . bağımsız değişkeni ve n ise girdi değişkeni sayısını ifade etmek suretiyle, rüzgar gücüne ait gecikme değişkeni $k = 4$ olacak şekilde Eşitlik (1)'deki gibi oluşturulmuştur.

$$Y(t) = f \left(\begin{array}{c} Y_{(t-1)}, \dots, Y_{(t-k)}, X_{1(t)}, X_{1(t-1)}, \dots, X_{1(t-k)}, \\ X_{2(t)}, X_{2(t-1)}, \dots, X_{2(t-k)}, \dots \\ X_{n(t)}, X_{n(t-1)}, \dots, X_{n(t-k)} \end{array} \right) \quad (1)$$



Şekil 6. a) ACF-1 yıl. b) ACF-1 hafta. c) PACF-1 gün.

Makine öğrenmesinde kullanılan algoritmaların performansı her problem için aynı olmayıp verinin boyutu, özellikleri ve problemin yapısına göre farklılıklar göstermektedir. Yapılan çalışmada performanslarını karşılaştırmak amacıyla Çoklu Doğrusal Regresyon, Lasso Regresyon, Ridge Regresyon, Elastik Net, Karar Ağaçları ve Rassal Orman Algoritması kullanılmıştır.

Çoklu Doğrusal Regresyon: Yapılan çalışmada Çoklu Doğrusal Regresyon için "En Küçük Kareler" yöntemi kullanılmıştır. Yöntem Eşitlik (2)'de görüldüğü gibi Artık Kareler Toplamı (Residual Sum of Squares)'ni en küçükleyecek şekilde çalışmaktadır.

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^m \beta_j X_{ij} \right)^2 \quad (2)$$

Ridge Regresyon: Hoerl ve Kennard (1970) tarafından En Küçük Kareler yöntemine alternatif olarak önerilmiştir. Çoklu Doğrusal Regresyon’da iki ya da daha fazla bağımsız değişken arasında doğrusal veya doğrusala çok yakın bir ilişki olması durumunda ortaya çıkan yanlılığı gidererek daha küçük varyanslı tahminler üretir. Eşitlik (3)’te görüldüğü gibi katsayılara ceza katsayısı uygulayarak; ceza ve katsayı karelerinin toplamının çarpımı da minimize edilmeye çalışılır. Eğer λ değeri 0 ise Ridge Regresyon; Çoklu Doğrusal Regresyon haline gelir, λ ’nın çok büyük olduğu durumda ise tüm katsayılar sıfıra yaklaşır ve model düz bir çizgi haline dönüşür (Melkumova and Shatskikh 2017). Sonuç olarak λ arttıkça varyans yani model tahmininin değişkenliği azalacak ancak yanlılık artacaktır. Yüksek yanlılığa sahip olan bir modelin eksik öğrenme (underfitting) olasılığı oldukça fazla iken yüksek varyansa sahip olan bir model ise aşırı öğrenme (overfitting) riski ile karşı karşıyadır (Belkin *et al.* 2019). Bu nedenle yanlılık ve varyans arasında bir ödünleşim sağlanarak λ değeri test koşullarında sonucunda varyans, yanlılığın karesi ve MSE (Mean Squared Error) performans kriterleri dikkate alınarak 1 olarak belirlenmiştir.

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^m \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^m \beta_j^2 \quad (3)$$

Lasso (Least Absolute Shrinkage and Selection Operator) Regresyon: Tibshirani (1996) tarafından geliştirilen yöntem Ridge Regresyon’da olduğu gibi çoklu doğrusal bağlantı problemini çözmek için önerilmiştir. Ridge Regresyon’dan farkı Eşitlik (4)’te görüldüğü katsayıların karelerini almak yerine mutlak değerini almasıdır. Mutlak değer alındığı için bazı öznitelikler tamamıyla ihmal edilir. Böylece Lasso Regresyon sadece aşırı öğrenmeyi azaltmak için değil aynı zamanda öznitelik seçiminde de önemli bir rol oynamaktadır. Yapılan çalışmada λ değeri 1 olarak alınmıştır.

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^m \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^m |\beta_j| \quad (4)$$

Elastik Net: Ridge Regresyon gibi cezalandırma, Lasso Regresyon gibi değişken seçimi yapma üzere Zou ve Hastie (2005) tarafından önerilmiştir. Eşitlik (5)’te görüldüğü gibi λ “0” olduğunda denklem Ridge; “1” olduğunda Lasso Regresyona dönüşür. Yapılan çalışmada λ değeri 0,5 olarak alınmıştır.

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^m \beta_j X_{ij} \right)^2 + (1 - \lambda) \sum_{j=1}^m \beta_j^2 + \lambda \sum_{j=1}^m |\beta_j| \quad (5)$$

Karar Ağaçları: Hem regresyon hem de sınıflama modelleri için uygulanabilir. Kök düğüm bağımlı değişkeni temsil ederken kökten yaprak düğümlere doğru gidildikçe veri seti küçük alt parçalara ayrılarak dallanır. Dallarda seçilen karara göre sonraki düğümlere geçiş yapılır. Model, tekrarlayan dallanma süreci boyunca tahmin için kullanılabilir bir dizi kural üretir (Tso and Yau 2007) ve nihai düğüme ulaşana kadar süreç tekrarlanır.

Rassal Orman Algoritması: Breiman (2001) tarafından karar ağaçlarının en büyük problemlerinden biri olan aşırı öğrenme problemini gidermek amacıyla önerilmiştir. Rassal olarak seçilen alt kümeler üzerinde birbiri arasında korelasyon olmayan karar ağaçları birlikte çalışmakta ve her bir ağaç bireysel tahminlerde bulunmaktadır. Regresyon modellerinde nihai sonucun elde edilmesinde bireysel tahminlerin ortalaması alınmaktadır. Rassal Orman algoritmasında kullanılan iki temel parametre bulunmaktadır. Bunlar karar ağacı sayısı ve her bir düğümde kullanılacak olan değişken sayılarıdır.

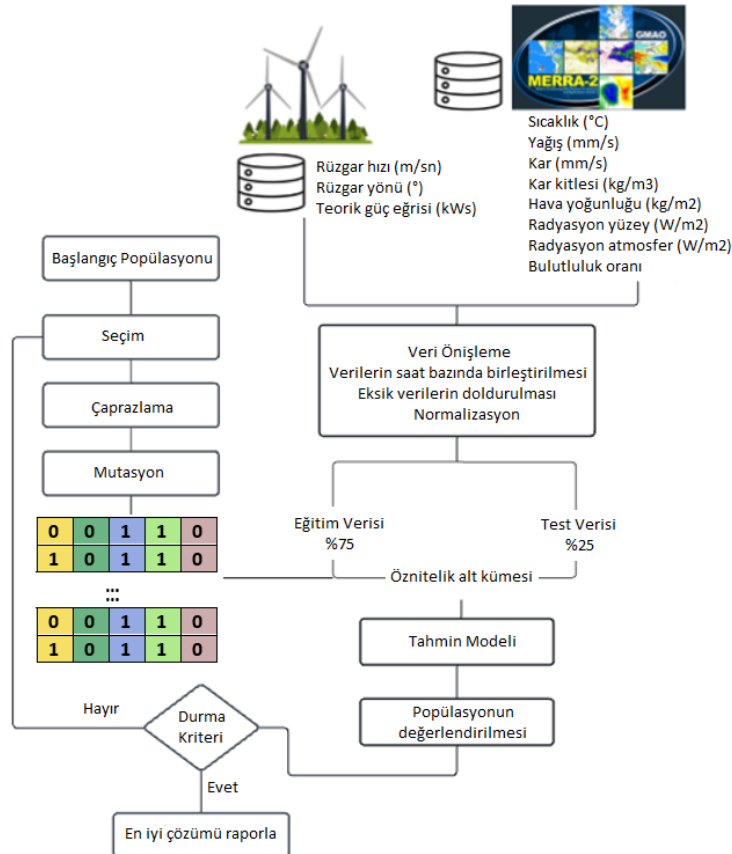
2.4. Öznitelik seçimi

Önceki bölümde bahsedildiği gibi 11 adet bağımsız değişken bulunan modelde gecikme değişkenleri de hesaba katılarak toplam 47 adet girdi değişkeni bulunmaktadır. Ancak bu şekilde büyük miktarda öznitelik kullanımı ile bilgi vermeyen değişkenlerin varlığı, tahminlerde belirsizlik oluşturabilir (Kuhn and Johnson 2013). Bazı özellikler bağımlı değişken için alakasız veya daha az önemli olabileceğinden, modele gereksiz şekilde dahil edilmeleri, model karmaşıklığını artırarak modelin yorumlanmasını

zorlaştırabilir. Gereksiz öznitelik kullanımının bir diğer dezavantajı ise modelin eğitilmesi için gereken süreyi artırarak modelin etkinliğini sınırlandırmasıdır. Bu nedenle yapılan çalışmada makine öğrenme algoritmalarından etkin düzeyde yararlanılması amacıyla öznitelik seçim prosedürü uygulanmıştır. Öznitelik seçimi için temelde iki yöntem kullanılmaktadır. Bunlar filtreleme yöntemleri ve sarmal (wrapper) yöntemlerdir. Filtreleme yöntemlerinde istatistiksel analizlere başvurularak anlamlılık düzeylerine göre öznitelikler belirlenir. Ancak etkisi anlamsız olan bir girdi değişkeninin modelden çıkarılması modelin performansını her zaman artırmaz. Modelden çıkarılan bu değişkenin başka bir girdi değişkeni ile ilişkisi olabilir. Bu dezavantajı gidermek için sarmal yöntemler önerilmiştir. Sarmal yöntemlerde, iteratif bir arama süreci gerçekleştirilerek farklı alt kümeler ile denemeler yapılır ve en uygun öznitelik kümesi oluşturulur. İncelenen modelde n adet öznitelik bulunması durumunda 2^n farklı olası alt küme oluşacaktır. En iyi kombinasyonun belirlenmesi problemi, NP-Zor yapıda olup çözümünde kısa sürede etkin çözümler üreten sezgisel algoritmaların kullanılması kaçınılmazdır. Bu nedenle sarmal

yöntemler, öznitelik seçiminde, genellikle sezgisel algoritmalara başvurmaktadır. Yapılan çalışmada sarmal yöntem olarak Genetik Algoritma (GA) yaklaşımına başvurulmuştur. GA, popülasyon tabanlı bir algoritma olup, Darwin'in doğal seleksiyon teorisine bağlı olarak geliştirilen bir stokastik arama yöntemidir (Holland 1992). GA'nın çalışma prensibinin daha iyi anlaşılabilmesi için öncelikle algoritmada yer alan temel kavramlar açıklanmıştır:

- Gen, probleme has en küçük bilgiyi taşıyan birim olup, yapısı probleme özgü olarak farklı şekillerde kodlanabilir (encoding).
- Kromozom, birden fazla genin bir araya gelerek oluşturduğu ve problem çözümünü temsil eden yapıdır.
- Popülasyon, uygun çözümleri temsil eden kromozomlar tarafından oluşturulan topluluktur. Popülasyonda yer alan kromozom sayısı ele alınan problemin yapısına göre karar verilen bir parametredir.
- Daha kaliteli nesiller üretmek için kullanılan genetik operatörler ise seçim, çaprazlama ve mutasyon operatörlerini içermektedir.



Şekil 7. Önerilen Tahmin Süreci

Şekil 7’de çalışma prensipleri verilen GA’ nın öznitelik seçim problemi için uygulama adımları izleyen bölümde detaylı bir şekilde açıklanmıştır:

- Algoritma parametreleri olan popülasyon büyüklüğü, turnuva büyüklüğü, çaprazlama oranı, mutasyon oranı ve maksimum iterasyon sayısı sırasıyla 20; 3; 0,90; 0,10 ve 100 olarak belirlenmiştir.
- Başlangıç popülasyonu oluşturulurken popülasyon büyüklüğü kadar birey tamamen rassal olarak oluşturulmuştur ve ikili kodlama (binary encoding) yapısı kullanılmıştır. Buna göre Şekil 8’de görüldüğü gibi kromozomda yer alan her bir gen bir özneliğe karşılık gelmektedir. Gen değerinin “1” olduğu durumda, ilgili öznelik modelde kullanılmakta, “0” olduğu durumda ise kullanılmamaktadır.

Öznitelikler:	X_1	X_2	X_3	X_4	X_5	X_6	...	X_n
	×	×	✓	✓	×	✓	...	×
Kromozom:	0	0	1	1	0	1	...	0

Şekil 8. Kromozom yapısı

- Başlangıç popülasyonu oluşturulduktan sonra popülasyonda yer alan bireyler uyum değerlerine göre değerlendirilir. Her birey özneliklerin oluşturduğu bir alt kümeyi temsil etmektedir. Uyum değeri olarak ise regresyon modelinin determinasyon katsayısı (R^2) kullanılmıştır.
- Seçim operatörü olarak turnuva seçim yöntemi kullanılmıştır. Buna göre turnuva büyüklüğü kadar birey popülasyondan rassal olarak seçilir ve uyum değeri en iyi olan yani R^2 değeri en yüksek olan çözüm ebeveyn olarak seçilir.
- Seçim sürecinden sonra seçilen ebeveynler yeni bireylerin oluşturulması amacıyla çaprazlanır. Çaprazlama sürecinde ise Şekil 9’da görüldüğü gibi ikili kodlamada sıklıkla kullanılan tek nokta çaprazlama kullanılmıştır. Buna göre kromozomda rassal olarak bir nokta seçilir. Bu noktadan önceki genler birinci sonraki genler ise ikinci ebeveynden gelecek şekilde iki yeni birey oluşturulur.

Ebeveyn 1:	0	0	1	1	0	1	1	0
Ebeveyn 2:	1	0	1	1	0	0	1	1
Çocuk 1:	0	0	1	1	0	0	1	1
Çocuk 2:	1	0	1	1	0	1	1	0

Şekil 9. Tek nokta çaprazlama

- Algoritmanın yerel en iyiye takılmasını önlemek amacıyla yeni üretilen bireyler için mutasyon oranına bağlı olarak mutasyon operatörü uygulanır. Mutasyon operatörü uygulanırken kromozom içerisinde rassal olarak seçilen bir gen “1” ise “0”; “0” ise “1” değerini alır.
- Genetik operatörler uygulandıktan sonra yeni oluşturulan bireyler, popülasyonda yer alan bireylerle karşılaştırılarak popülasyon güncellenir.
- Durma kriteri sağlanana kadar yukarıdaki adımlar uygulanır, durma kriteri sağlanınca uyum değeri en iyi olan birey en iyi çözüm olarak raporlanır. İncelenen problem için durma kriteri olarak maksimum iterasyon sayısı kullanılmıştır.

3. Bulgular

3.1. Performans Ölçütleri

Geliştirilen modeller determinasyon katsayısı (R^2), ortalama mutlak hata (Mean Absolute Error, MAE) ve ortalama hata karesi kökü (Root Mean Squared Error, RMSE) performans ölçütlerine göre karşılaştırılmıştır.

- Determinasyon Katsayısı (R^2): Korelasyon katsayısının karesi olan determinasyon katsayısı, bağımlı değişkendeki değişkenliğin ne kadarının bağımsız değişkenlerdeki değişkenlikle açıklanabildiğini ifade eder. İki değişken arasındaki doğrusal ilişkinin gücünü belirleyen determinasyon katsayısı 0-1 aralığında değer almakta olup ilişkinin gücü 1’e yaklaştıkça artmaktadır.
- Ortalama Mutlak Hata (Mean Absolute Error, MAE): Model tahmini ile hedef değer arasındaki mutlak farkın ortalamasını verir.

- Kök Ortalama Hata Karesi (Root Mean Squared Error, RMSE): Tahmin hatalarının standart sapmasıdır.
- y_i : i . gözlemin hedef değeri; \hat{y}_i : i . gözlemin tahmin edilen değeri; \bar{y} : hedef değerlerin ortalaması, n : örneklem büyüklüğü ve e_i : i . gözleme ilişkin hata ve $e_i = y_i - \hat{y}_i$ olmak üzere, modeller karşılaştırılırken kullanılan performans ölçütleri Çizelge 3'te verilmiştir.

Çizelge 3. Performans Ölçütleri

Açıklama	Formül
R^2	$= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
MAE	$= \frac{1}{n} \sum_{i=1}^n e_i $
RMSE	$= \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2}$

3.2. Karşılaştırmalı sonuçlar

Bu bölümde öncelikle öznitelik seçimi yapılmadan Çizelge 4'te verilen modeller karşılaştırılmıştır. Çizelge 4'te görüldüğü gibi ilk dört modelde gecikme değişkenleri kullanılmazken beşinci modelde Eşitlik (1) de belirtildiği gibi gecikme değişkenleri de ilave edilerek toplam 47 öznitelik kullanılmıştır.

Çizelge 4. Önerilen Modeller

Model	Değişkenler	Değişken sayısı
1	rüzgar hızı, rüzgar yönü	2
2	rüzgar hızı, rüzgar yönü, teorik güç eğrisi	3
3	rüzgar hızı, rüzgar yönü, meteorolojik değişkenler	10
4	rüzgar hızı, rüzgar yönü, teorik güç eğrisi, meteorolojik değişkenler	11
5	Model 4+ gecikme (lag) değişkenleri	47

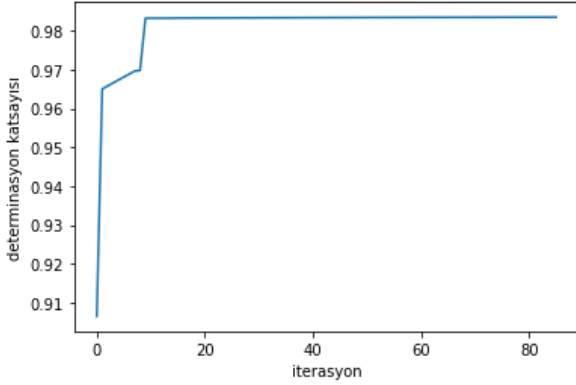
Çizelge 5'te görüldüğü gibi determinasyon katsayısı en yüksek olan model gecikme değişkenlerinin modele dahil edildiği beşinci modeldir. Yine beşinci modelde ortalama mutlak hata ve kök ortalama hata karesi değerleri diğer modellere göre daha düşük bulunmuştur.

Çizelge 5. Modellerin karşılaştırılması

Model 1:	R^2	MAE	RMSE
Çoklu Doğrusal Regresyon	0,844	353,583	502,740
Lasso Regresyon	0,844	353,618	502,738
Ridge Regresyon	0,844	353,585	502,740
Elastik Net	0,843	359,159	503,657
Karar Ağaçları	0,849	190,242	494,296
Rassal Ormanlar	0,904	169,702	393,838
Model 2:	R^2	MAE	RMSE
Çoklu Doğrusal Regresyon	0,909	173,155	383,387
Lasso Regresyon	0,909	173,089	383,392
Ridge Regresyon	0,909	173,154	383,387
Elastik Net	0,909	172,359	383,658
Karar Ağaçları	0,842	193,356	506,329
Rassal Ormanlar	0,906	165,469	390,059
Model 3:	R^2	MAE	RMSE
Çoklu Doğrusal Regresyon	0,856	347,524	482,809
Lasso Regresyon	0,852	353,297	489,667
Ridge Regresyon	0,854	351,091	486,466
Elastik Net	0,848	360,071	496,420
Karar Ağaçları	0,928	116,940	341,406
Rassal Ormanlar	0,957	99,554	263,208
Model 4:	R^2	MAE	RMSE
Çoklu Doğrusal Regresyon	0,921	169,411	357,754
Lasso Regresyon	0,916	180,118	368,278
Ridge Regresyon	0,919	175,621	363,171
Elastik Net	0,913	186,291	374,894
Karar Ağaçları	0,935	109,528	323,198
Rassal Ormanlar	0,959	95,722	256,888
Model 5:	R^2	MAE	RMSE
Çoklu Doğrusal Regresyon	0,983	76,454	164,878
Lasso Regresyon	0,983	75,044	164,242
Ridge Regresyon	0,983	75,984	164,282
Elastik Net	0,983	73,372	165,008
Karar Ağaçları	0,966	89,633	233,145
Rassal Ormanlar	0,982	70,273	170,941

Kullanılan yöntemler değerlendirildiğinde ise Çoklu Doğrusal Regresyon, Lasso Regresyon, Ridge Regresyon ve Elastik Net sonuçları birbirine yakın bulunurken değişken sayısının az olduğu Model 1, Model 2, Model 3 ve Model 4 için Rassal Orman Algoritması daha iyi performans göstermiştir. Model 5'te ise Karar Ağaçları dışında diğer yöntemler

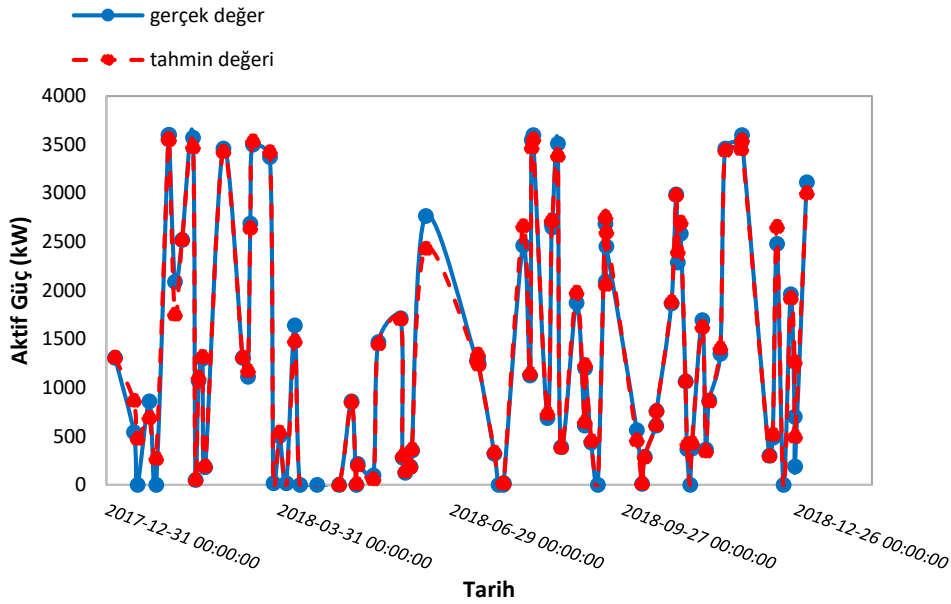
birbirine yakın performans gösterirken Karar Ağaçları diğer yöntemlere kıyasla daha başarısız bulunmuştur.



Şekil 10. GA performansı

İkinci aşamada ise Model 5 için öznitelik seçiminde GA'nın performansı değerlendirilmiştir. GA ile elde edilen modelde sadece dokuz değişken kullanılarak Şekil 10'da görüldüğü gibi ilk yirmi iterasyonda determinasyon katsayısı 0,906'dan 0,983 değerine ulaşmıştır. Böylece modele gereksiz şekilde dahil edilmeleri, model karmaşıklığını artırarak modelin yorumlanmasını zorlaştırabilecek olan değişkenler modelden uzaklaştırılmıştır.

Şekil 11 'de öznitelik seçim sürecinin ardından önerilen nihai modelin performansı değerlendirilmiştir. Şekil 11'de görüldüğü gibi test verisi ve tahmin değerleri karşılaştırıldığında tahmin değerlerinin gerçek değerlere oldukça yakın olduğu görülmektedir.



Şekil 11. Tahmin performansı

4. Tartışma ve Sonuç

Yapılan çalışmada makine öğrenme teknolojisi, son zamanlarda popülaritesi artan, yenilenebilir enerji kaynaklarından rüzgar enerjisi tahmini için kullanılmıştır.

Çalışmanın ilk aşamasında kullanılan bağımsız değişkenlere göre farklı regresyon modelleri geliştirilmiştir. Sadece SCADA değişkenlerinin kullanıldığı model için Çoklu Doğrusal Regresyon modelinde R^2 değeri 0,909 olarak bulunurken, NASA tarafından paylaşılan meteorolojik değişkenlerin modele dahil edilmesiyle R^2 değeri 0,921 olarak elde edilmiştir. Gecikmeli değişkenlerin de kullanıldığı

toplam 47 değişkenin bulunduğu modelde ise R^2 değeri 0,983'e çıkmıştır.

Çalışmada ayrıca farklı makine öğrenme algoritmalarının performansları da karşılaştırılmıştır. Değişken sayısının az olduğu modellerde Rassal Orman algoritması diğer yöntemlere göre daha iyi performans göstermiştir. Gecikme değişkenlerinin kullanıldığı modelde ise Karar Ağaçları dışındaki yöntemler oldukça benzer performans gösterirken Karar Ağaçlarının performansı diğer yöntemlere göre düşük bulunmuştur.

Regresyon modelinde modelin performansına etki etmeyen değişkenlerin modelde kullanılması hem modelin eğitilmesi için gereken süreyi artırarak modelin etkinliğini sınırlandırmakta hem de büyük miktarda öznitelik kullanımı ile tahminlerde

belirsizlik oluşabilmektedir. Bu nedenle yapılan çalışmada makine öğrenme algoritmalarından etkin düzeyde yararlanılması amacıyla öznitelik seçim prosedürü uygulanmıştır. Öznitelik seçim sürecinde GA'dan yararlanılmıştır. Algoritma sonuçlarına göre ilk 20 iterasyon gibi oldukça kısa bir sürede popülasyonda yer alan en iyi bireyin R^2 değeri 0,906 dan 0,983 değerine ulaşmıştır. Böylece 47 değişkenli modelin R^2 değerine sadece 9 değişkenin kullanıldığı model ile ulaşılarak etkin ve güçlü bir regresyon modeli geliştirilmiştir.

Gelecekte yapılması planlanan çalışmalarda kullanılan makine öğrenme algoritmalarına ilişkin hiperparametre (hyperparameter) değerleri sezgisel algoritmalarla belirlenerek model performansları karşılaştırılabilir. Lasso ve Ridge Regresyon yöntemlerinde kullanılan λ değerinin belirlenmesi için çapraz doğrulama yöntemi kullanılarak sonuçlar detaylı bir şekilde analiz edilebilir. Bunun dışında benzer bir tahmin süreci, güneş enerjisi gibi farklı yenilenebilir enerji kaynakları için de uygulanabilir.

5. Kaynaklar

- Azimi R, Ghofrani M, Ghayekhloo M, 2016, A hybrid wind power forecasting model based on data mining and wavelets analysis. *Energy conversion and management*, **127**, 208-225.
- Belkin M, Hsu D, Ma S, Mandal S, 2019, Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, **116**, **32**, 15849-15854.
- Breiman L, 2001, Random forests. *Machine learning*, **45**, **1**, 5-32.
- Demolli H, Dokuz AS, Ecemis A, Gokcek M, 2019, Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Conversion and Management*, **198**, 111823.
- Deng Y-C, Tang X-H, Zhou Z-Y, Yang Y, Niu F, 2021, Application of machine learning algorithms in wind power: a review. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 1-22.
- Foley AM, Leahy PG, Marvuglia A, McKeogh EJ, 2012, Current methods and advances in forecasting of wind power generation. *Renewable energy*, **37**, **1**, 1-8.
- Heinermann J, Kramer O, 2016, Machine learning ensembles for wind power prediction. *Renewable Energy*, **89**, 671-679.
- Higashiyama K, Fujimoto Y, Hayashi Y, 2018, Feature extraction of NWP data for wind power forecasting using 3D-convolutional neural networks. *Energy Procedia*, **155**, 350-358.
- Hoerl AE, Kennard RW, 1970, Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**, **1**, 55-67.
- Holland JH, 1992. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, A Bradford Book MIT Press, p.
- Kuhn M, Johnson K, 2013. *Applied predictive modeling*, New York: Springer, 26, 13.
- Le TT, Fu W, Moore JH, 2020, Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics*, **36**, **1**, 250-256.
- Liu H, Chen C, 2019, Data processing strategies in wind energy forecasting models and applications: A comprehensive review. *Applied Energy*, **249**, 392-408.
- Liu X, Zhang H, Kong X, Lee KY, 2020, Wind speed forecasting using deep neural network with feature selection. *Neurocomputing*, **397**, 393-403.
- Lu P, Ye L, Zhao Y, Dai B, Pei M, Li Z, 2021, Feature extraction of meteorological factors for wind power prediction based on variable weight combined method. *Renewable Energy*, **179**, 1925-1939.
- Melkumova L, Shatskikh SY, 2017, Comparing Ridge and LASSO estimators for data analysis. *Procedia Engineering*, **201**, 746-755.
- Olson RS, Moore JH, TPOT: A tree-based pipeline optimization tool for automating machine learning, *Workshop on automatic machine learning*, 66-74.
- Renani ET, Elias MFM, Rahim NA, 2016, Using data-driven approach for wind power prediction: A comparative study. *Energy Conversion and Management*, **118**, 193-203.
- Salcedo-Sanz S, Pastor-Sánchez A, Prieto L, Blanco-Aguilera A, García-Herrera R, 2014, Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization–Extreme learning machine approach. *Energy Conversion and Management*, **87**, 10-18.
- Shi J, Wang L, Lee W-J, Cheng X, Zong X, 2019, Hybrid Energy Storage System (HESS) optimization enabling very short-term wind power generation scheduling

based on output feature extraction. *Applied Energy*, **256**, 113915.

Sun S, Qiao H, Wei Y, Wang S, 2017, A new dynamic integrated approach for wind speed forecasting. *Applied Energy*, **197**, 151-162.

Theofilatos A, Chen C, Antoniou C, 2019, Comparing machine learning and deep learning methods for real-time crash prediction. *Transportation research record*, **2673**, **8**, 169-178.

Tibshirani R, 1996, Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**, **1**, 267-288.

Tso GK, Yau KK, 2007, Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, **32**, **9**, 1761-1768.

Wang H, Sun J, Sun J, Wang J, 2017, Using random forests to select optimal input variables for short-term wind speed forecasting models. *Energies*, **10**, **10**, 1522.

Wang K, Qi X, Liu H, Song J, 2018, Deep belief network based k-means cluster approach for short-term wind power forecasting. *Energy*, **165**, 840-852.

WCED SWS, 1987, World commission on environment and development. *Our common future*, **17**, **1**, 1-91.

Zhang S, 2012, Nearest neighbor selection for iteratively kNN imputation. *Journal of Systems and Software*, **85**, **11**, 2541-2552.

Zou H, Hastie T, 2005, Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, **67**, **2**, 301-320.

internet kaynakları

- 1- <https://sdgs.un.org/goals>, (16.04.2022)
- 2- <https://www.globalgoals.org/goals/7-affordable-and-clean-energy/>, (16.04.2022)
- 3- https://www.ren21.net/wp-content/uploads/2019/05/GSR2021_Full_Report.pdf, (14.04.2022)
- 4- <https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset>, (16.05.2022)
- 5- <https://www.tureb.com.tr/>, (16.05.2022)
- 6- <https://www.renewables.ninja/>, (16.05.2022)