

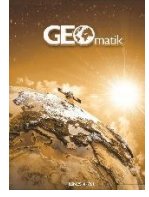
PAPER DETAILS

TITLE: Farkli makine öğrenmesi yaklasimlari ile Türkiye'nin solar radyasyon tahmini

AUTHORS: Taha Demirgöl,Vahdettin Demir,Mehmet Faik Sevimli

PAGES: 106-122

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/3467843>



Farklı makine öğrenmesi yaklaşımları ile Türkiye'nin solar radyasyon tahmini

Taha Demirgöl ^{*1}, Vahdettin Demir ¹, Mehmet Faik Sevimli ¹

¹ KTO Karatay Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, İnşaat Mühendisliği Bölümü, Konya, Türkiye, demirgultaha@gmail.com; vahdettin.demir@karatay.edu.tr; mehmet.faik.sevimli@karatay.edu.tr

Kaynak Göster: Demirgöl, T., Demir, V., & Sevimli, M. F. (2024). Farklı makine öğrenmesi yaklaşımları ile Türkiye'nin solar radyasyon tahmini. Geomatik, 9 (1), 106-122

<https://doi.org/10.29128/geomatik.1374383>

Anahtar Kelimeler

Solar Radyasyon
HELIOSAT
Makine Öğrenmesi
3600 Grid
Türkiye

Araştırma Makalesi

Geliş: 11.10.2023
Revize: 16.11.2023
Kabul: 20.11.2023
Yayınlanma: 05.02.2024



Öz

Solar radyasyon (SR), enerji dönüşümü, yeşil bina konsepti, meteoroloji, küresel iklim değişikliği, tarım ve hayvancılık ile ilgili çalışmalar için önemli bir parametredir. Solar radyasyonun belirlenebilmesi için gereken alıcıların tüm noktalar için temin edilememesinden dolayı bu parametrenin çeşitli yöntemlerle tahmin edilmesi gerekmektedir. Bu çalışmada, Türkiye'nin 3600 grid noktasının 2004-2021 yıllarına ait yıllık ortalama solar radyasyon değerleri (kWsa/m^2) kullanılarak ülke çapında farklı test noktaları için solar radyasyon tahmini gerçekleştirilmiştir. Solar radyasyon değerleri, çok değişkenli uyarlanabilir regresyon eğrileri (MARS) ve en küçük kareler destek vektör regresyonu (LSSVR) olmak üzere 2 farklı makine öğrenmesi tekniği kullanılarak MATLAB platformunda tahmin edilmiştir. Solar radyasyon haritaları için ise ters mesafe ağırlıklı enterpolasyon tekniği kullanılmıştır. Tahmin edilen veriler ArcMap ortamında haritalandırılmıştır. Solar radyasyon, komşu ölçüm grid noktalarına ait konum bilgileri kullanılarak tahmin edilmiştir. Modellerde kullanılan veriler, Meteoroloji Genel Müdürlüğü'nden (MGM) temin edilmiş uydu tabanlı model olan heliospheric optical satellite model (HELIOSAT) verileridir. Farklı kombinasyonlar kullanılarak test noktalarından elde edilen SR tahminleri gözlenen verilerle karşılaştırılmıştır. Bu karşılaştırmalarda, karekök ortalama karesel hata, ortalama mutlak hata, ortalama mutlak bağıl hata, Nash-Sutcliffe modeli verimlilik katsayısı ve determinasyon katsayısı yöntemleri kullanılmıştır. Tahmin edilen SR değerlerine ait gidiş, saçılma grafikleri, Taylor ve Violin diyagramları oluşturulmuştur. Ayrıca Kruskal-Wallis testi ile Wilcoxon testi uygulanmıştır. Makine öğrenmesi yöntemlerinden LSSVR çok başarılı tahmin sonuçları vermiştir. Böylece, makine öğrenme algoritmalarının literatürde yer alan kabul görmüş geleneksel yöntemlere göre daha kolay ve alternatif bir yöntem olabileceği gösterilmiştir.

Türkiye's solar radiation forecasting with different machine learning approaches

Keywords

Solar Radiation
HELIOSAT
Machine Learning
3600 Grid
Türkiye

Research Article

Received: 11.10.2023
Revised: 16.11.2023
Accepted: 20.11.2023
Published: 05.02.2024

Abstract

SR is an important parameter for studies related to energy conversion, green building concept, meteorology, global climate change, agriculture and animal husbandry. Since the receivers required to determine solar radiation are not available for all points, this parameter needs to be estimated by various methods. In this study, the annual average solar radiation values (kWh/m^2) of Türkiye's 3600 grid points for the years 2004-2021 were used to estimate solar radiation for different test points across the country. SR values were estimated using two different machine learning techniques, namely MARS and LSSVR on MATLAB platform. Inverse distance weighting interpolation technique was used for solar radiation maps. The estimated data were mapped in ArcMap environment. SR was estimated using the location information of neighboring measurement grid points and the periodicity component of year values. The data used in the models are the data from HELIOSAT, obtained from the MGM. SR estimates obtained from the test points using different combinations were compared with the observed data. In these comparisons, root mean square error, mean absolute error, mean absolute relative error, Nash-Sutcliffe model efficiency coefficient and coefficient of determination methods were used. Grid-based variation, scatter graphs, Taylor and Violin diagrams of the estimated SR values were created. In addition, Kruskal-Wallis test and Wilcoxon test were applied. LSSVR, one of the machine learning methods, gave very successful prediction results. Thus, it is shown that machine learning algorithms can be an easier and alternative method compared to the traditional methods accepted in the literature.

1. Giriş

Dünyamız için elzem bir enerji kaynağı olan güneş yüksek enerji potansiyeli olan çevreci bir enerji kaynağıdır (Kılıç ve Kumaş, 2019). Geçmişten günümüze kullanılmakta olan yersel ölçüm istasyonları Solar Radyasyon (SR) değerlerini ölçüm aletleri vasıtasıyla ölçmektedir. Fakat, bu aletlerin maddi ve insan gücü maliyetleri ve kalibrasyon ihtiyacı yüzünden global çapta çoğu bölgede ölçümü yapılamamaktadır. Solar (Güneş) enerji, gezegenimizde bulunan enerjinin büyük bir kısmının kaynağıdır ve günümüzde yaşanan fosil enerji kaynaklarına ulaşımın güçlüğü bazı faktörlerden ötürü arttığından yakın gelecekte, bilhassa da gelişmekte olan ülkelerde çok büyük bir önem kazanacağı beklenmektedir (Guermoui ve ark., 2020). Güneşten çıkıp atmosferi geçerek yeryüzüne ulaşan SR, ulaştığı alanın iklim koşullarını etkiler. Yapılarda enerji sistemlerinin tasarımı, güneş enerjisi sistemlerinin kurulacağı alanın doğru tespiti ve yapılardaki ısı ve ışık miktarının belirlenebilmesi için SR bilgisine sahip olmak gerekmektedir. Bu amaç için en iyi veri kaynağı, yapılması planlanan güneş sisteminin lokasyonunda uzun dönem gözlenen verileridir. Ancak SR gözlem istasyonlarının yetersizliği, bu parametreyi elde etmekte güçlük çıkarmakta ve SR ölçüm yöntemlerinin geliştirilmesi ihtiyacının her geçen gün önemini arttırmaktadır (Ahmad ve Tiwari, 2011). Güneşten yeryüzüne ulaşan elektromanyetik enerji olarak tanımlanabilen SR, günümüzde güneş ışınları ile doğrudan ya da dolaylı ilişkisi olan, başta güneş panelleri ile elektrik üretiminde kullanılan santrallerde, tarım ve hayvancılıkta, meteorolojik araştırmalarda, doğal aydınlatma sistemlerinde ve seralarda olmak üzere pek çok farklı alanda kullanılmaktadır (Jiang, 2009; Şeker, 2021). SR, tarımda bitkilerin gelişme aşaması, büyümesi ve biyokütle birikimi için referans terleme-buharlaşmanın hesaplanmasında önemli bir etken olup (Çitakoğlu, 2015), günümüzde iklim değişimi ile ilgili olası senaryoların incelenmesinde SR verileri kullanılmaktadır (Bora ve ark., 2015). Güneş enerjisi santrallerinin tasarım aşamasında SR değerlerinin bilinmesi elzemdir (Çitakoğlu ve Demir, 2021). Bu nedenle SR'nin alternatif yöntemlerle tahmin edilebilmesi, güneş enerjisi santrallerinin tasarlanma ve geliştirilme aşamaları için büyük bir öneme sahiptir. Yapılarda ısınma veya soğutma, atmosferik enerji dengesi çalışmaları doğal veya yapay yollarla biriken temiz su kaynaklarının buharlaşma seviyesinin dengede tutulması ve meteorolojik tahminler gibi pek çok konuda SR değerlerinin bilinmesi gerekmektedir (Özgören ve ark., 2012). Çalışılacak olan coğrafi bölgelerde solar radyasyon değerlerinin doğru tahmin edilmesi, birçok mühendislik, akademik, kamu kurumları ve özel sektör çalışmaları için oldukça önem arz etmektedir. Türkiye'nin yıllık ortalama SR verileri incelendiğinde 2005 yılındaki günlük ışınlam değeri 3,6 kWsa/m² olarak kaydedilmiş olup, toplam yıllık radyasyon dönemi yaklaşık 2610 saat olarak hesaplanmıştır (Sözen ve Arcaklioğlu, 2005). 2022 yılında ise T.C. Enerji ve Tabii Kaynakları Bakanlığı resmî web sitesinde yayınlanan, Türkiye Güneş Enerjisi Potansiyeli Atlasına (GEPA) göre, günlük ışınlam değeri 4,2 kWsa/m² olarak gözlenmiş olup,

ortalama yıllık toplam güneşlenme süresi 2.741 saat olarak hesaplanmıştır (Enerji İşleri Genel Müdürlüğü, 2023). Gözlem yapılan coğrafi bölgelerde ölçülen güneşlenme süresi, rüzgâr hızı, hava basıncı, nispi nem, buhar basıncı ve hava sıcaklığı gibi meteorolojik değerlerin SR'nin en doğru şekilde ölçülmesi için kullanılabilecek önemli parametreler olduğu belirtilmektedir (Kılıç ve Kumaş, 2019). Ancak SR'nin ölçülmesi yukarıda anlatılan sebeplerden ötürü zaman, insan gücü ve ekipman/bakım maliyetleri açısından çeşitli dezavantajlara sahiptir. Bu nedenle, uydu tabanlı veya gözlem istasyonlarından elde edilen SR değerlerinden faydalanılarak tahmin yöntemlerinin geliştirilmesine ve test edilmesine olan ilgi gün geçtikçe artmaktadır (Gül ve Çelik, 2017). SR, günümüzde başta mühendislik olmak üzere pek çok alanda yapılan bilimsel faaliyetler için ihtiyaç duyulan önemli bir değişken haline gelmiştir. Gözlenen verinin yetersizliğinden ötürü, bu gözlenen verilerden yola çıkarak SR gözlemi yapılamayan istasyonların verilerinin elde edilmesine ihtiyaç duyulmaktadır. Son yıllarda SR tahmini için gözlem verileri referans alınarak çalışılan metotlar artmaktadır (Alizamir ve ark., 2020). SR değerleri nem, sıcaklık, rüzgâr hızı ve bulut örtüsü gibi günümüzde kolaylıkla elde edilebilen iklimsel değişkenler kullanılarak tahmin edilebilmektedir. Günümüzde, SR değerlerini tahmin etmek için pek çok yöntem önerilmiştir. Bunlardan geleneksel olanları matematiksel denklemlerle elde edilen ampirik yöntemlerdir. Ampirik modellerin avantajlı yönü kolay hesaplanabilir olması ve SR değerlerini tahmin etmede faydalı bir yöntem olarak kabul görmüş olmasıdır. Ampirik yöntemlerin dezavantajlı yönü ise kısa dönem SR değerlerini tahmin ederken bulut örtüsü, nem miktarı, rüzgâr hızı ve yağışlı geçen süreler gibi hava koşullarındaki ani değişimler sebebiyle kısa dönem güneş radyasyonu verilerini isabetli bir şekilde tahmin edemiyor olmasıdır. Ayrıca, literatürde bazı araştırmacıların bu yöntemlerin, yağışlı günlerde yoğun bulutlardan SR'nin büyük ölçüde etkilendiğini, nem oranı yüksek bölgelerde hem bağımlı hem de bağımsız parametreler arasındaki karmaşık ve nonlineer ilişkileri aktaramadıklarını belirtmişlerdir (Ağbulut, ve ark., 2021). Bu sebeplerden ötürü meteorolojik değişkenler yerine coğrafi değişkenlerle yapılan tahminler (enlem, boylam, periyodiklik ve yükseklik gibi) daha kolay ve ucuzdur. Günümüzde ampirik modellere alternatif olan makine öğrenimi modelleri gibi modeller isabetli sonuçlar vermesi sebebinden ötürü SR tahminleri için alternatif bir yol olarak görülmektedir. Literatürde yaygın olarak tercih edilen yöntemler yapay zekâ ve makine öğrenmesi uygulamalarıdır. Çeşitli yenilenebilir enerji kaynakları arasında, çevre dostu ve yüksek potansiyel barındıran güneş enerjisi, endüstriyel, kentsel ve tarımsal üretim için umut verici bir enerji kaynağı olarak kabul edilmektedir.

3600 grid noktasıyla yapılan akademik çalışmalar için Google Akademik, TRDizin, Web of Science, Scopus ve Ulusal Tez Merkezi de dahil olmak üzere literatür taraması yapılmış fakat bir çalışma bulunamamıştır. Bu makalenin özgünlüğü 3600 grid noktası ile tahmin yapılmış olmasıdır. SR tahmini ile ilgili yapılan bazı önemli çalışmalar incelendiğinde, Bhardwaj ve ark.

(2013), SR tahmini için gizli markov modeli ve genelleştirilmiş bulanık modelin bir kombinasyonunu kullanarak şekil tabanlı kümeleme modelini kullandıkları görülmektedir. Araştırmacılar, 6 giriş (ortam sıcaklığı, nispi nem, atmosfer basıncı, rüzgâr hızı, rüzgâr yönü, güneş spektrumu) ve 1 çıkış (SR tahmini) kullanmışlardır. Çalışma sonucunda kullanılan modelin başarılı olduğunu belirtmişlerdir. Wang ve ark. (2017), Çin'deki 21 değişik bölge için günlük küresel SR tahmininde grid bölmeli adaptive-network based fuzzy inference systems (ANFIS)-SC, ANFIS-GP ve M5-Tree yöntemlerini çalışmışlar ve 7 giriş (nispi nem, güneşlenme süresi, ortalama sıcaklık, hava basıncı, maksimum sıcaklık, minimum sıcaklık ve su buharı basıncı) 1 çıkış (SR tahmini) belirlemişlerdir. Çalışma sonucunda üç farklı model içinde en iyi sonucu M5-Tree yönteminin verdiğini ve en başarılı sonucu veren bölgenin ortalama mutlak hata (OMH) değerinin 2,1 (MJ/m²gün), karekök ortalama karesel hata (KOKH) değerinin 2,79 (MJ/m²) ve determinasyon katsayısının (R²) 0,85 olduğunu, en başarısız sonuç veren istasyonun ise OMH değerinin 2,95 (MJ/m²gün), KOKH değerinin 3,87 (MJ/m²) ve R²'nin 0,74 olduğu görülmektedir. Cornejo ve ark. (2019), İspanya'nın Toledo şehri için 1 yıllık saatlik SR değerleriyle destek vektör regresyonu (SVR), aşırı öğrenme makineleri (ELM), gauss süreç regresyonu (GPR) ve çok katmanlı algılayıcı (MLP) son teknoloji makine öğrenmesi yöntemleriyle yapmış oldukları çalışmada farklı makine öğrenmesi regresyon tekniklerinin başarı seviyelerini analiz etmişler ve kıyaslanan yöntemlerden en iyi sonucu ELM'nin verdiğini tespit etmişlerdir. Demirgöl ve ark. (2023) yapmış oldukları çalışmada Türkiye'nin aylık ortalama SR değerlerini (kWsa/m²), M5 model ağacı (M5-tree), sezgisel regresyon tekniği kullanarak yapmış oldukları SR modellemesinde komşu ölçüm istasyonlarına ait konum bilgileri ve periyodiklik bileşeni olan ay değerlerini kullanarak tahminler gerçekleştirmişlerdir. Modellerde için MGM'den temin edilen uydu tabanlı hibrit bir model olan HELIOSAT model verilerinin 2004-2018 yıllarını kapsayan uzun dönem aylık ortalama SR verilerini kullanan araştırmacılar çalışmalarında Türkiye'nin 81 iline ait ölçüm ortalamalarını kullanmıştır. 81 noktadan alınan 12 aylık verilerin %75'i eğitim sürecinde kullanılırken %25'i test sürecinde kullanan araştırmacılar rastgele seçilen test istasyonlarından elde edilen solar radyasyon tahminleri gözlenen verilerle karşılaştırmışlardır. Karşılaştırma kriteri olarak KOKH, OMH, Ortalama Mutlak Bağlı Hata (OMBH) ve R² kullanılmışlardır. Güzel ve ark. (2023), Türkiye'nin Isparta ili için MGM'den temin ettikleri meteorolojik verileri kullanarak SR tahmini yapmışlardır. Yapmış oldukları çalışmada rastgele orman, k-en yakın komşu (k-nearest neighbor (k-NN)), yapay sinir ağları ve derin öğrenme yöntemlerini kullanmışlardır. Ele ettikleri sonuçlar içinde en iyi sonuçları yapay sinir ağları ve derin öğrenme yöntemlerinin verdiğini belirtmişlerdir.

Bu çalışmanın amacı, Meteoroloji Genel Müdürlüğü (MGM)'den temin edilmiş olan 2004-2021 yılları arasında kapsayan, heliospheric optical satellite (HELIOSAT) modeli ile işlenen, 3600 grid noktasına ait uydu verileri kullanılarak Türkiye'de çeşitli veri paketleri için SR

değerlerinin tahmin edilmesi üzerinedir. Çalışmada, solar radyasyon değerlerinin tahmin edilmesi için 2004-2021 yılları arasındaki 3600 grid noktasına ait veriler 2 farklı makine öğrenmesi yönteminde uygulanmıştır. SR değerlerinin tahmin edilmesi için kullanılan makine öğrenmesi yöntemleri, çok değişkenli uyarlanabilir regresyon eğrileri (MARS) ve en küçük kareler destek vektör regresyonu (LSSVR)'dur. Karşılaştırma ve hata kriteri olarak ise literatürde tahmin konulu çalışmalarda çokça tercih edilen kriterlerden KOKH, OMH, OMBH, R² ve Nash-Sutcliffe eşitliği (NSE) kullanılmıştır. SR haritaları için ise ters mesafe ağırlıklı (IDW) enterpolasyon yöntemi kullanılmıştır. Bu çalışmanın hedefi ise MGM'den temin edilen veri setlerinin Türkiye'nin uzun dönem yıllık ortalama SR değerlerini makine öğrenmesi yöntemlerinin eğitilmesi ve test edilmesi için kullanarak, kullanılacak yöntemlerin avantajlı ve dezavantajlı yönlerinin ortaya konulması, geçmiş (eksik verilerin tamamlanması) yıllar için SR değerlerinin tahmin edilmesidir. Türkiye'nin 3600 grid noktasına ait uzun dönem yıllık ortalama solar radyasyon değerlerinin çeşitli kombinasyonları kullanılarak elde edilen tahmin değerleri beş farklı karşılaştırma ve hata kriteri ile değerlendirilmiş ve bu tahminler kullanılarak IDW enterpolasyon yöntemi ile ArcMap platformunda Türkiye'nin SR haritaları oluşturulmuştur. İki farklı makine öğrenmesi yöntemi ile elde edilen en iyi paketleri ve bu paketlere ait KOKH değerleri ile detaylı KOKH haritaları oluşturulmuştur. KOKH haritaları ile hataların dağılımı incelenmiş ve yorumlanmıştır. Bu sayede yakın geleceğimiz için hayati risk oluşturabilecek kuraklık, su kıtlığı gibi doğal afetler için önceden tedbir alınabilmesinde ve yenilenebilir enerjilerin başında gelen güneş enerjisi santralleri için en verimli bölgelerin belirlenebilmesinde fikir oluşturması hedeflenmektedir. Böylece, makine öğrenme algoritmalarının literatürde yer alan kabul görmüş geleneksel yöntemlere göre daha kolay ve alternatif bir yöntem olabileceği gösterilmiştir.

Bu çalışmada, makine öğrenmesi yöntemleri ile modellerin etkisinin araştırılması için elde edilen sonuçların grafiksel karşılaştırma (Taylor, Violin diyagramları, gidiş ve saçılma grafikleri) yoluyla değerlendirilmesi yapılmıştır. Ayrıca Kruskal-Wallis ve Wilcoxon testleri yapılmıştır. Çalışmanın hedefleri detaylandırılacak olursa; temin edilen 3600 grid noktasına ait verilerin eğitim ve test grubu olarak sınıflandırılma işlemi için uygun formata dönüştürülmesi sağlanmıştır. Gruplandırma için MGM'den temin edilen 3600 grid noktasına ait SR değerleri %75 eğitim (2700 grid) %25 test (900 grid) olacak şekilde 4 farklı veri paketi (M1, M2, M3 ve M4) oluşturulmuştur. Makine öğrenmesi yöntemlerinin karşılaştırılması yapılarak avantaj veya dezavantajları incelenmiştir. Farklı şekilde ölçülmüş veri paketlerine ait modellerin karşılaştırması yapılarak hangi parametrelerin ve hangi yöntemin daha başarılı olduğu incelenmiştir. Son olarak elde edilen bulguların tartışılması avantajlı ve dezavantajlı olduğu durumların belirlenmesi ve yorumlanması ile konumsal olarak SR değerlerinin nasıl değişebileceğinin makine öğrenmesi yöntemleri kullanılarak tahmin edilmesi hedeflenmiştir.

2. Materyal ve Method

Bu çalışmada Türkiye'nin 3600 grid noktası için SR tahmini yapılacaktır. Bu bölümde çalışma alanı ve temin edilen verilere ait bilgiler, makine öğrenmesi yöntemleri ve enterpolasyon yöntemine ait bilgiler verilecektir.

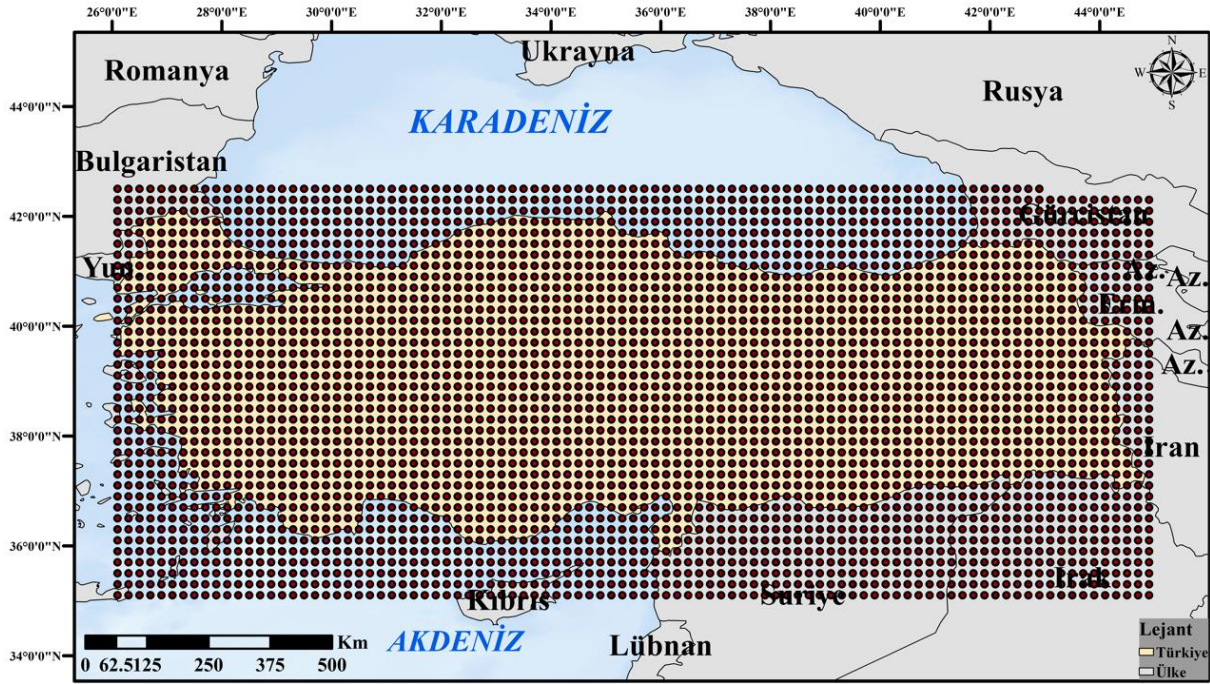
2.1. Çalışma Alanı

26°-45° doğu boylamları ve 36°-42° kuzey enlemleri arasında yer alan Türkiye, tipik Akdeniz iklimine sahip olmasından dolayı her mevsim güneş alan ve güneş enerjisi potansiyeli oldukça yüksek olan bir ülkedir (Sözen ve Arcaklıoğlu, 2005). Çalışma alanı olan

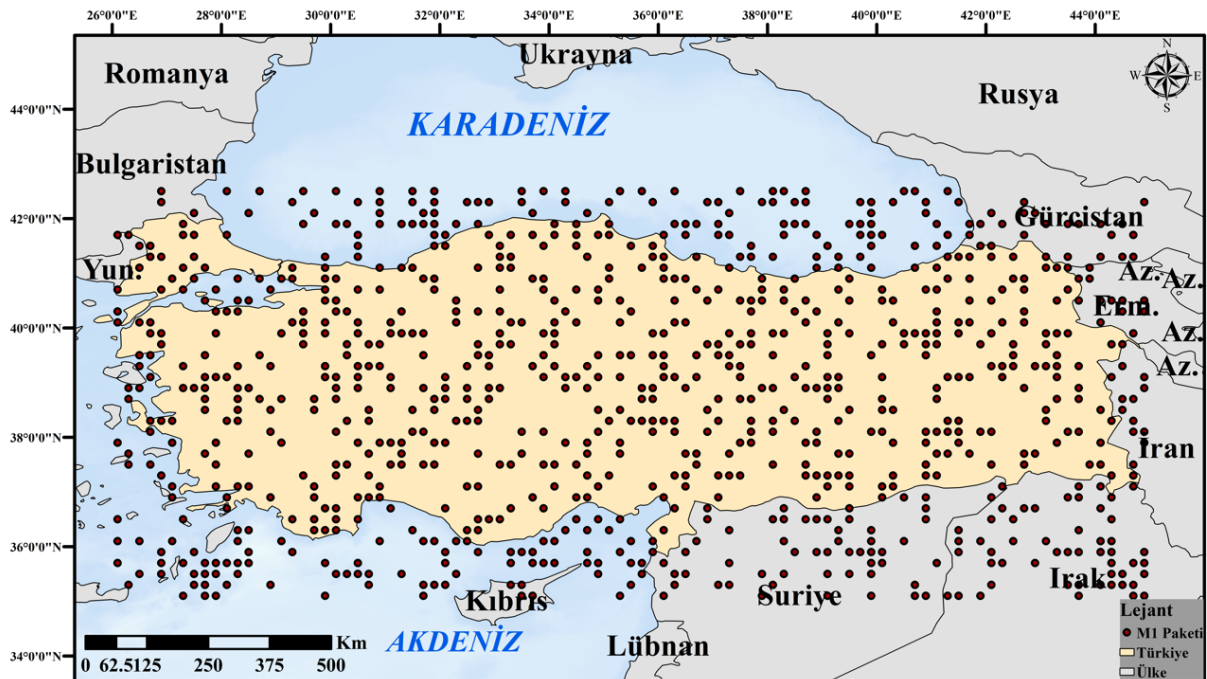
Türkiye'de yer alan 3600 grid noktasına ait noktasal konumlar Şekil 1'de verilmiştir. Bu grid noktalarına ait ortalama yıllık SR verileri MGM'den temin edilmiştir (Meteoroloji Genel Müdürlüğü, 2022).

Şekil 1'de görülen çalışma alanı 0,2°x0,2° çözünürlükte 3600 gridden (ızgara) oluşmaktadır. Çalışmada 3600 grid noktasının 900 gridden oluşan ve rastgele seçilerek hazırlanan M1, M2, M3 ve M4 test modellerine ait haritalar Şekil 2-5'te verilmiştir.

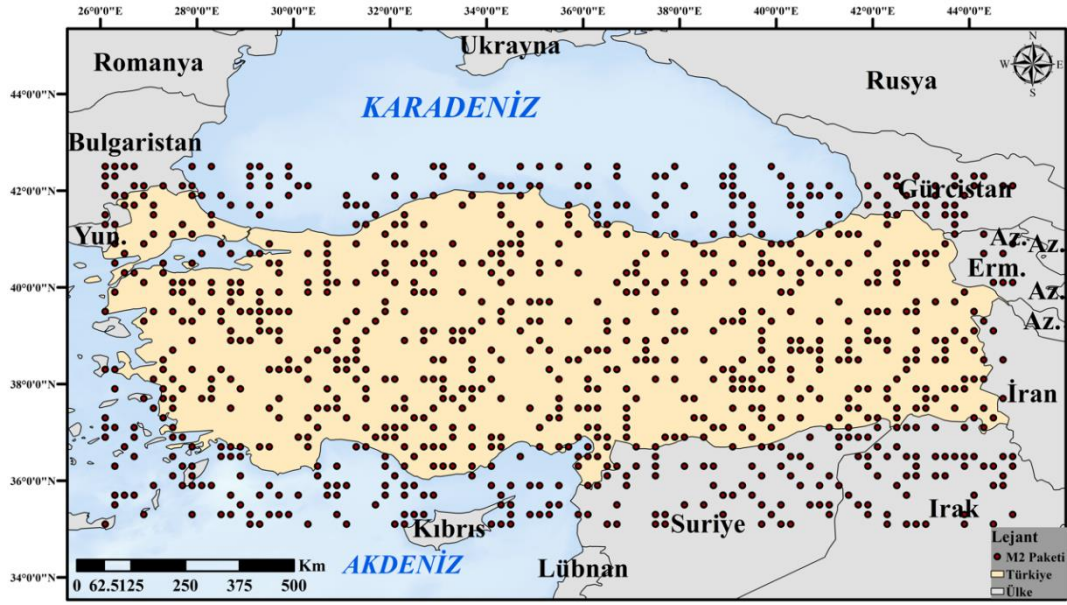
Şekil 2-5'te M1, M2, M3 ve M4 paketlerine ait gridlerin konumlarının gösterildiği haritalar görülmektedir. Gridler rastgele seçilerek hazırlanmıştır ve tüm paketlerin çalışma alanına homojen dağıldığı görülmektedir.



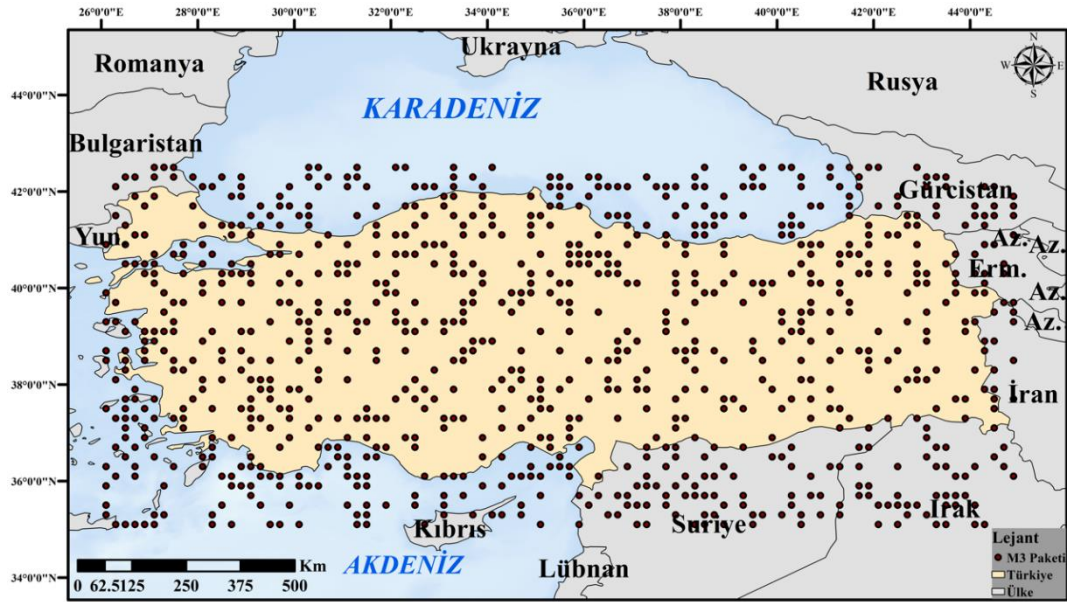
Şekil 1. Türkiye'nin 3600 grid noktasından oluşan çalışma alanı.



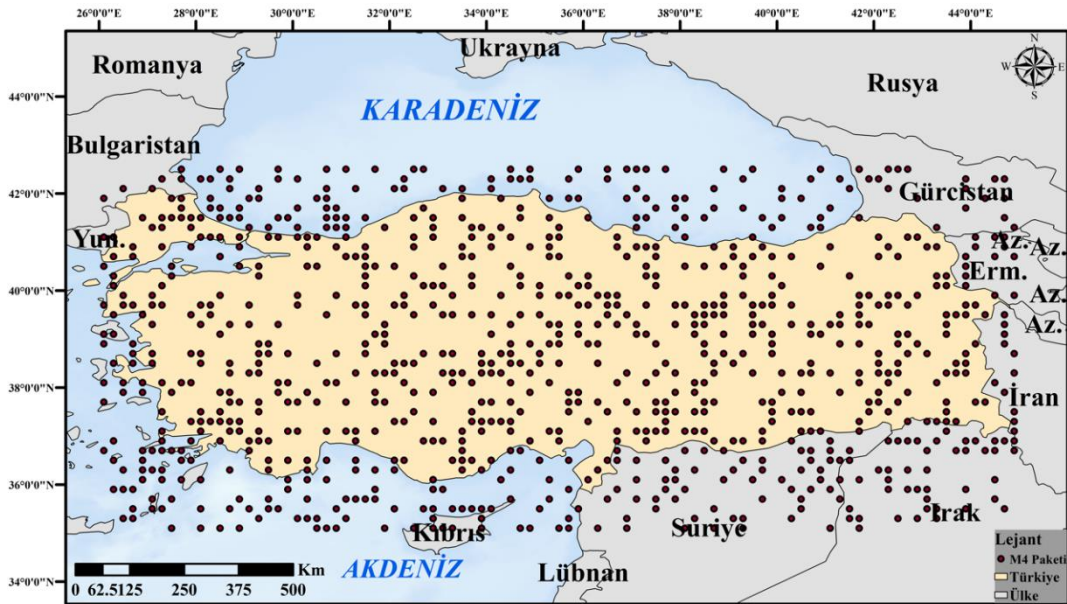
Şekil 2. Türkiye'nin M1 modeline ait çalışma alanı.



Şekil 3. Türkiye'nin M2 modeline ait çalışma alanı.



Şekil 4. Türkiye'nin M3 modeline ait çalışma alanı.



Şekil 5. Türkiye'nin M4 modeline ait çalışma alanı.

2.2. Veri

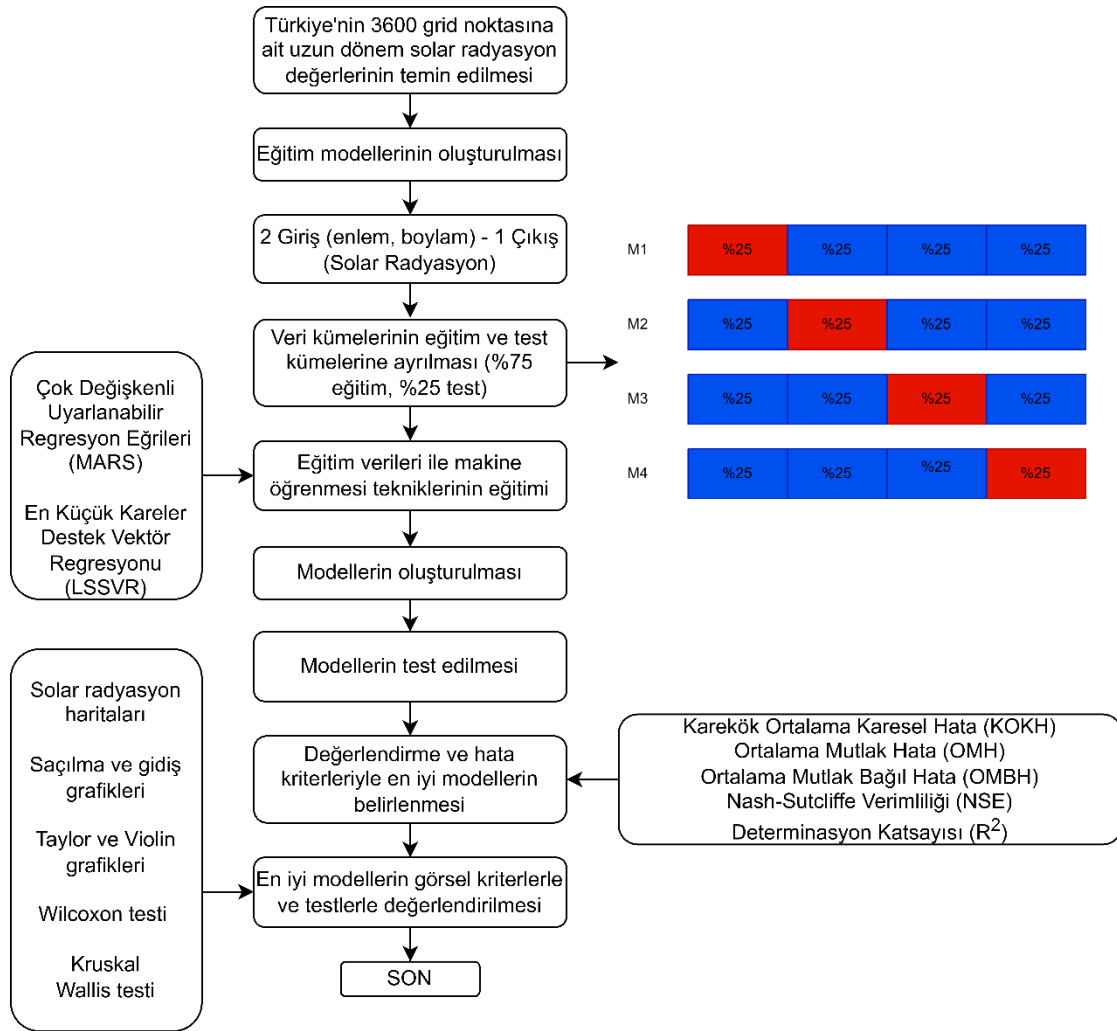
Bu çalışmada enlem ve boylam parametrelerinin tercih edilmesinin sebebi sadece coğrafi parametreler ile yapılan tahmin çalışmalarının başarılı olup olmayacağını görmek ve literatüre bu parametrelerle yapılan SR tahmin sonuçları kazandırılarak katkıda bulunmak istenmesidir. Çalışmada kullanılacak olan verilere ait tanımlayıcı istatistik bilgileri Tablo 1’de verilmiştir.

Tablo 1 incelendiğinde standart hata değeri incelendiğinde 3600 grid noktasına ait değerlerin ana kütleyle yakın olduğu görülmektedir. Standart sapma ve standart hata değerleri incelendiğinde 3600 grid noktasına ait değerlerin iyi olduğu görülmektedir. 3600 grid noktasına ait veri setinde çarpıklık katsayısı $-1 < -0.0028 < +1$ aralığında olduğu için normal dağılımdan sapmadığı ve sola çarpık olduğu görülmektedir. Basıklık katsayısı sıfırdan küçük olduğu için yassı basık olduğu görülmektedir. Ancak 3600 grid sayısına ait veri setinin

basıklık katsayısı sıfıra çok yakın olduğu için kusursuz eğriye çok yakındır. Genel olarak basıklık ve çarpıklık olmadığı görülmektedir. Çalışmaya ait iş akış şeması Şekil 6’da verilmiştir.

Tablo 1. 3600 gride noktasına ait 2004-2021 uzun dönem yıllık ortalama solar radyasyon değerlerinin (kWsa/m^2) tanımlayıcı istatistik bilgileri.

Parametre	Değer
Ortalama	4.5459
Standart Hata	0.0075
Ortanca	4.5673
Standart Sapma	0.4470
Basıklık	-0.496
Çarpıklık	-0.003
Varyans	0.1998
En büyük	5.6422
En küçük	3.2977
Veri sayısı	3600



Şekil 6. İş akış şeması.

2.3. Çok değişkenli uyarlanabilir regresyon eğrileri (MARS) yöntemi

Çok değişkenli uyarlanabilir regresyon eğrileri (multivariate adaptive regression splines) (MARS) yöntemi 1991 yılında Jerome Friedman tarafından geliştirilmiştir. Bu yöntemin amacı bağımsız değişkenlerin bağımlı değişkene olan etkisini

çözümlemekte kullanmaktır (Friedman, 1991). Bu yöntemi kullanabilmek için isminden de anlaşılacağı üzere birden fazla değişken olması gerekir. Bu yöntemin çalışma prensibi bağımlı ve bağımsız değişkenler arasında regresyon analizi yaparak ortaya çıkan modeller arasından optimum modeli seçmek üzerinedir (Yüksel, 2016). Bu yöntem neticesinde elde edilen eşitlikler doğrusal regresyon olmadığı için bağımlı ve

bağımsız değişkenler arasındaki ilişkinin daha doğru tespit edilebilmesi için düzleştirme eğrileri kullanılmaktadır (Yüksel, 2016). MARS, sürekli doğrusal olmayan (sezgisel) sayısal çıktıları tahmin etmek için kullanılan bir sezgisel regresyon modelidir. MARS algoritmasının iki unsuru vardır: ileri ve geri adımlar. İleri adım yöntemi, ilgili girdi değişkenlerinin bir koleksiyonunu seçmek için kullanılır (De Andrés ve ark., 2011). Geri adım yöntemini kullanarak önceden seçilmiş koleksiyondaki gereksiz değişkenleri kaldırır. Aşağıdaki temel denklem, X değişkeninden (giriş) Y değişkenine (çıkış) bir fonksiyon çizmek için kullanılır. Yeni Y değerleri, giriş aralığındaki sapma noktasında tanımlanan iki temel fonksiyon veya her iki değişken değeri kullanılarak elde edilir (Sharda ve ark., 2006). MARS modelinin formülasyonu Eşitlik 1’de verilmiştir (Demir ve Yaseen, 2023).

$$Y = B_0 + \sum_{n=2}^k a_n B_n(X_t) + \varepsilon \quad (1)$$

Bu formülde bağımlı değişken Y ile ifade edilmektedir. B_0 model sabit terimini, k temel fonksiyon adedini, a_n n’inci temel fonksiyonun katsayısını, $B_n(X_t)$ t’inci bağımsız değişken için n’inci temel fonksiyonu, X bağımsız değişkeni ve ε hata terimini göstermektedir (Friedman, 1991).

2.4. En küçük kareler destek vektör regresyonu (LSSVR) yöntemi

1995 yılında Vladimir Vapnik ve arkadaşları, AT&T Bell Laboratuvarlarında giriş parametreleri ile çıkış parametreleri arasındaki lineer olmayan ilintiyi minimum hatayla hesap etmek için tasarlanan en küçük kare destek vektör makine modellerini (LSSVR) geliştirdiler (Cortes ve Vapnik, 1995; Suykens, 2001; Smola ve Schölkopf, 2004). LSSVR, SVR’den (destek vektör regresyonu) türetilen, gerçek hayatta karşılaşılan problemleri regresyon, fonksiyon tahmini ve sınıflandırma kombinasyonu ile çözmek için uygulanabilecek mükemmel uygunlukta bir yöntemdir. 1999 yılında Suykens ve Vandewalle tarafından LSSVR modeli projelendirilerek (Suykens ve Vandewalle, 1999), karmaşık zaman serisi tahminine uygulandı. LSSVR ve SVR arasındaki temel fark, denklemlerin dikkate alınmasıdır; eğitim aşamasında, LSSVR doğrusal denklemler kullanırken, SVR ikinci dereceden optimizasyonda kullanır. Geri yayımlı sinir ağları (BPNN), kısmi en küçük kare regresyon (PLS) ve çok değişkenli doğrusal regresyon (MLR) gibi diğer geleneksel modeller, LSSVR’den hesaplama açısından daha kapsamlıdır. Dolayısıyla bu modeli uygulamak diğerlerine göre daha kolaydır. Mevcut zaman serisindeki veriyi önceki zaman serilerindeki verilerle istatistiksel olarak karşılaştırarak giriş (X) ve çıkış (Y) arasındaki en iyi işlevi bulmak için kullanılır (Yaseen ve ark., 2016). Bu prosedürü, çok boyutlu bir özellik uzayı ve doğrusal olmayan bir ilişki fonksiyonu kullanarak gerçekleştirir. Regresyon fonksiyonu Eşitlik 2 ile ifade edilebilir (Demir ve Yaseen, 2023).

$$q(p) = \text{sign}[w^T p + b] \quad (2)$$

İki sınıfın verileri ayrılabilir olduğunda, Eşitlik 3’teki gibi ifade edilebilir.

Girdi verileri $p_k \in R_n$ ve çıktı verileri $q_k \in R$ olan belirli bir eğitim kümesi düşünülmelidir. R sınıf etiketleri $q_k \in \{-1, +1\}$ ve doğrusal sınıflandırıcı ile Eşitlik 2’teki gibi elde edilmektedir.

$$\begin{cases} w^T p_k + b \geq +1, & \text{eğer } q_k = +1 \\ w^T p_k + b \leq -1, & \text{eğer } q_k = -1 \end{cases} \quad (3)$$

2.5. IDW Enterpolasyon Yöntemi

Ters Mesafe Ağırlıklı Enterpolasyon (IDW) Metodu, en çok tercih edilen jeostatistik olmayan ve örnek noktalar tarafından örneklenemeyen noktalara veri oluşturmak için kullanılan bir enterpolasyon metodudur. Yalnızca komşu noktalardan tahminler ürettiği için, yerel bir ara değer tahmini yapar (Yılmaz ve ark., 2023; Doğan ve ark., 2023). Bu verilerin oluşturulması, çeşitli noktalarla komşuluklar değerlendirilerek uygulanan enterpozisyonel mesafeye ve formüle bağlıdır (Taylan ve Damçayiri, 2016; Demirgöl ve ark., 2023). Yönteme ait genel Eşitlikler 4-5’de yer almaktadır.

$$f(x, y) = \sum_{i=1}^n w_i f_i \quad (4)$$

$$w_i = \frac{h_i^{-p}}{\sum_{j=1}^n h_j^{-p}} \quad (5)$$

Burada; w_i ağırlıkları temsil eder ve değerlerinin toplamı Eşitlik 5’teki gibi olmalıdır. f_i bilinen yükseklik değerini temsil eder, p bir güç parametresi olarak alınır ve üs ile gösterilir, h_i örnek noktalar ile enterpolasyon noktaları arasındaki uzamsal mesafeyi temsil eder (Krige, 1951; Shepard, 1968).

3. Bulgular

Farklı veri setlerinin farklı kombinasyonlarının farklı yöntemlerin tahminlerinde çalışılmasıyla beraber ortaya çıkan tahminlerin birbirleriyle karşılaştırılması sonucunda avantaj ve dezavantajlı yönlerinin incelenmesi elzemdir. Bu çalışmada kriterleri belirlerken literatürde en yaygın kullanılan 5 kriter tercih edilmiştir. Hidrolojik ve meteorolojik tahminler değerlendirilirken gözlenmiş değer ile modelde tahmin edilen değer arasındaki hata ölçütlerinin kullanımına sıklıkla rastlanılmaktadır (Yaseen, 2021). Literatürde yaygın olarak kullanılan karşılaştırma kriterlerinden Karekök Ortalama Karesel Hata (KOKH), Ortalama Mutlak Hata (OMH), Ortalama Mutlak Bağıl Hata (OMBH), determinasyon katsayısı (R^2)’dir ve bu çalışmada da benzer karşılaştırma kriterleri kullanılmıştır.

Karşılaştırma kriterlerinin Eşitlik 6-9'da gösterilmektedir. Eşitlik 6-9'da kullanılan N veri sayısı, SR_p tahmin edilen solar radyasyon değeri, SR_o gözlenen solar radyasyon değeridir. Denklem 10'da ise G_i gözlenen verileri, M_i modelin tahmin ettiği verileri, $\bar{G}$ gözlenen verilerin ortalamasını, n veri adedini göstermektedir. Değerlendirme kriterlerinin performans değerleri Tablo 2'de verilmiştir.

$$KOKH = \sqrt{\frac{1}{N} \sum_{i=1}^N (SR_p - SR_o)^2} \quad (6)$$

$$OMH = \frac{1}{N} \sum_{i=1}^N |SR_p - SR_o| \quad (7)$$

$$OMBH = \frac{1}{N} \left(\frac{\sum_{i=1}^N |SR_p - SR_o|}{SR_o} * 100 \right) \quad (8)$$

$$R^2 = \left(\frac{N * (\sum SR_o * SR_p) - (\sum SR_o) * (\sum SR_p)}{\sqrt{(N * \sum SR_o^2) - (\sum SR_o)^2} * \sqrt{(N * \sum SR_p^2) - (\sum SR_p)^2}} \right)^2 \quad (9)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (G_i - M_i)^2}{\sum_{i=1}^n (G_i - \bar{G})^2} \quad (10)$$

Tablo 2. Değerlendirme kriterleri performans değer aralıkları (Demirgöl, 2023).

Performans Kriteri	KOKH/OMH/OMBH	R ²	NSE
Mükemmel uyum	0	0.9 < R ² < 1	0.75 < NSE < 1
Çok iyi uyum	0 < KOKH < 0.05	0.75 < R ² < 0.9	0.65 < NSE < 0.75
İyi uyum	0.05 < KOKH < 0.1	0.5 < R ² < 0.75	0.5 < NSE < 0.65
Orta uyum	0.1 < KOKH < 0.2	0.25 < R ² < 0.5	0.25 < NSE < 0.5
Zayıf uyum	0.2 < KOKH < 0.5	0 < R ² < 0.25	0 < NSE < 0.25
Kabul edilemez uyum	KOKH > 0.5	0	0

Tablo 2'de verilen performans aralıklarından KOKH, OMH ve OMBH kriterleri tahmin değerlerimiz ile gözlenen değerlere yakınlığını belirlemek için kullandığımız kriterlerdir. Bu yüzden bu kriterler sıfıra ne kadar yaklaşırsa tahmin değerlerimiz ile gözlem değerlerimiz arasındaki fark o kadar azalmış olur. R² ve NSE için ise tam tersine veriler arasındaki uyum ne kadar fazlaysa 1 değerine o kadar yaklaştığı söylenebilir (Bayram ve Çitakoğlu, 2023). Bu çalışmada kullanılan veriler MGM'den temin edilen SR verileridir ve birimler kWsa/m² cinsindendir. Karşılaştırma kriterlerinden Karekök Ortalama Karesel Hata, Ortalama Mutlak Hata hata sonuçları olduğundan modeldeki hata sonuçlarının birimleri de aynı şekilde kWsa/m²'dir. Hata sonuçlarının yüksek olması modelin gerçek veriye göre uzakta yani kötü sonuç verdiğini gösterir. Düşük olması ise modelin gerçek veriye göre yakın, hatası düşük ve modelin uygun olduğunun göstergesidir. J. E. Nash ve J. V. Sutcliffe tarafından bulunan Nash-Sutcliffe verimliliği (NSE), oluşturulan modellerin tahmin yeteneğini test eden gösterge değerini verir (Nash ve Sutcliffe, 1970). NSE değeri -∞ ile 1 değerleri arasında değişebilir. Negatif bir değer aldığında modelin başarısız olduğu kabul edilir. 0 ile 1 arasındaki değerler kabul edilebilir değerlerdir. Eğer sonuç 1 veya 1'e çok yakın çıkarsa modelin tahmin ettiği değer ile gözlenen değer arasında çok iyi bir uyum olduğu anlaşılmaktadır (Aydın ve Düzen, 2011). Çalışmanın önemli noktalarından bir tanesi de eğitim ve test veri paketlerinin hazırlanması adıdır. Makine öğrenmesi ile tahmin çalışmalarında eğitilen veya test edilen verinin fazlalığı, homojenliği, eksiksizliği,

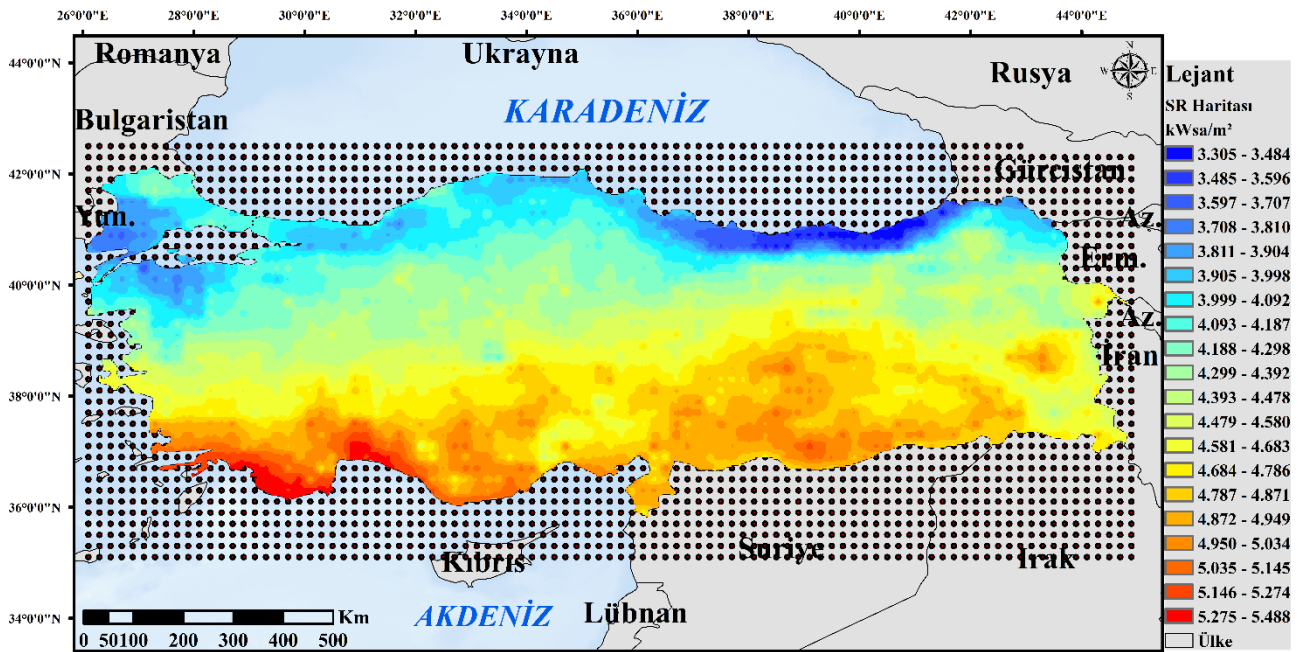
periyodikliği gibi etmenler doğrudan sonuca etki etmektedir (Alan ve Karabatak, 2020). Bu nedenle veri setlerinin/kombinasyonlarının hazırlanması son derece elzemdir. Bu çalışmada birçok farklı eğitim ve test veri paketi kullanılarak modellemeler gerçekleştirilmiştir.

3.1. 3600 Grid Noktası İçin Uzun Dönem SR Modelleme

MARS ve LSSVR makine öğrenmesi yöntemlerinde kullanılmak üzere 3600 grid noktasının 2004-2021 yıllarına ait yıllık ortalama SR değerleri %75 eğitim %25 test istasyonu olacak şekilde M1, M2, M3 ve M4 veri paketleri oluşturulmuştur. MGM'den elde edilen veri seti uzun dönem yıllık ortalama SR değerlerine ait olduğu için bu veri paketinde periyodiklik parametresi kullanılamamıştır. MGM tarafından HELIOSAT yöntemiyle elde edilen 2004-2021 uzun dönem yıllık ortalama solar radyasyon değerlerine ait SR haritası ise Şekil 7'de verilmiştir.

Şekil 7'de 3600 grid için 2004-2021 uzun dönem yıllık ortalama solar radyasyon değerleri ile IDW enterpolasyon tekniği kullanılarak hazırlanmış SR haritası görülmektedir. 3600 grid noktasına ait değerlerle oluşturulan veri paketlerine ait tanımlayıcı istatistik bilgileri Tablo 3'te verilmiştir.

2 giriş (enlem-boylam), 1 çıkış (SR tahmini) olacak şekilde iki farklı makine öğrenmesi yöntemi için modellemelerin yapıldığı toplamda 8 farklı model oluşturulmuş ve değerlendirme kriterleri Tablo 4-7'de verilmiştir.



Şekil 7. 3600 grid noktasının HELIOSAT yöntemiyle elde edilen değerlerine ait SR haritası.

Tablo 3. 3600 grid noktalarına ait tanımlayıcı istatistiki bilgiler.

3600 Grid Noktası	M1	M2	M3	M4
Ortalama	4.5288	4.5541	4.5243	4.5762
Standart Hata	0.0149	0.0152	0.0149	0.0146
Ortanca	4.5190	4.6056	4.5181	4.6128
Standart Sapma	0.4459	0.4550	0.4459	0.4392
Basıklık	-0.5357	-0.4303	-0.4855	-0.4849
Çarpıklık	0.0420	-0.1171	0.1238	-0.0595
Varyans	0.1988	0.2070	0.1988	0.1929
En büyük	5.6422	5.6209	5.6001	5.6245
En küçük	3.2977	3.3030	3.3002	3.3726
Veri sayısı	900	900	900	900

Tablo 4. 3600 grid MARS eğitim modellerine ait değerlendirme kriterleri.

Değerlendirme Kriterleri	M1	M2	M3	M4
KOKH	0,1943	0,1921	0,1930	0,1952
OMH	0,1567	0,1546	0,1552	0,1579
OMBH	3,4680	3,4179	3,4324	3,5028
NSE	0,8113	0,8129	0,8137	0,8112
R ²	0,8113	0,8129	0,8137	0,8112

Tablo 5. 3600 grid MARS test modellerine ait değerlendirme kriterleri

Değerlendirme Kriterleri	M1	M2	M3	M4
KOKH	0,1922	0,1992	0,1960	0,1912
OMH	0,1554	0,1616	0,1577	0,1528
OMBH	3,4464	3,5979	3,5043	3,3445
NSE	0,8142	0,8085	0,8070	0,8104
R ²	0,8142	0,8095	0,8084	0,8128

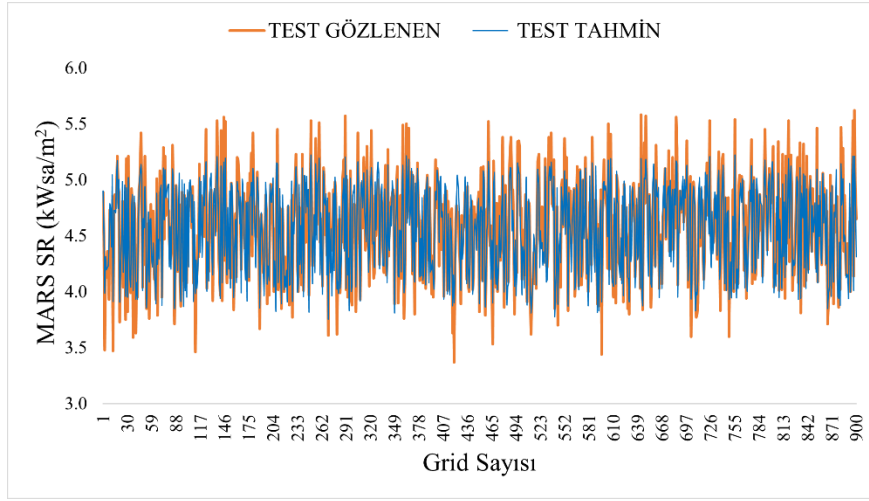
MARS makine öğrenmesi yöntemiyle MATLAB platformunda 2 giriş (enlem-boylam), 1 çıkış (SR değeri) olarak analiz edilen eğitim aşamasında en iyi sonucu M2 paketi verirken en kötü sonucu ise M4 veri paketinin verdiği görülmüştür. Ancak tüm veri paketleri birbirlerine çok yakın değerler verirken R^2 ve NSE kriterlerinin kötü olduğu görülmektedir.

MARS makine öğrenmesi yöntemiyle MATLAB platformunda 2 giriş (enlem-boylam), 1 çıkış (SR değeri) olarak analiz edilen test aşamasında en iyi sonucu M4 paketi verirken en kötü sonucu ise M2 paketinin verdiği görülmüştür. Ancak tüm veri paketlerinde R^2 ve NSE

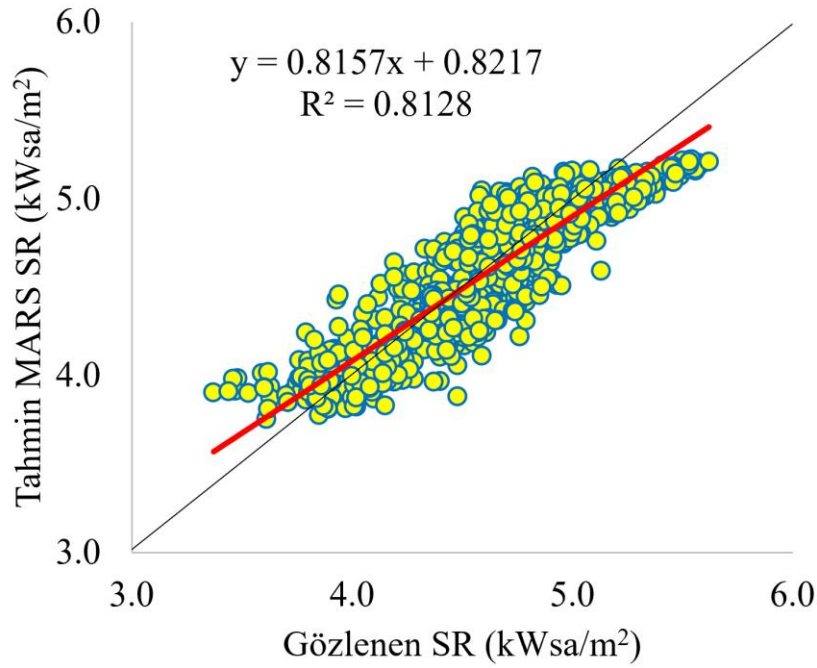
kriterlerinin kötü olduğu görülmektedir. 3600 grid noktası için MARS test modeline ait 900 grid noktası için gidiş ve saçılma grafiği **Şekil 8** ve **Şekil 9**'da verilmiştir.

Şekil 8’de görüldüğü gibi model genel olarak gözlenen değerlere yakın değerler tahmin edememiştir.

Şekil 9'da görüldüğü gibi model genel olarak doğrusal çizginin yakınında toplanmıştır. Ancak LSSVR yöntemine göre daha başarısız olduğu görülmektedir. MARS yöntemiyle en iyi sonucu veren M4 modelinin eğitim ve test tahmin değerleri ile ArcMap platformunda IDW enterpolasyon yöntemiyle hazırlanan SR haritası **Şekil 10'**da verilmiştir.



Şekil 8. 3600 grid M4 paketi MARS modeline ait gidiş grafiği.



Şekil 9. M4 paketi 900 grid noktasının MARS test modeline ait saçılma grafiği.

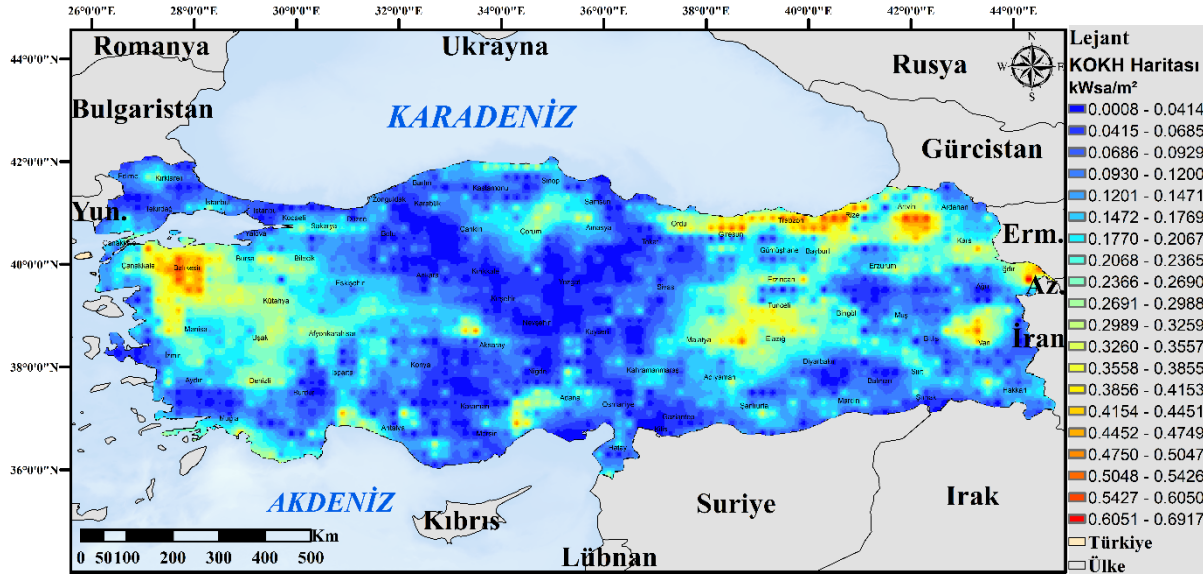


Şekil 10. 3600 grid M4 paketi MARS modeline ait SR haritası.

Şekil 10'da Türkiye'nin 3600 gridinin 2004-2021 uzun dönem yıllık ortalama SR değerlerinin M4 paketi ile eğitilen MARS yönteminin tahmin ettiği SR değerlerinin IDW enterpolasyon yöntemiyle ArcMap ortamında oluşturulmuş SR haritası görülmektedir. Haritanın Şekil 7'deki gözlem değerleri ile oluşturulmuş SR haritasına benzerlik gösterdiği ancak LSSVR kadar başarılı olamadığı ve çalışma alanında paraleller oluşturarak tahminlerde bulunduğu yani nispeten başarısız tahminler yaptığı görülmektedir. Bu başarı oranının daha net görülebilmesi için karekök ortalama karesel hata değerleri ile bir harita daha oluşturulmuştur. M4 paketi

MARS yöntemi ile tahmin edilen eğitim ve test verilerine ait KOKH haritası Şekil 11'de verilmiştir.

Şekil 11'de M4 paketi MARS yöntemi ile elde edilen SR tahmin değerlerinin gözlenen değerlerle KOKH değerlerine ait KOKH haritası verilmiştir. KOKH için başarı ölçütü sıfıra olan yakınlığıdır. Harita incelendiğinde Türkiye'nin İç Anadolu ve Güneydoğu Anadolu bölgelerinin genelinde homojen ve başarılı tahminler yapıldığı görülmektedir. Ancak MARS yönteminde KOKH değerlerinin yüksek olduğu noktaların bölgeler oluşturduğu görülmektedir.



Şekil 11. 3600 grid M4 paketi MARS modeline ait KOKH haritası.

Tablo 6. 3600 grid LSSVR eğitim modellerine ait değerlendirme kriterleri.

Değerlendirme Kriterleri	M1	M2	M3	M4
KOKH	0,0794	0,1072	0,1091	0,1101
OMH	0,1420	0,0784	0,0801	0,0802
OMBH	1,7932	1,7705	1,8069	1,8165
NSE	0,9420	0,9417	0,9405	0,9400
R ²	0,9420	0,9417	0,9405	0,9400

Tablo 7. 3600 grid LSSVR test modellerine ait değerlendirme kriterleri.

Değerlendirme Kriterleri	M1	M2	M3	M4
KOKH	0,1137	0,1154	0,1096	0,1065
OMH	0,0821	0,0844	0,0815	0,0790
OMBH	1,8571	1,9100	1,8469	1,7668
NSE	0,9350	0,9357	0,9396	0,9411
R ²	0,9350	0,9359	0,9399	0,9414

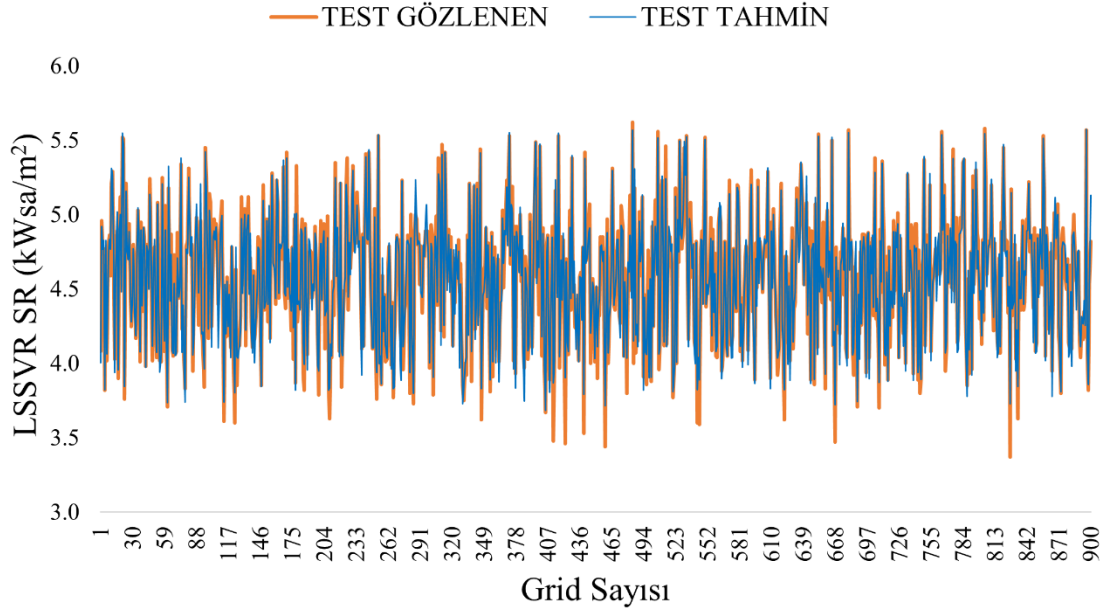
LSSVR makine öğrenmesi yöntemiyle MATLAB platformunda 2 giriş (enlem-boylam), 1 çıkış (SR değeri) olarak analiz edilen eğitim aşamasında en iyi sonucu M1 paketi, en kötü sonucu ise M4 veri paketinin verdiği gözlenmiştir. Ancak tüm veri paketlerinde birbirine yakın değerler olduğu görülmektedir.

LSSVR makine öğrenmesi yöntemiyle MATLAB platformunda 2 giriş (enlem-boylam), 1 çıkış (SR değeri) olarak analiz edilen eğitim aşamasında en iyi sonucu M4 paketi, en kötü sonucu ise M2 veri paketinin verdiği gözlenmiştir. Ancak tüm veri paketlerinde birbirine yakın değerler olduğu görülmektedir. 3600 grid noktası için LSSVR test modeline ait 900 grid noktası için gidiş ve saçılma grafiği Şekil 12 ve Şekil 13'te verilmiştir.

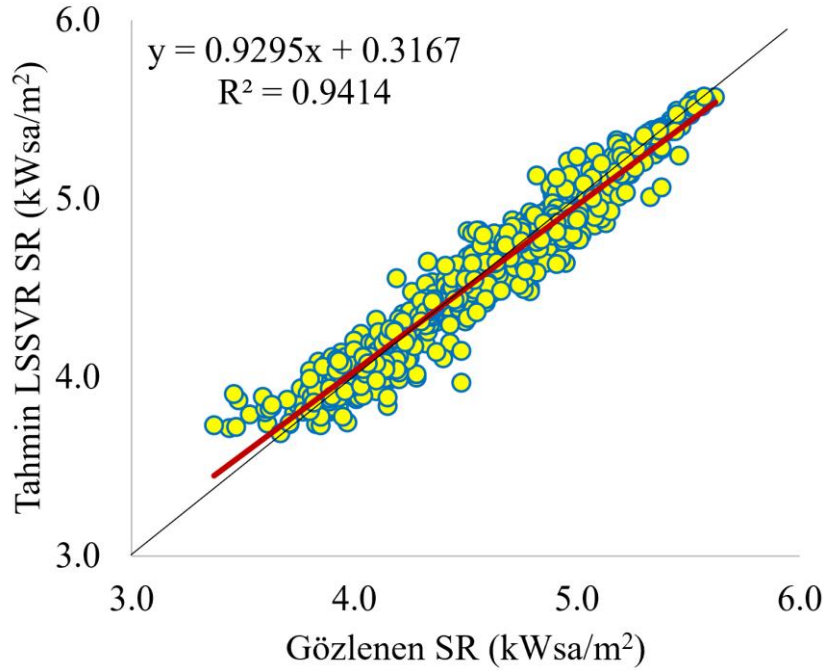
Şekil 12'de görüldüğü gibi model genel olarak gözlenen değerlere yakın değerler tahmin etmiştir.

Şekil 13'te görüldüğü gibi model genel olarak doğrusal çizginin yakınında toplanmıştır. LSSVR yöntemiyle en iyi sonucu veren M4 modelinin eğitim ve test tahmin değerleri ile ArcMap platformunda IDW enterpolasyon yöntemiyle hazırlanan SR haritası Şekil 14'te verilmiştir.

Şekil 13'te görüldüğü gibi model genel olarak doğrusal çizginin yakınında toplanmıştır. LSSVR yöntemiyle en iyi sonucu veren M4 modelinin eğitim ve test tahmin değerleri ile ArcMap platformunda IDW enterpolasyon yöntemiyle hazırlanan SR haritası Şekil 14'te verilmiştir.



Şekil 12. 3600 grid M4 paketi LSSVR modeline ait gidiş grafiği.



Şekil 13. M4 paketi 900 grid noktasının LSSVR test modeline ait saçılma grafiği.

Şekil 14'te Türkiye'nin 3600 gridinin 2004-2021 uzun dönem yıllık ortalama SR değerlerinin M4 paketi ile eğitilen LSSVR yönteminin tahmin ettiği SR değerlerinin IDW enterpolasyon yöntemiyle ArcMap ortamında oluşturulmuş SR haritası görülmektedir. Haritanın Şekil 7'deki gözlem değerleri ile oluşturulmuş SR haritasına çok benzediği yani oldukça başarılı tahminler yaptığı görülmektedir. Bu başarı oranının daha net görülebilmesi için karekök ortalama karesel hata değerleri ile bir harita daha oluşturulmuştur. M4 paketi LSSVR yöntemi ile tahmin edilen eğitim ve test verilerine ait KOKH haritası Şekil 15'te verilmiştir.

Şekil 15'te M4 paketi LSSVR yöntemi ile elde edilen SR tahmin değerlerinin gözlenen değerlerle KOKH değerlerine ait KOKH haritası verilmiştir. KOKH için

başarı ölçütü sıfıra olan yakınlığıdır. Harita incelendiğinde Türkiye genelinde homojen ve çok başarılı tahminler yapıldığı görülmektedir. 3600 gride ait tahminlerin Doğu Karadeniz'in uç kısımları ile Akdeniz'in iki gridinde hata oranının yüksek olduğu görülmektedir. 3600 gride ait gözlem verileri ile MARS ve LSSVR tahminlerinin Taylor ve Violin grafikleri Şekil 16 ve Şekil 17'de verilmiştir.

Şekil 16'da Taylor diyagramının tam ve yakınlaştırılmış kesitine ait görsel verilmiştir. Bu diyagramda gözlem değerine en yakın, karekök ortalama karesel sapma değerinin en düşük ve korelasyon katsayısının en büyük olduğu yöntem en başarılı yöntem olarak kabul edilmektedir. Bu görsele göre en başarılı sonucu LSSVR yönteminin verdiği görülmektedir. Ancak

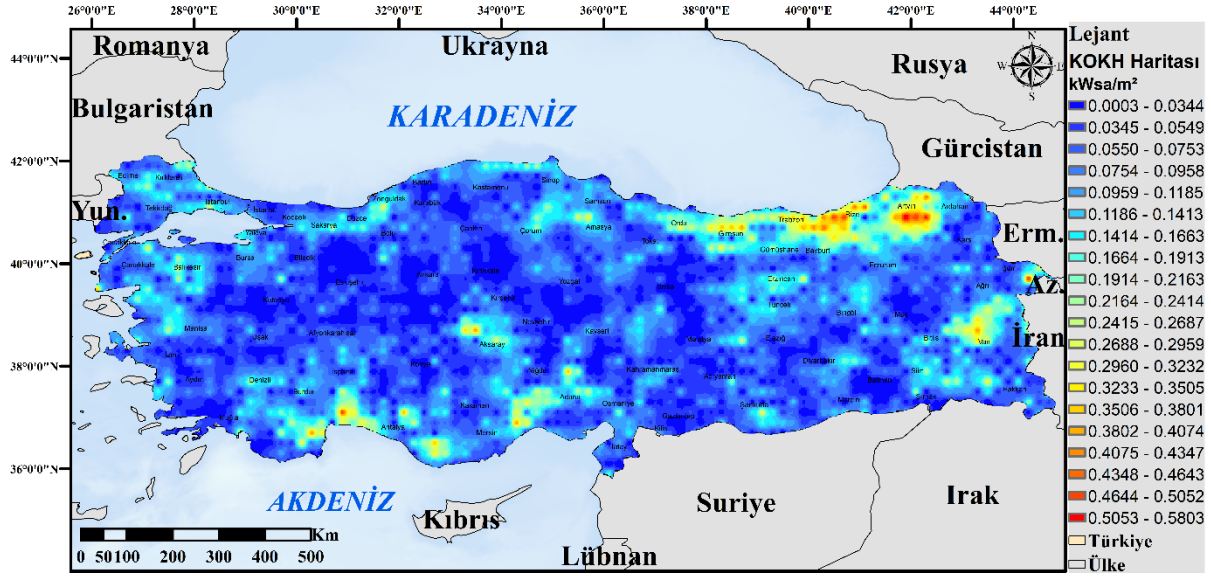
MARS yöntemi ile elde edilen değerlerin LSSVR yöntemine göre başarısız olduğu görülmektedir.

Violin diyagramı verilerin dağılımını ve olasılık yoğunluğunu görsellere bakarak değerlendirme

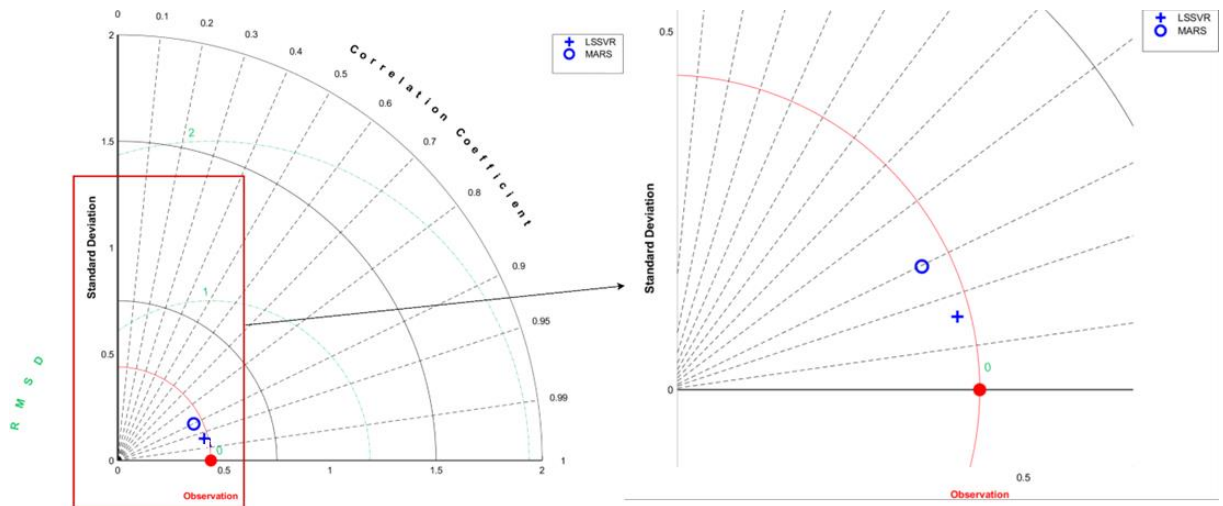
yapmamız içindir. LSSVR yöntemini ile gözlem değerlerinin birbirine oldukça benzediği görülmektedir. Ancak MARS yönteminin LSSVR ve gözlem değerleri ile nispeten ayrıştığı görülmektedir.



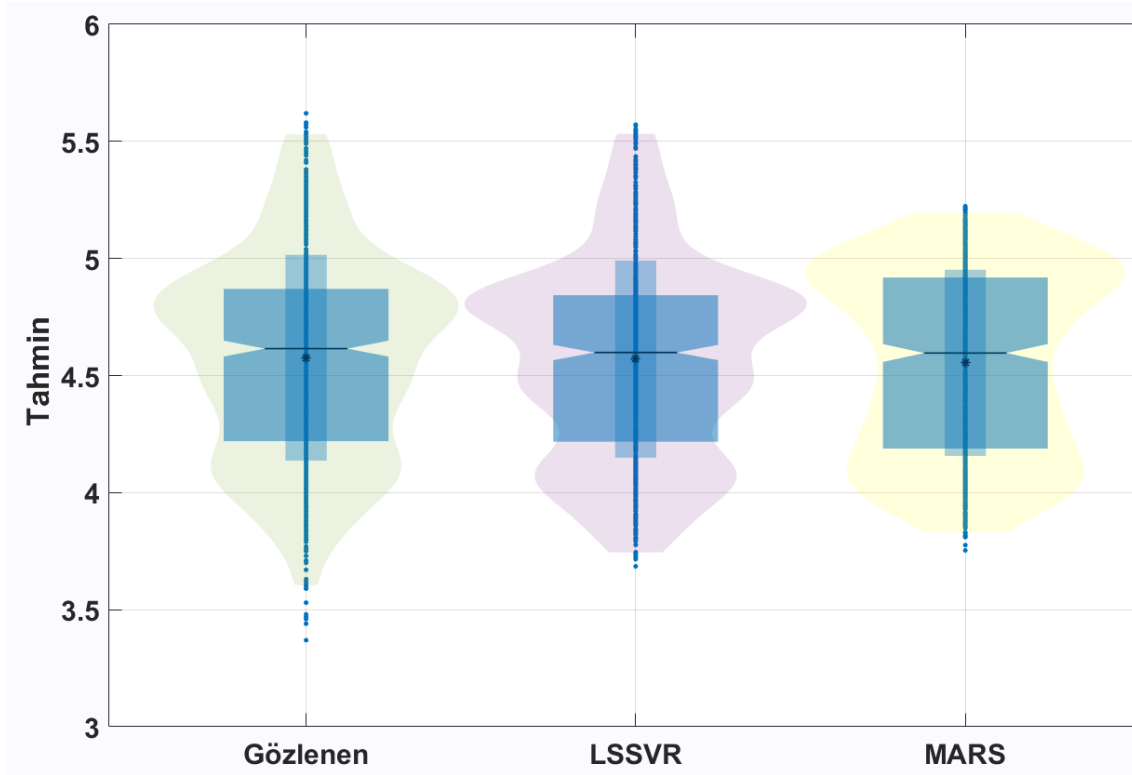
Şekil 14. 3600 grid M4 paketi LSSVR modeline ait SR haritası.



Şekil 15. 3600 grid M4 paketi LSSVR modeline ait KOKH haritası.



Şekil 16. 3600 grid gözlem/tahmin paketlerine ait Taylor diyagramı.



Şekil 17. 3600 grid gözlem/tahmin paketlerine ait Violin diyagramı.

3600 grid noktasına ait veri seti modellerine Wilcoxon testi ve Kruskal Wallis testi yapılarak değerlendirilmiştir.

Tek kuyruklu Wilcoxon işaretli sıralar testi parametrik olmayan veri setlerinde bir modelin diğer modellerden daha doğru tahmin yapıp yapmadığını belirlemekte kullanılan bir test yöntemidir. Eğer iki model aynı tahmin hatasına sahipse ($e_1 = e_2$) sıfır hipotezi (H_0) ile değerlendirilmektedir. Birinci model ikinci modelden daha küçük bir hataya sahipse ($e_1 < e_2$) ise alternatif hipotez (H_1) kullanılır (Uncuoglu ve ark., 2022). Bu çalışmada %95 anlamlılık düzeyi için Wilcoxon işaretli sıralar testi uygulanmıştır. Eğer p-value 0.05'ten büyük ise H_0 değeri geçerlidir. Eğer p-value 0.05'ten küçük ise H_0 değeri %95 anlamlılık düzeyinde güvenle reddedilir.

Wilcoxon testinin sonuçları incelendiğinde 3600 grid noktasına ait veriler kullanılarak MARS yöntemi ile yapılan tahminlerin p-value değeri 0,001 ve LSSVR yöntemi ile yapılan tahminlerin p-value değeri 0,059 olduğu görülmektedir. Modeller içinde MARS yöntemi 0,05'ten küçük olduğu için anlamlı düzeyinde güvenle reddedildiği görülmektedir. Fakat bu sonuç LSSVR için tahminlerle gözlenen değerlerin benzer popülasyondan geldiğini veya gözlenen değerlerin tahmin edilen değerlerle uyumlu olduğunu (temsil edildiğini) göstermektedir.

3600 grid noktasına ait verilerle tahmin edilen 2 modele istatistiksel anlamlılık testi olan Kruskal Wallis testi yapılmıştır. Kruskal Wallis testine ait sonuçlar Tablo 8'de verilmiştir.

Tablo 8. Kruskal-Wallis testi.

Modeller	3600 Grid Noktası	
	p-value	H_0
MARS	0.8470	Ret
LSSVR	0.5947	Ret

Tablo 8'de verilen Kruskal-Wallis testi %95 anlamlılık düzeyinde gerçekleştirilmiştir (Uncuoglu ve ark., 2022). Tablo 8'de görüldüğü üzere yöntemlerin tamamında gözlem değeri ile aralarında anlamlı farklılık bulunamamıştır.

4. Tartışma ve Sonuç

Bu çalışmada, MGM'den temin edilen Türkiye'nin 3600 grid noktasına ait 2004-2021 yıllarını kapsayan uzun dönem yıllık ortalama SR değerleri için önce rastgele karıştırma yapıp daha sonra 4 eşit parçaya bölünmüş ve her biri %25 test, %75 eğitim seti olmak üzere M1, M2, M3 ve M4 paketleri oluşturulmuştur. Karşılaştırma kriteri olarak KOKH, OMH, OMBH, R^2 ve NSE kullanılmıştır. Ayrıca yöntemlerin yapmış olduğu tahmin verileri ile gözlem verilerinin gidiş ve saçılma grafikleri ile Taylor ve Violin diyagramları oluşturulmuştur. Çalışmada, günümüz veya geleceğe ait SR değerlerinin tahmin edilmesinin yanı sıra yıllık ortalama SR verilerine göre farklı yöntem ve algoritmalarla gözlem verisi olmayan istasyonların değerlerinin tahmini de amaçlanmıştır.

Bu çalışmanın yaptığı tahminlerin karşılaştırması yapılması için literatür taraması yapıldığında 3600 grid noktasıyla yapılan akademik çalışmalar için literatür taranmış fakat bir çalışma bulunamamıştır. Ancak farklı veri setleri ile Türkiye'de yapılan Solar radyasyon tahminlerine ait veriler incelendiğinde; Sözen ve Arcaklıoğlu (2005), Türkiye'deki 12 istasyona ait ortalama günlük SR değerleri tahmini için yapay sinir ağları (YSA) yöntemi ile üç farklı algoritma (Scaled konjugat gradyan, Levenberg-Marquardt öğrenme algoritmaları ve Pola-Ribiere eşlenik gradyan) ve lojistik sigmoid transfer fonksiyonunu kullanmış ve giriş parametresinde 6 giriş (enlem, boylam, yükseklik,

periyodiklik, ortalama güneşlenme süresi ve ortalama sıcaklık) çıkış parametresinde 1 çıkış (SR tahmini) kullanarak yapmış oldukları bu çalışmada 12 istasyon içinden en başarılı sonuç veren istasyonun Artvin istasyonu ($R^2=0,999$) olduğu, en başarısız sonuç veren istasyonun Çanakkale istasyonu ($R^2=0,998$) olduğunu belirtmiştir. Özgören ve ark. (2012), Türkiye'nin herhangi bir noktasına ait aylık ortalama günlük global SR değerini tahmin etmek için çoklu non-lineer regresyon (MNL) metodu ile bir YSA modeli oluşturarak Türkiye'nin 31 istasyonundan elde edilen parametreler ile (1 giriş 1 çıkışla başlayıp 10 giriş 1 çıkışlı kombinasyona kadar) oluşturdukları 10 model içinde en başarılı tahmini veren modelin R^2 değerinin 0,9872, en başarısız sonucu veren modelin ise R^2 değerinin 0,9771 olduğu ifade etmişlerdir. Kuncan ve Şahin (2017), Türkiye'deki aylık ortalama günlük SR tahmini gerçekleştirdikleri çalışmalarında 53 istasyon için tercih ettikleri yapay sinir ağıları yönteminde 20 farklı model elde etmişlerdir. Çalışmada 5 giriş parametresi (SR ay değerleri, yeryüzü sıcaklığı, yükseklik, enlem ve boylam) ve 1 çıkış parametresi (SR tahmin) kullanan Kuncan ve Şahin 20 model içinden en iyi sonucu veren modelin KOKH değerinin 1,550 (MJ/m^2), en başarısız modelin ise KOKH değerinin 1,782 (MJ/m^2) olduğunu ifade etmişlerdir. Çitakoğlu ve Demir (2021), Türkiye'nin aylık ortalama SR değerlerini üç değişik derin öğrenme algoritması kullanarak [uyarlanabilir moment tahmini (ADAM), momentumlu stokastik gradyan inişi (SGDM) ve kök ortalama kare yayılımı (RMSProp)] 4 giriş (enlem, boylam, yükseklik ve periyodiklik) ile 1 çıkış (SR tahmini) parametresi uygulamışlardır. 81 il istasyonuna ait SR tahminlerinde en iyi sonucu veren istasyonun OMH değerinin 0,17 ($\text{MJ}/\text{m}^2\text{gün}$), KOKH değerinin 0,228 (MJ/m^2) ve R^2 değerinin 0,987, en başarısız sonucu veren istasyonun ise OMH değerinin 0,174 ($\text{MJ}/\text{m}^2\text{gün}$), KOKH değerinin 0,238 (MJ/m^2) ve R^2 değerinin 0,985 olduğunu ifade etmişlerdir. Ağbulut ve ark. (2021), Türkiye'de farklı SR ortalamalarına sahip dört il istasyonunun (Kırklareli, Tokat, Nevşehir ve Karaman) günlük küresel SR değerlerini tahmin etmek için 4 farklı makine öğrenme yöntemi [destek vektör makinesi, yapay sinir ağıları, k en yakın komşu (k-NN) ve derin öğrenme] kullanılmış olup, çalışmada 5 giriş parametresi (maksimum sıcaklık, minimum sıcaklık, bulut örtüsü, atmosfer dışı SR, gün uzunluğu) kullanarak 1 çıkış parametresi (SR tahmini) elde etmişlerdir. 4 farklı makine öğrenmesi yönteminden en iyi sonucu veren modelin KOKH değerinin 2,157 (MJ/m^2) ve R^2 değerinin 0,936, en başarısız sonucu veren modelin KOKH değerinin 2,776 ve R^2 değerinin 0,883 olduğunu belirlemişlerdir. Demirel ve ark. (2023), Türkiye'nin 81 ili için yapmış oldukları SR tahmini çalışmalarında M5-tree kullanılarak elde edilen modellerde en başarılı sonuçlar; KOKH= 0.3604, OMH= 0.1451, OMBH= 3.6029 ve $R^2= 0.9879$ olarak elde edildiği belirtilmiştir. Böylece M5-tree yönteminin literatürde yer alan yöntemlere alternatif bir yöntem olabileceğini belirtmişlerdir.

Bu çalışmada, elde edilen sonuçlar incelendiğinde; MARS yönteminde en iyi sonuç M4 paketinde (KOKH=0,1912, OMH=0,1528, OMBH=3,3445, NSE=0,8104, $R^2= 0,8128$), en kötü sonuç ise M2 paketinde (KOKH=0,1992, OMH=0,1616, OMBH=3,5979,

NSE=0,8085, $R^2=0,8095$) görülmüştür. LSSVR yönteminde ise en iyi sonuç M4 paketinde (KOKH=0,1065, OMH=0,0790, OMBH=1,7668, NSE=0,9411, $R^2=0,9414$), en kötü sonuç ise M2 paketinde (KOKH=0,1154, OMH=0,0844, OMBH=1,9100, NSE=0,9357, $R^2= 0,9359$) görülmüştür. Genel olarak 3600 gride ait 2004-2021 uzun dönem yıllık ortalama SR değerleri ile yapılan tahminlerde LSSVR yönteminin çok başarılı olduğu ancak MARS yönteminin daha başarısız olduğu görülmektedir. Bunun sebebinin MARS yönteminin daha az sezgisel olması olduğu düşünülebilir.

Ayrıca 3600 gride ait modellemelerde periyodiklik değişkeni olmadığı halde nokta sayısının yani veri sayısının çok daha fazla olmasının yarattığı avantaj bu eksikliği başarılı bir şekilde kapatamamıştır. Makine öğrenmesi yöntemleri ile SR tahmini yapmak için veri sayısının ve değişken sayısının artırılmasıyla çok daha başarılı sonuçlar vereceği söylenebilir. Tüm bu çalışma ışığında Türkiye'nin eksikliğini büyük ölçüde hissettiği SR yersel ölçüm istasyonlarının yerini makine öğrenmesi yöntemlerinin rahatlıkla doldurabileceği söylenebilir. Çalışmanın sonucunda solar radyasyonun makine öğrenmesi yöntemiyle oldukça başarılı bir şekilde tahmin edilebildiği gözlenmiş olup, makine öğrenmesi yöntemlerinin meteorolojide ve enerji sektöründe kullanımının yaygınlaşmasının güneş enerjisinden daha etkili bir şekilde yararlanılmasına neden olacağı düşünülmektedir. Bu sonuçlar literatürle karşılaştırıldığında literatürde kullanılan yöntemlerle uyumlu olduğu görülmüştür.

Bilgilendirme/Teşekkür

Bu çalışma Taha Demirel'in "Farklı Makine Öğrenmesi Yaklaşımlarını Kullanarak Türkiye'de HELIOSAT Tabanlı Güneş Radyasyonunun Modellenmesi" başlıklı yüksek lisans tez çalışmasından hazırlanmıştır.

Yazarlar desteklerinden dolayı KTO Karatay Üniversitesine teşekkür etmektedir. Bu çalışma, KTO Karatay Üniversitesi BAP Komisyonu tarafından (Proje no:10042305) desteklenmiştir

Araştırmacıların katkı oranı

Taha Demirel: Literatür taraması, Arazi çalışması, Modelleme, Makale yazımı; **Vahdettin Demir:** Kurgu, Düzenleme, **Mehmet Faik Sevimli:** Son kontrol. **Çalışma Beyanı**

Herhangi bir çıkar çatışması bulunmamaktadır.

Kaynakça

- Ağbulut, Ü., Gürel, A. E., & Biçen, Y. (2021). Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. *Renewable and Sustainable Energy Reviews*, 135, 110114. <https://doi.org/10.1016/j.rser.2020.110114>
- Ahmad, M. J., & Tiwari, G. N. (2011). Solar radiation models—A review. *International Journal of Energy Research*, 35(4), 271-290.

- <https://doi.org/10.1002/er.1690>
- Alan, A., & Karabatak, M. (2020). Veri seti-sınıflandırma ilişkisinde performansa etki eden faktörlerin değerlendirilmesi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 32(2), 531-540.
<https://doi.org/10.35234/fumbd.738007>
- Alizamir, M., Kim, S., Kisi, O., & Zounemat-Kermani, M. (2020). A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: Case studies of the USA and Turkey regions. *Energy*, 197, 117239.
<https://doi.org/10.1016/j.energy.2020.117239>
- Aydın, H., & Düzen, H. (2011). Comparison of measured and calculated evaporation values of open water surface: The case of Lake Van (Turkey). 5th Atmospheric Science Symposium, 275-284.
- Bayram, S., & Çitakoğlu, H. (2023). Modeling monthly reference evapotranspiration process in Turkey: application of machine learning methods. *Environmental Monitoring and Assessment*, 195(1), 67.
<https://doi.org/10.1007/s10661-022-10662-z>
- Bhardwaj, S., Sharma, V., Srivastava, S., Sastry, O. S., Bandyopadhyay, B., Chandel, S. S., & Gupta, J. R. P. (2013). Estimation of solar radiation using a combination of Hidden Markov Model and generalized Fuzzy model. *Solar Energy*, 93, 43-54.
<https://doi.org/10.1016/j.solener.2013.03.020>
- Bora, E., Tekeli, M., & Etöz, M. (2015). Menemen Ovasında Referans Bitki Su İhtiyacı ve Ölçülen Solar Radyasyonun Zamansal Değişimi. *Anadolu Ege Tarımsal Araştırma Enstitüsü Dergisi*, 25(2), 65-76.
- Citakoglu, H. (2015). Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation. *Computers and Electronics in Agriculture*, 118, 28-37.
<https://doi.org/10.1016/j.compag.2015.08.020>
- Cornejo-Bueno, L., Casanova-Mateo, C., Sanz-Justo, J., & Salcedo-Sanz, S. (2019). Machine learning regressors for solar radiation estimation from satellite data. *Solar Energy*, 183, 768-775.
<https://doi.org/10.1016/j.solener.2019.03.079>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297.
<https://doi.org/10.1007/BF00994018>
- Çitakoğlu, H., & Demir, V. (2021). Solar Radyasyonun Derin Öğrenme Kullanılarak Tahmin Edilmesi: Türkiye Örneği. *International World Energy Conference*, 167-175.
- De Andrés, J., Lorca, P., de Cos Juez, F. J., & Sánchez-Lasheras, F. (2011). Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). *Expert Systems with Applications*, 38(3), 1866-1875.
<https://doi.org/10.1016/j.eswa.2010.07.117>
- Demir, V., & Yaseen, Z. M. (2023). Neurocomputing intelligence models for lakes water level forecasting: a comprehensive review. *Neural Computing and Applications*, 35(1), 303-343.
<https://doi.org/10.1007/s00521-022-07699-z>
- Demirgöl T. (2023). Farklı makine öğrenmesi yaklaşımlarını kullanarak Türkiye'de HELİOSAT tabanlı güneş radyasyonunun modellenmesi. [Yüksek Lisans Tezi. KTO Karatay Üniversitesi].
- Demirgöl, T., Demir, V., & Sevimli, M. F. (2023). Model-Ağacı (M5-tree) yaklaşımı ile HELIOSAT tabanlı güneş radyasyonu tahmini. *Geomatik*, 8(2), 124-135.
<https://doi.org/10.29128/geomatik.1137687>
- Doğan, A., Başeğmez, M., & Aydın, C. C. (2023). Geniş çalışma alanlarında jeofizik ve jeoteknik ölçümlerin yerine geçebilecek vekil özelliklerin CBS ile belirlenmesi. *Geomatik*, 8(3), 293-305.
<https://doi.org/10.29128/geomatik.1161434>
- Enerji İşleri Genel Müdürlüğü (2023). <https://enerji.gov.tr/eigm-yenilenebilir-enerji-kaynaklar-gunes>
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 19(1), 1-67.
<https://doi.org/10.1214/aos/1176347963>
- Guermoui, M., Melgani, F., Gairaa, K., & Mekhalfi, M. L. (2020). A comprehensive review of hybrid models for solar radiation forecasting. *Journal of Cleaner Production*, 258, 120357.
<https://doi.org/10.1016/j.jclepro.2020.120357>
- Gül, M., & Çelik, E. (2017). ANFIS kullanılarak Tunceli ili için global güneş radyasyonu tahmini. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 8(4), 891-899.
- Güzel, B., Seveli, O., & Okatan, E. (2023). Predicting Solar Radiation Based on Meteorological Data Using Machine Learning Techniques: A Case Study of Isparta. *International Journal of Engineering Research and Development*, 15(2), 704-713.
<https://doi.org/10.29137/umagd.1268055>
- Jiang, Y. (2009). Computation of monthly mean daily global solar radiation in China using artificial neural networks and comparison with other empirical models. *Energy*, 34(9), 1276-1283.
<https://doi.org/10.1016/j.energy.2009.05.009>
- Kılıç, B., & Kumaş, K. (2019). Burdur ili güneşlenme değerlerinin yapay sinir ağları metodu ile tahmini. *Teknik Bilimler Dergisi*, 6(1), 38-44.
- Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6), 119-139.
- Kuncan, F., & Şahin, M. (2017, September). Yapay sinir ağı ve uydu dataları kullanılarak güneş radyasyonunun tahmini. In *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 1-10.
<https://doi.org/10.1109/IDAP.2017.8090288>
- Meteoroloji Genel Müdürlüğü. (2022). https://mgm.gov.tr/kurumci/radyasyon_iller.aspx
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, 10(3), 282-290.
[https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Ozgoren, M., Bilgili, M., & Sahin, B. (2012). Estimation of global solar radiation using ANN over Turkey. *Expert systems with applications*, 39(5), 5043-5051.
<https://doi.org/10.1016/j.eswa.2011.11.036>
- Sharda, V. N., Patel, R. M., Prasher, S. O., Ojasvi, P. R., & Prakash, C. (2006). Modeling runoff from middle Himalayan watersheds employing artificial intelligence techniques. *Agricultural Water*

- Management, 83(3), 233-242. <https://doi.org/10.1016/j.agwat.2006.01.003>
- Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. Proceedings of the 1968 23rd ACM national conference, 517-524. <https://doi.org/10.1145/800186.810616>
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and computing, 14, 199-222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- Sozen, A., & Arcaklioglu, E. (2005). Solar potential in Turkey. Applied Energy, 1(80), 35-45. <https://doi.org/10.1016/j.apenergy.2004.02.003>
- Suykens, J. A. (2001). Support vector machines: a nonlinear modelling and control perspective. European Journal of Control, 7(2-3), 311-327. <https://doi.org/10.3166/ejc.7.311-327>
- Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural processing letters, 9, 293-300. <https://doi.org/10.1023/A:1018628609742>
- Şeker, M. (2021). Yapay Sinir Ağı (YSA) Kullanılarak Meteorolojik Verilere Dayalı Solar Radyasyon Tahmini. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 23(69), 923-935. <https://doi.org/10.21205/deufmd.2021236920>
- Taylan, E. D., & Damçayırı, D. (2016). Isparta bölgesi yağış değerlerinin IDW ve Kriging enterpolasyon yöntemleri ile tahmini. Teknik Dergi, 27(3), 7551-7559.
- Uncuoglu, E., Citakoglu, H., Latifoglu, L., Bayram, S., Laman, M., Ilkentapar, M., & Oner, A. A. (2022). Comparison of neural network, Gaussian regression, support vector machine, long short-term memory, multi-gene genetic programming, and M5 Trees methods for solving civil engineering problems. Applied Soft Computing, 129, 109623. <https://doi.org/10.1016/j.asoc.2022.109623>
- Wang, L., Kisi, O., Zounemat-Kermani, M., Zhu, Z., Gong, W., Niu, Z., ... & Liu, Z. (2017). Prediction of solar radiation in China using different adaptive neuro-fuzzy methods and M5 model tree. International Journal of Climatology, 37(3), 1141-1155. <https://doi.org/10.1002/joc.4762>
- Yaseen, Z. M. (2021). An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: Review, challenges and solutions. Chemosphere, 277, 130126. <https://doi.org/10.1016/j.chemosphere.2021.130126>
- Yaseen, Z. M., Kisi, O., & Demir, V. (2016). Enhancing long-term streamflow forecasting and predicting using periodicity data component: application of artificial intelligence. Water resources management, 30, 4125-4151. <https://doi.org/10.1007/s11269-016-1408-5>
- Yilmaz, C. B., Bodu, H., Yüce, E. S., Demir, V., & Sevimli, M. F. (2023). Türkiye'nin uzun dönem ortalama sıcaklık (°C) değerlerinin üç farklı enterpolasyon yöntemi ile tahmini. Geomatik, 8(1), 9-17. <https://doi.org/10.29128/geomatik.984310>
- Yüksel, S. (2016). Türkiye'de cari işlemler açığının belirleyicileri: Mars yöntemi ile bir inceleme. Bankacılar Dergisi, 96(27), 102-121.

