

UZAKTAN ALGILAMA VE COĞRAFİ BİLGİ SİSTEMLERİ ALANINDA WEKA YAZILIMI KULLANIMINA BİR ÖRNEK

Metin ERSİN¹, Bekir Yiğit YILDIZ², Hacı Mustafa KANDIRMAZ³, Ozan ŞENKAL⁴,

Nuri EMRAHOĞLU⁵

¹Lisansüstü Öğrencisi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Anabilim Dalı, 01330 Balcalı Kampüsü, Adana, mersin@cu.edu.tr

²Doç. Dr., Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Anabilim Dalı, 01330 Balcalı Kampüsü, Adana,

³Prof. Dr., Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Anabilim Dalı, 01330 Balcalı Kampüsü, Adana,

⁴Prof. Dr., Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Anabilim Dalı, 01330 Balcalı Kampüsü, Adana,

⁵Doç. Dr., Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Anabilim Dalı, 01330 Balcalı Kampüsü, Adana.

ÖZET

Uzaktan algılama ve coğrafi bilgi sistemleri alanında araştırmacılar son yıllarda veri madenciliği ve makine öğrenmesi algoritmalarını daha çok kullanır olmuşlardır. Klasik istatistik ve analitik yöntemler yerini yapay zeka tabanlı yöntemlere bırakmıştır. Ancak programlama deneyimi az olan araştırmacıların veri analizinde bu yöntemleri hazır yazılım paketleri sayesinde kullanarak profesyonel sonuçlara ulaşabilmeleri mümkün olmuştur. Weka yazılımı bu türden araştırmalarda kullanılacak yazılımlara örnek teşkil etmektedir. Bu çalışmada Weka yazılımı ile güneş enerjisi tahmininin nasıl yapılacağı incelenmiş, Weka yazılımı tanıtılmıştır. Uzaktan algılama ve coğrafi bilgi sistemleri tabanlı uydu ve meteoroloji verileri analiz edilmiş, makine öğrenmesi algoritmalarından destek vektör regresyonu ile çok katmanlı algılayıcı yöntemleri, doğrusal regresyon tahmini sonuçları ile karşılaştırılmış ve % 89.2 lere varan tahmin doğruluğu elde edilmiştir.

Anahtar Sözcükler: Güneş enerjisi tahmini, uzaktan algılama, Weka yazılımı.

ABSTRACT

A SAMPLE OF USING WEKA SOFTWARE IN THE FIELD OF REMOTE SENSING AND GEOGRAPHICAL INFORMATION SYSTEMS

Researchers in the field of remote sensing and geographic information systems have used data mining and machine learning algorithms more in recent years. Classical statistical and analytical methods have been replaced by artificial intelligence-based methods. However, it has been possible for researchers with less programming experience to achieve professional results by using these methods in data analysis thanks to ready-made software packages. Weka software is an example of software that can be used in this type of research. In this study, how to make solar energy forecasting with Weka software is examined and introduced. Remote sensing and geographic information systems-based satellite and meteorology data were analyzed, support vector regression and multilayer sensor methods from machine learning algorithms were compared with linear regression prediction results, and a prediction accuracy of up to 89.2% was obtained.

Keywords: solar radiation forecasting, remote sensing, Weka software.

1.GİRİŞ

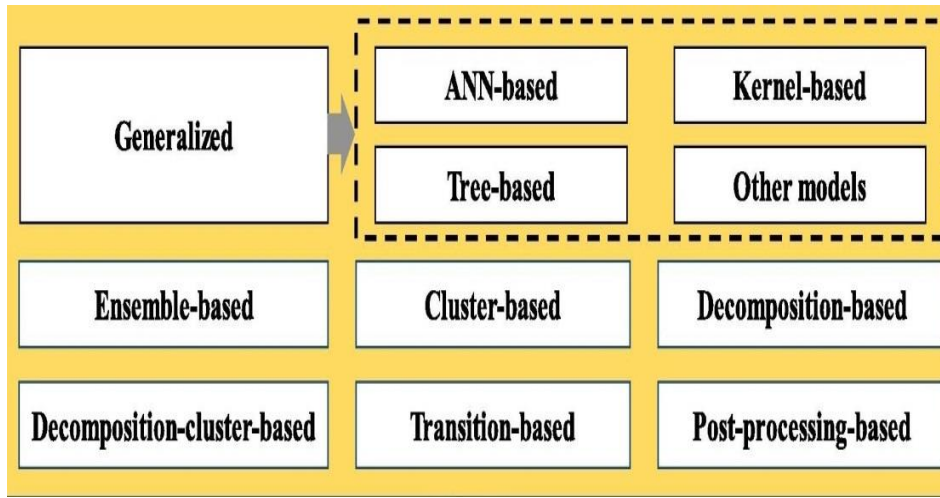
Sanayi devrimiyle birlikte yıllar içinde, fosil yakıt kullanımı artmış çok kısa sürede ciddi bir çevre kirlenmesi yaratmaya başlamıştır. Bu süreçte karbon emisyonu küresel ısınma oluşturacak boyutlara ulaşmış olup, bu durum insanlığı temiz enerji kaynakları aramaya yöneltmiştir. Başta güneş enerjisi (solar radiation) olmak üzere rüzgâr, dalga, hidroelektrik vb kaynakların nasıl kullanılabilineceği araştırılmıştır. Güneş enerjisinin üretimi, özellikle günlük döngüye ilişkin coğrafi faktörlere ve mevsimsel değişimlere bağlı olan üretim planlamasıyla ilgili önemli zorluklara sahiptir. Belirli bir bölgedeki kurulacak bir santral için üretim hacmi tahmini ve diğer üretim değişkenlerinin belirlenmesi ve stratejik planlanması, yatırım açısından büyük önem taşımaktadır. Çok sayıda ekonomik, sosyal, coğrafi ve mühendislikle ilgili değişken, bu kritik kararlar üzerinde etkilidir.

Günümüzde güneş enerjisinin elektrik üretiminde kullanılması ağırlıklı olarak fotovoltaik sistemler ile mümkün olmaktadır. Güneş enerjisi karasal yüzeye elektromanyetik radyasyon olarak gelir ve bu enerjinin ölçülmesi, belirli coğrafi alanlar için potansiyelinin tespiti spektrometre olarak bilinen aletler tarafından gerçekleştirilir. Bu tür aletler stratejik olarak karasal meteoroloji istasyonlarında konumlandırılmıştır. Her farklı konumda hem zamansal

hem de uzamsal deęişkenlerden bağımsız ölçümler yapmak doğası gereęi zordur. Finansal maliyetleri de ayrıca bir kısıt oluşturmaktadır.

Uzaktan algılama (remote sensing), nesnelere ile fiziki temas olmaksızın herhangi bir uzaklıktan yapılan ölçümler ile nesnelere hakkında bilgi edinme yöntemi şeklinde tanımlanmaktadır. (Yıldız 2009) Gözlem uydularındaki sensörler sayesinde atmosfer ile ilgili bir çok parametre hakkında bilgi edinmek mümkündür. Ölçülebilen parametrelerden biri de güneş radyasyonudur. Dünyayı etkileyen güneş radyasyonunun yarısı yansımaya tabidir ve bu radyasyonun %30'u uzaya geri yansır. Kalan %70, zeminin belirli koşullarına baęlı olarak emilir. Dünya'ya dik olarak çarpan ortalama güneş enerjisine güneş sabiti denir, yaklaşık 1368 W/m² deęerindedir. Bu deęer, yıllık döngü içinde yaklaşık % 0.1 dalgalanan deęişkenlik gösterir. Su ve kayanın diferansiyel absorpsiyon ve yansımaya özellikleri göz önüne alındığında, meteorologlar, yüzey sıcaklığı olarak bilinen gölgeli bir ortamda ekili yüzeyden 1,5 metre yükseklikte kaydedilen sıcaklığı kullanırlar (Ackerman ve Knox, 2015).

Literatürde güneş enerjisinin tahmini için çeşitli yöntemler, meteoroloji verileri ve de uydu verileri ile denenmiştir. Ülkemizde de yapılan araştırmalarda bu veri kaynakları ve algoritmalar çeşitli kombinasyonlar dahilinde kullanılmıştır. (Kaba ve ark. 2018; Karaman ve ark. 2021; Şenkal ve ark. 2009, 2010, 2018.) Son yıllarda klasik regresyon yaklaşımları yerine makine öğrenmesi algoritmaları kullanılmaktadır. Yaygın bazı yaklaşımlar Şekil.1'de verilmiştir.



Şekil 1. Güneş Enerjisi Tahmininde Kullanılan Makine Öğrenmesi Yöntemlerinin Sınıflandırması (Zhou ve ark., 2021).

Bu yaklaşımlardan yapay sinir aęları ile çekirdek tabanlı yaklaşımların araştırmalarda iyi sonuçlar verdiği görülmüştür. Ancak uygulamalarda seçilen programlama dilleri ya da hazır yazılım seçenekleri de araştırmacılar açısından önemlidir. Bu çalışmada seçilen algoritmaları çalıştırmak için veri madencilięi / makine öğrenmesi yazılımı olarak bilinen WEKA yazılımı tanıtılmış ve kullanılmıştır. Adana bölgesinde güneş enerjisi tahmini için uydu ve meteoroloji verileri hazırlanıp seçilen algoritmalar ile tahminler yapılmıştır.

2.MATERYAL ve METOD

2.1.Kullanılan Veriler ve Deęerlendirme Kriteri

Bu araştırmada, Adana bölgesine ait 2010'dan 2016'ya kadar olan dönemi kapsayan toplam yedi yıllık NOAA uydu serisinin AVHRR sensörünün 5 adet band verileri ki bunlardan band 1 görünür ışık, band 2,3 yakın kızıl ötesi, band 4,5 kızıl ötesi dalga boylarında algılama verileri (Yıldız, 2009) ile Adana bölgesine ait meteoroloji istasyonundan ilgili tarihlere ait yoğunlaşabilir su buharı, hava sıcaklığı ve yer yüzüne gelen güneş enerjisi ölçüm verileri alındı, uyumlu hale getirildi. Sobrino algoritması kullanılarak hesaplanan arazi yüzey sıcaklığı (LST) ve bandlardan T4/T5 fark deęerleri hesaplandı. Araştırma için uygun görülen çeşitli günlerden, 932 günlük uydu ve meteorolojik veriden oluşan kapsamlı bir veri kümesi çalışma için girdi ve çıktı parametreleri olarak belirlendi. Eğitim veri kümesi olarak 2010, 2011, 2012 ve 2013 yıllarına ait 596 günlük kayıtlı tam verileri içeren bir veri kümesi, 2014, 2015 ve 2016'daki 336 günlük dięer veri kümesi de test veri kümesi olarak kullanıldı. Verilerin dağılımı, varsayılan olarak Weka tarafından, eğitim için% 66 ve test için% 34'lük dağılıma sahip tekil bir veri kümesi belirlendi. Yıllık veriler sistematik olarak iki ayrı gruba ayrıldı; yılın sıralı gününü belirten Julian gün deęerinin ek bir veri alanı olarak dahil edilmesi ve böylece yıllık verilerin sürekli olarak iki kategoriye ayrılmasını zorunlu kılmıştır. Eğitim veri seti 4 yılı kapsayan verileri içerirken, test veri seti ise 3 yıllık verileri içermektedir.

Veri seti, uydu verisi, uydu verilerinden hesaplanmış veri, meteoroloji verisi ve Julien gün verisini içerdiği için makine öğrenmesi modellerini elde etmede en iyi veri kombinasyonunu sağlamağa yönelik literatürde henüz bir uzlaşma tespit edilemediğinden, elde edildiği kaynakla ilişkili olarak, verilerden çeşitli kombinasyonlar oluşturulmuştur. Oluşturulan 8 adet veri kombinasyonundan 1 adeti sadece meteoroloji, 1 adeti de sadece uydu verilerinden oluşturulmuştur.

Çalışmada tahmin edilen güneş enerjisi değerlerinin tahmin doğruluğu için Hata Karelerinin Ortalamasının Kare Kökü (RMSE) ile Korelasyon Katsayısı (R) ölçütleri baz alınmıştır. Weka işlemler sonunda, bu kriterlere ait değerleri otomatik rapor etmektedir.

Çizelge 1. Kullanılan Veri Setinin İçeriği

| Alan Adı | Veri Türü | İçerdiği Bilgi |
|----------|-----------|-----------------------------------------------------|
| JDATE | Nümerik | Yılın Kaçınıcı Günü Olduğu Verisi (Julian Günü) |
| BAND1 | Nümerik | Görünür Alan Yansıması (Uydu Verisi) |
| BAND2 | Nümerik | Yakın Kıızıl Ötesi Alan Işıması (Uydu Verisi) |
| BAND3 | Nümerik | Yakın Kıızıl Ötesi Alan Işıması (Uydu Verisi) |
| BAND4 | Nümerik | Kıızıl Ötesi Alan Işıması (Uydu Verisi) |
| BAND5 | Nümerik | Kıızıl Ötesi Alan Işıması (Uydu Verisi) |
| T4-T5 | Nümerik | Hesaplanan Değer |
| LST | Nümerik | Hesaplanan Değer |
| PWAT | Nümerik | Yoğuşabilir Su Buharı (Meteoroloji Verisi) |
| TEMP | Nümerik | Hava Sıcaklığı (Meteoroloji Verisi) |
| SOLR | Nümerik | Yeryüzüne Gelen Güneş Enerjisi (Meteoroloji Verisi) |

Çizelge 2. Veri Seti Kombinasyonları

| A Grubu | B Grubu | C Grubu | D Grubu | E Grubu | F Grubu | G Grubu | H Grubu |
|---------|---------|---------|---------|---------|---------|---------|---------|
| JDATE | JDATE | JDATE | JDATE | JDATE | JDATE | JDATE | JDATE |
| BAND1 | BAND1 | BAND1 | | BAND1 | BAND1 | | |
| BAND2 | BAND2 | BAND2 | | BAND2 | BAND2 | | |
| BAND3 | BAND3 | BAND3 | | BAND3 | | BAND3 | |
| BAND4 | BAND4 | BAND4 | | BAND4 | | BAND4 | |
| BAND5 | BAND5 | BAND5 | | BAND5 | | BAND5 | |
| T4-T5 | T4-T5 | | T4-T5 | | T4-T5 | | |
| LST | LST | | LST | | LST | | |
| PWAT | | | PWAT | PWAT | PWAT | PWAT | PWAT |
| TEMP | | | TEMP | TEMP | TEMP | TEMP | TEMP |
| SOLR | SOLR | SOLR | SOLR | SOLR | SOLR | SOLR | SOLR |

2.2.Kullanılan Algoritmalar

Makine öğrenmesini gerçekleştirebilecek birçok algoritma vardır. Bu algoritmalar genel olarak üç ana başlıkta toplanır. Bunlar; denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme algoritmalarıdır (Alpaydın, 2012). Denetimli öğrenme yapılırken makine öğrenmesi algoritması eşleştirilmiş verinin uygun bir fonksiyonla, ölçülmüş bir çıktı verisine bağlanması üzerinden bu fonksiyonu tahmin eder. Bu fonksiyon katsayı değerlerini test verisi üzerinde uygular, gerekirse düzeltme yapar. Veri setinde bulunduğu ağırlık katsayılarından çıkarım yapmaya çalışır ve uygun tahminler ile modeli üretir.

WEKA makine öğrenmesi yazılımında Yapay Sinir Ağları / Çok Katmanlı Algılayıcı (MLP), Destek Vektör Makinesi / Regresyonu (SVR) ve Doğrusal Regresyon (LR) algoritmaları kullanılarak Adana bölgesi için dünya yüzeyine gelen güneş radyasyon değerleri tahmin edilmeye çalışıldı. Bu algoritmalar kısaca şöyle özetlenebilir.

Yapay Sinir Ağları (ANN), doğrusal olmayan çözümler üretebilmesi için insan beyninin bilişsel işleme mekanizmalarını kopyalamaya çalışan matematiksel bir temsil görevi görür. MLP, genellikle yapay sinir hücreleri olarak adlandırılan birbirine bağlı nöronlardan oluşan birkaç katmandan oluşur. Tipik olarak, bu katmanlar sistematik olarak bir giriş katmanı, bir veya daha fazla ara gizli katman ve bir nihai çıktı katmanı şeklinde düzenlenir. Bu mimarinin çerçevesi, MLP'ye özgü doğrusal olmayan aktivasyon fonksiyonuna dayanır; burada

çoklu nöronlar, geri yayılma olarak bilinen bir öğrenme algoritması kullanan hiyerarşik bir şekilde birbirine bağlanır.

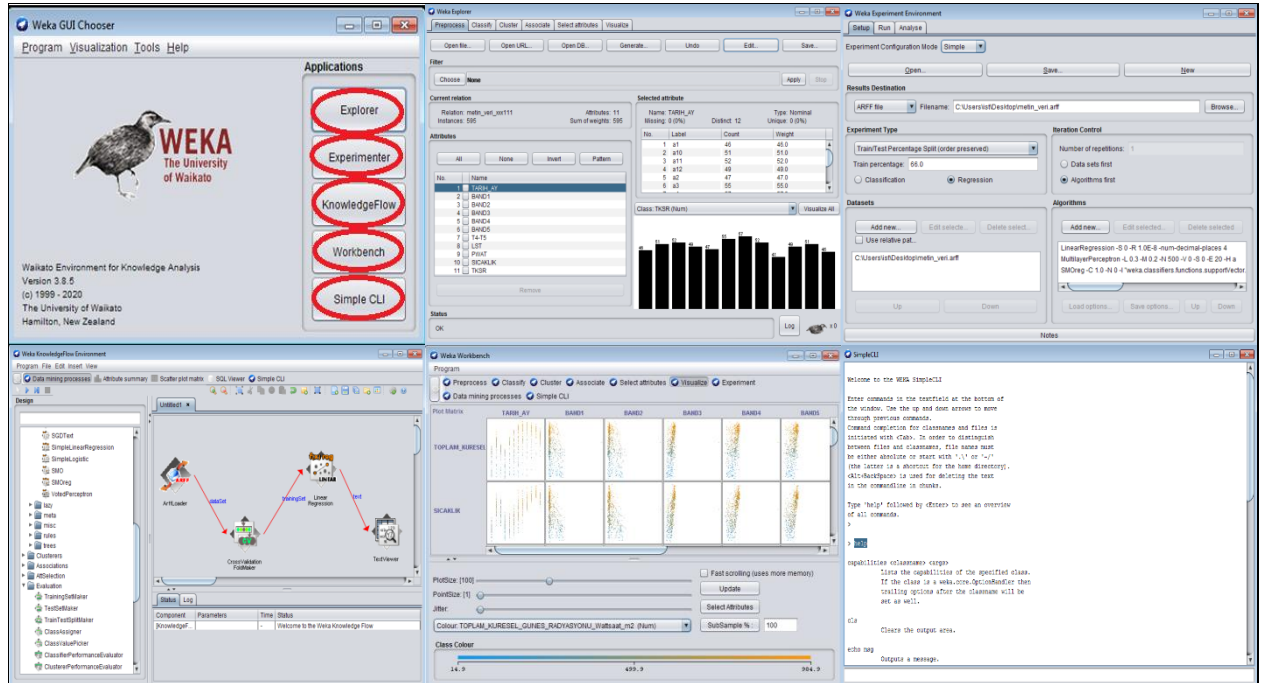
Destek Vektör Makinesi (Support Vector Machine SVM) yöntemi, (Cortes ve Vapnik, 1995). yapısal risk minimizasyonu prensibine dayanır. VM'nin amacı, veri örneklerinin mesafelerini sınıfları tanımlayan hiper düzleme dik olacak şekilde optimize etmek, böylece sınıflandırma işlemi sırasında farklı kategorilerin en yakın örneklerini tanımlamaktır. Hiper düzlem, veri kümesi genelinde etkinliğini korurken çeşitli konfigürasyonlarda tezahür edebilir. SVM, doğası gereği, ayırıcı hiper düzlemin her iki sınıftan eşit uzaklıkta olmasını sağlar ve her birine mümkün olan maksimum mesafeyi korur. Böylece, SVM hiper düzlemi, ilgili sınıflardan eşit bir maksimum ayrımda konumlandırılır.

Regresyon analizi, iki veya daha fazla değişken arasındaki doğrusal veya doğrusal olmayan ilişkileri ölçmek için kullanılan istatistiksel bir teknik oluşturur. Analiz tekil bir değişken kullanılarak gerçekleştirildiğinde, buna tek değişkenli regresyon analizi denir, çoklu değişkenlerin dahil edilmesi ise çok değişkenli regresyon analizi olarak adlandırılır.

2.3.WEKA Yazılımı ve Uygulama

WEKA, veri madenciliği ve analitik süreçler amacıyla kullanılan açık kaynaklı bir yazılım uygulamasıdır. 1993 yılında Yeni Zelanda'daki Waikato Üniversitesi'nden kaynaklanan adı, "Bilgi Analizi için Waikato Ortamı" ifadesinin ilk harflerini içeren adı, aynı anda Yeni Zelanda'ya özgü bir kuş faunası türüne atıfta bulunmakta ve logo resmi olarak kullanılmaktadır. Yazılım Java programlama dilinde oluşturulmuştur ve kodun uygulanmasıyla çalıştırılabilir. Yazılımın kullanıcı dostu yapısı, kodlama yeterliliğini gerektirmeden gelişmiş makine öğrenimi metodolojilerinin yayılmasını önemli ölçüde kolaylaştırmıştır. Harici olarak entegre edilebilen belirli algoritmalarla çok sayıda makine öğrenimi algoritmasını kapsar (Şeker, 2013).

Yazılımın resmi web portalına <http://www.cs.waikato.ac.nz/~ml/weka/> adresinden erişilebilir ve Linux, Mac OS ve Windows dahil olmak üzere çeşitli işletim sistemlerinde indirme ve kurulum için kullanılabilir. Ücretsiz kullanıma izin veren bir kamu lisansı altında yayılır. Yüklendikten sonra, kullanıcıya belirli bir yazılım arayüzü sunulur. Bu çalışma kapsamında yazılımın 3.8.5 sürümü kullanılmıştır.



Şekil 2. Soldan Sağa; Weka Ana Menu, Explorer, Experimenter, KnowledgeFlow, Workbench, Simple CLI

Ana ara yüzde Applications başlığı altında 5 düğme yer almaktadır. Her biri ayrı görevleri olan ara yüzlere erişim sağlamaktadır. Bunlardan en çok kullanılanı Explorer ara yüzüdür. Explorer ara yüzü; Veri ön işleme (Pre-Process), sınıflandırma (Classify), kümeleme / bölütleme (Cluster), ilişkilendirme (Associate), özellik / filtre seçim (Select attributes) ve görselleştirme (Visualize) işlemlerine erişim sağlamaktadır. Veri ön yükleme, veri girişi ve yapılandırma ayarları süreçlerini kapsar. En sık kullanılan arayüzü temsil eder. Sınıflandırma sekmesi, çeşitli

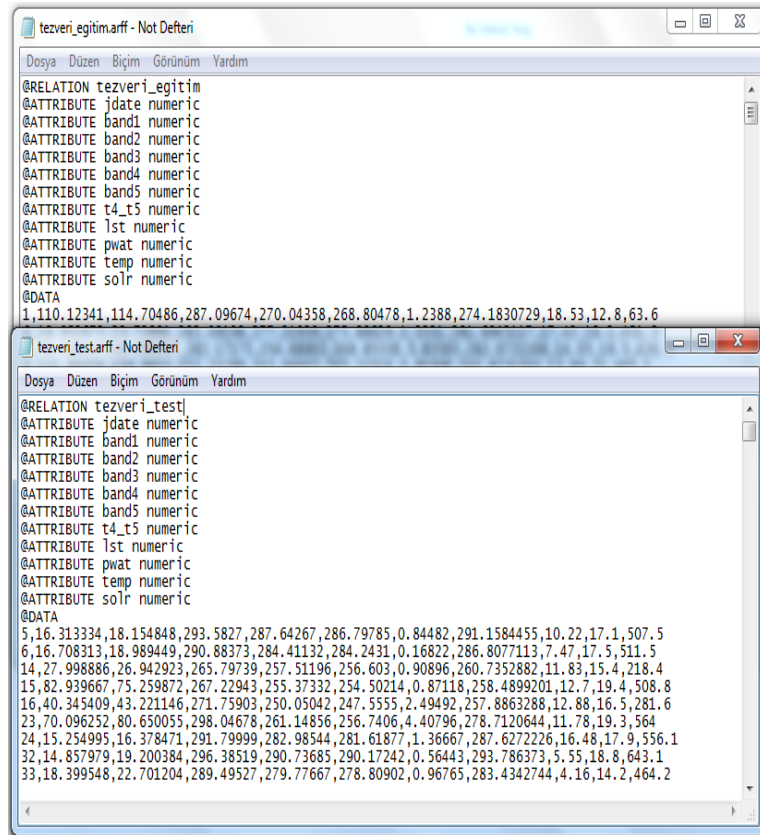
algoritmaların ve filtrelerin seçimi ve yapılandırılması için yer görevi görür. Denetimli öğrenme algoritmalarını içeren sınıflandırmanın yönü, regresyon tabanlı işlemlerin oluşturulmasıyla ilgilidir. Denetimsiz öğrenme algoritmalarının uygulanması Kümeleme sekmesinde gerçekleştirilebilir. İlişkilendirme ve Özellik seçimi sekmeleri, veri alanları üzerinde belirli filtreleme işlemlerinin yürütüldüğü ve veriler arasındaki karşılıklı ilişkilerin ayırt edildiği alanları oluşturur (Aydemir, 2018).

Experimenter ara yüzü; Explorer kullanılarak da yürütülebilen karşılaştırmalı deneysel tasarımları ve operasyonel görevleri kolaylaştırır. Örneğin, aynı veya farklı veri kümelerine uygulandığında çeşitli algoritmaların performans farklılıkları gözlemlenebilir. Arayüz üç ayrı sekmede yapılandırılmıştır: Deneysel tasarımla ilgili veri ve algoritmaların seçimi için belirlenmiş kurulum; Deneyle ilgili yürütülmesi ve hataların bildirimi için kullanılan çalıştır; ve deneysel sonuçların incelenmesine adanmış analiz. Ayrıca, veriler kümesindeki algoritmaların karşılaştırmalı performansını değerlendirmek için istatistiksel t-testi kullanılabilir.

Knowledge Flow ara yüzü; görsel varlıklar aracılığıyla temsil edilen iş süreçlerinin bağlantı mantığına uygun olarak, Explorer içinde yürütülebilen ilişkilendirme işlemlerinin temel prensibi ile uyumlu olarak çalışır. Farklı kodların arka planda yürüteceği işlemler, açık nesnelere arasındaki ilişkilerin görsel olarak oluşturulabileceği segmenti temsil eder.

Workbench ara yüzü; Bu arayüz, diğer tüm arayüzleri tekil bir varlığa entegre eden ve eşzamanlı kullanımlarını kolaylaştıran kapsamlı bir platform görevi görür. Birleştirilmiş bir birimden gelişmiş profesyonel işlemlerin yürütülmesini sağlar. Çeşitli tanımlayıcı istatistiksel teknikler aracılığıyla verileri normalleştirmeye, standartlaştırmaya, görselleştirmeye ve sunmaya yardımcı olan çok çeşitli araçları kapsar.

Simple CLI ara yüzü; komut satırı aracılığıyla kod girişi yoluyla işlemleri yürüten işlevsel bir arayüz oluşturur. Yardım komutunu kullanarak ek komutlarla ilgili bilgi edinilebilir. Weka yazılımının geliştirilmesinde, Java programlama dili temel yazılım olarak hizmet eder ve görsel arayüzlerdeki tüm işlemlerin kod aracılığıyla yürütülmesini sağlar ve özellikle profesyonel kullanıcılara hitap eder. Java API yazılım nesnesi sayesinde, programcılara Weka programının işlevlerini ilgili uygulamalarına entegre etme fırsatı verilir. Son gelişmeler, R istatistik yazılımının eşzamanlı kullanımını kolaylaştırarak veri analizi ve madencilik amaçları için faydalı bir işbirliğini sağlar.



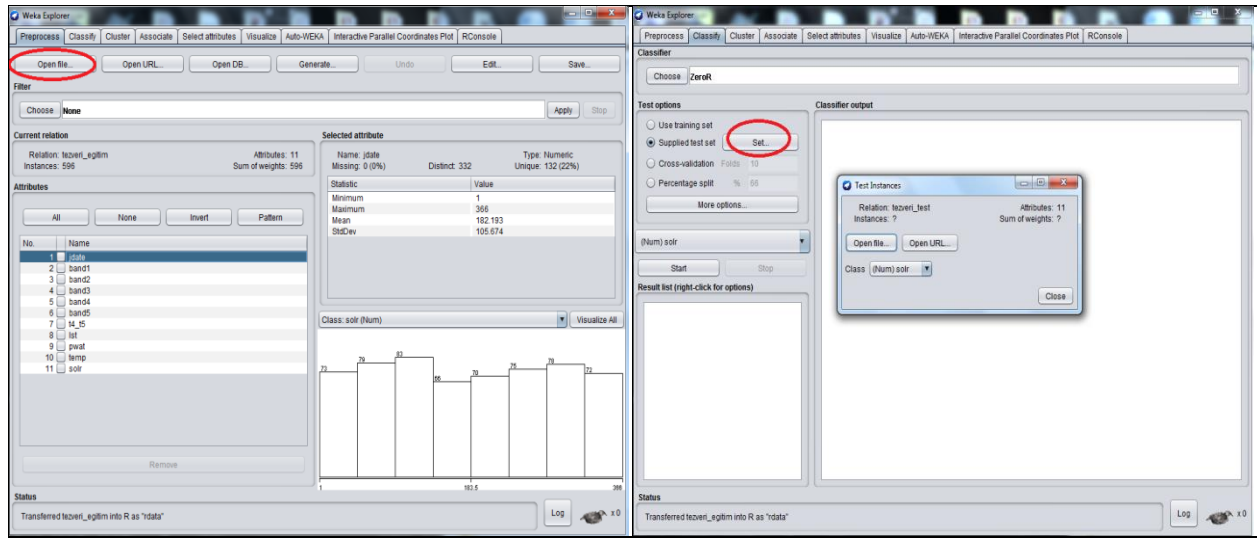
```
tezveri_egitim.arff - Not Defteri
Dosya Düzen Biçim Görünüm Yardım
@RELATION tezveri_egitim
@ATTRIBUTE jdate numeric
@ATTRIBUTE band1 numeric
@ATTRIBUTE band2 numeric
@ATTRIBUTE band3 numeric
@ATTRIBUTE band4 numeric
@ATTRIBUTE band5 numeric
@ATTRIBUTE t4_t5 numeric
@ATTRIBUTE lst numeric
@ATTRIBUTE pwat numeric
@ATTRIBUTE temp numeric
@ATTRIBUTE solr numeric
@DATA
1,110.12341,114.70486,287.09674,270.04358,268.80478,1.2388,274.1830729,18.53,12.8,63.6

tezveri_test.arff - Not Defteri
Dosya Düzen Biçim Görünüm Yardım
@RELATION tezveri_test
@ATTRIBUTE jdate numeric
@ATTRIBUTE band1 numeric
@ATTRIBUTE band2 numeric
@ATTRIBUTE band3 numeric
@ATTRIBUTE band4 numeric
@ATTRIBUTE band5 numeric
@ATTRIBUTE t4_t5 numeric
@ATTRIBUTE lst numeric
@ATTRIBUTE pwat numeric
@ATTRIBUTE temp numeric
@ATTRIBUTE solr numeric
@DATA
5,16.313334,18.154848,293.5827,287.64267,286.79785,0.84482,291.1584455,10.22,17.1,507.5
6,16.708313,18.989449,290.88373,284.41132,284.2431,0.16822,286.8077113,7.47,17.5,511.5
14,27.998886,26.942923,265.79739,257.51196,256.603,0.90896,260.7352882,11.83,15.4,218.4
15,82.939667,75.259872,267.22943,255.37332,254.50214,0.87118,258.4899201,12.7,19.4,508.8
16,40.345409,43.221146,271.75903,250.05042,247.5555,2.49492,257.8863288,12.88,16.5,281.6
23,70.096252,80.650055,298.04678,261.14856,256.7406,4.40796,278.7120644,11.78,19.3,564
24,15.254995,16.378471,291.79999,282.98544,281.61877,1.36667,287.6272226,16.48,17.9,556.1
32,14.857979,19.200384,296.38519,290.73685,290.17242,0.56443,293.786373,5.55,18.8,643.1
33,18.399548,22.701204,289.49527,279.77667,278.80902,0.96765,283.4342744,4.16,14.2,464.2
```

Şekil 3. ARFF Dosya Yapısı

WEKA, çeşitli veritabanlarından veri ayıklama yeteneğine sahip olsa da, csv gibi metin dosyalarını da işleyebilmesine rağmen, birincil veri dosyası olarak hizmet veren arff formatındaki dosyaların kullanılması tavsiye edilir. Temel olarak bir metin dosyası olan arff formatında, yapı, dosya adı için @relation kodlaması, her veri alanı için @attribute, virgülle ayrılan, @data kodlaması altındaki sütun sayısına göre düzenlenen değerlerle tanımlanır. Bu şekilde oluşturulmuş bir kayıt, belirli bir güne ilişkin olan, beraberindeki tabloda belirtilen niteliklere karşılık gelen bir veri girişini temsil eder. Veri setinde, 11'i de sayısal (nümerik) özellikte olan veri alanı yer almaktadır. Yapılan hesaplamalardan ve hazır olarak ölçüm verilerinden elde edilen veri alanları aynı güne gelenleri bir satırda olacak şekilde bir Arff dosyasında toplanmıştır.

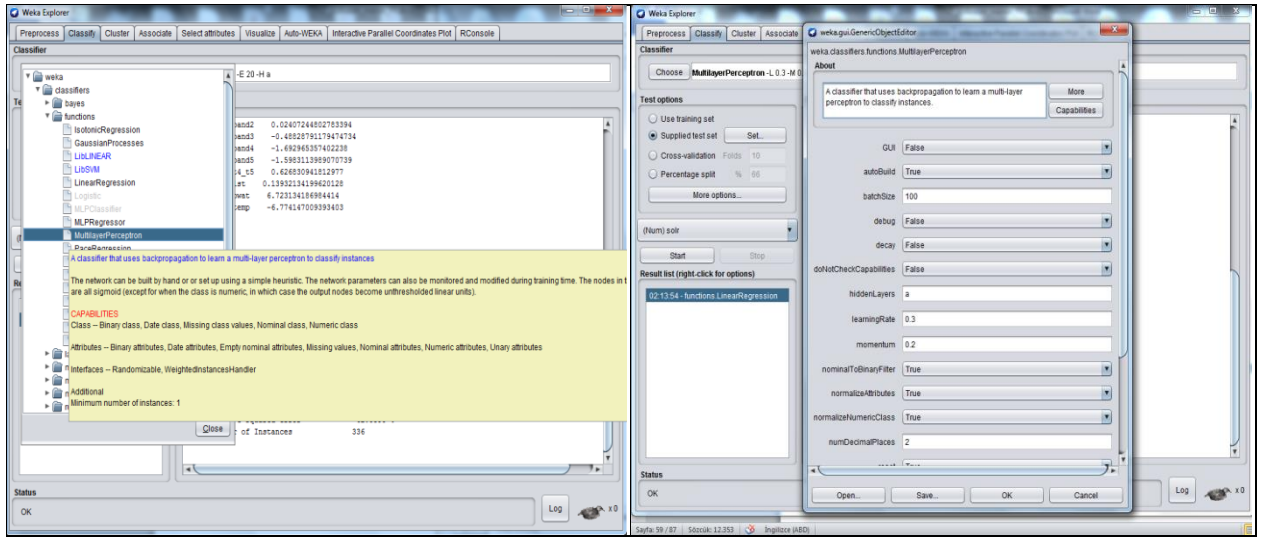
Eğitim ve test uygulamaları için belirlenmiş veri kümeleri, WEKA yazılım ortamında bağımsız olarak kurulur. Bu veri kümesi girişinin seçimi, kullanıcı tarafından belirtilen veri miktarının belirlenmesine ek olarak, kullanıcı özelliklerine göre yazılım tarafından özerk olarak ta yürütülebilir. Eğitim veri setini ön işlem sekmesindeki Weka explorer arayüzü üzerinden yükleme işlemi Şekil 4'te gösterilmektedir. Test veri seti ise Classify sekmesi altında yer alan Supplied test set isimli seçenek üzerinden yüklenerek makine öğrenmesi metodlarını çalıştırmaya uygun hale getirilmiştir.



Şekil 4. ARFF Eğitim ve test verisi yükleme ekranı

ANN tabanlı algoritmalar içinde en yaygın kullanımı olan MLP hem sınıflandırma hem de tahmin problemlerinde başarılı sonuçlar ortaya koyduğu için güneş enerjisi tahmininde araştırmacıların sıklıkla başvurduğu bir yöntemdir (Şenkal,2019).

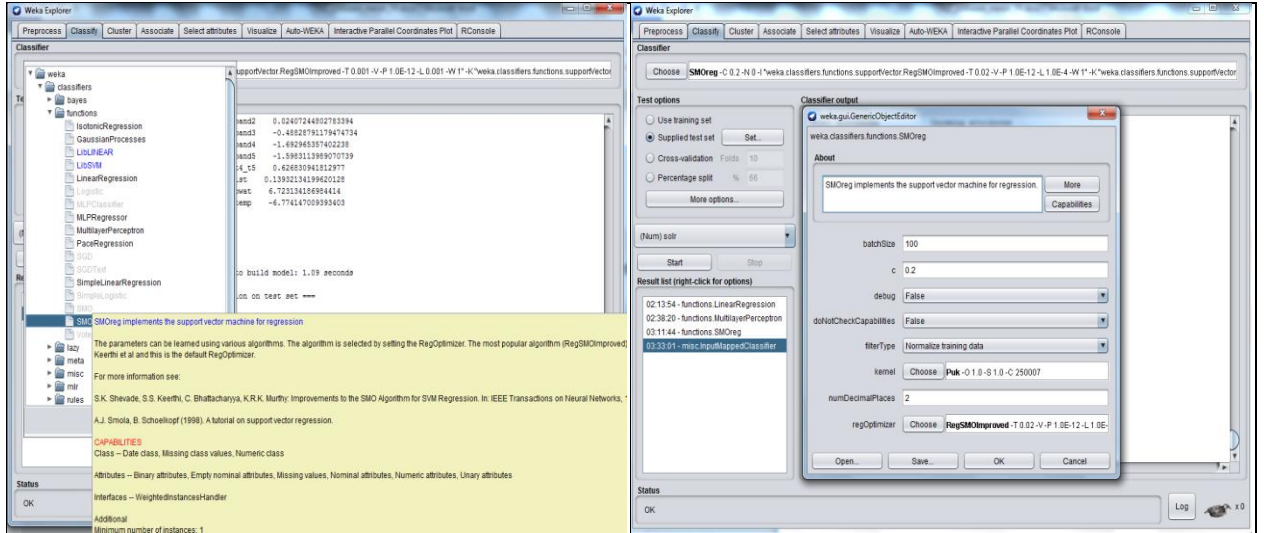
Bu çalışmada kullanılan ANN algoritmalarından biri olan MLP metodu kullanılmıştır. Metodunun seçimi için Explorer ara yüzü altında, Preprocess sekmesinden, veri seti yüklenip, Classify sekmesinden, functions grubundan MLP algoritması seçilip, çalıştırılır. Şekil 5. MLP seçim ve ayar ekranını göstermektedir. Bu çalışmada yapılan birçok deneme neticesinde WEKA ayarlarının (weka.classifiers.misc.InputMappedClassifier-I-trim-W weka.classifiers.functions.MultilayerPerceptron -- -L 0.234 -M 0.123 -N 500 -V 0 -S 0 -E 20 -H 21 -output-debug-info) olarak seçilip, algoritma özelliklerinden Debug aktif hale getirilip, gizli katmandaki nöron sayısını otomatik yerine 21 adet seçilmesi, öğrenme oranını 0.234'e ve momentum değeri 0.123'a ayarlanması durumunda en iyi MLP sonuçlarına ulaşılmıştır.



Şekil 5. MLP algoritması seçim ve ayar ekranı

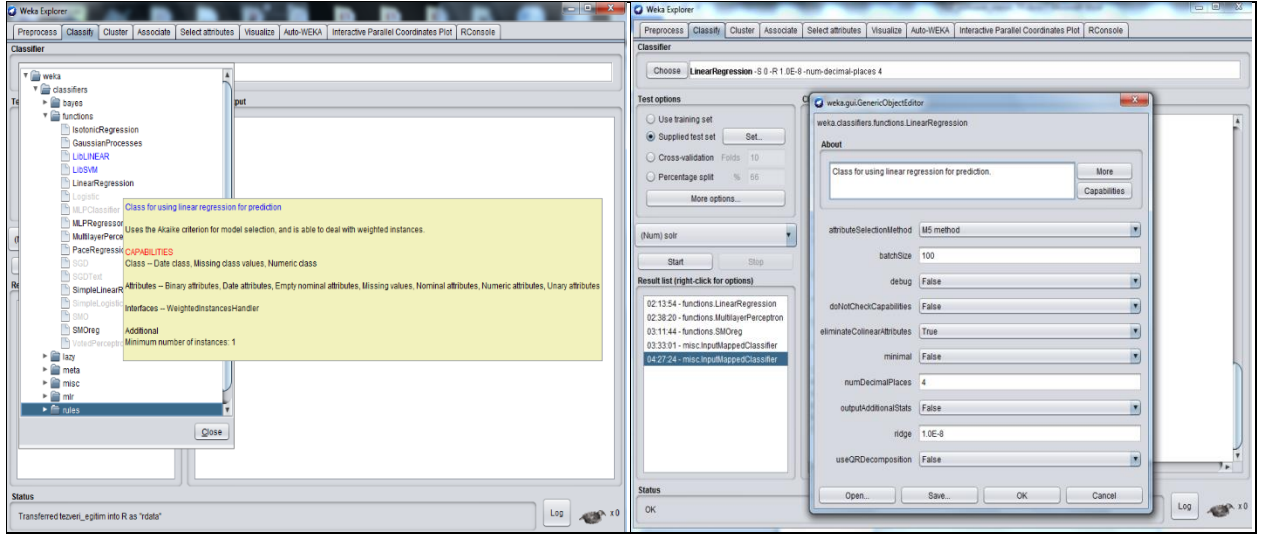
SVR algoritması WEKA altında Sequential Minimal Optimization Regressor (SMOReg) adıyla functions grubu altında bulunmaktadır. Verileri ikililer şeklinde alarak, optimizasyonu yaparak sonuca ulaşan bu algoritma çeşitli çekirdek fonksiyonlarına sahiptir (Baitharu ve ark 2015). Şekil 6’da SVR için seçim ekranı görülmektedir.

Explorer ara yüzü altında, Preprocess sekmesinden veri seti yüklenip, Classify sekmesinden, functions grubundan SVR algoritması (“weka.classifiers.misc.InputMappedClassifier -I -trim -W weka.classifiers.functions.SMOReg --- C 0.4 -N 0-I weka.classifiers.functions.supportVector.RegSMOImproved -T 0.025 -V -P 1.0E-12 -L 1.0E-4 -W 1” - K “weka.classifiers.functions.supportVector.Puk-O 0.6-S 1.5-C 250007”) şeklinde seçilip, algoritma özelliklerinden optimize edici olarak RegSMOImproved ile çekirdek türü olarak Puk tipi seçilip, çalıştırılmıştır.



Şekil 6. SVR Algoritması Seçim ve Ayar Ekranı

Explorer ara yüzü altında, Preprocess sekmesinden veri seti yüklenip, Şekil 7’de görüldüğü gibi Classify sekmesinden, functions grubundan lineer regresyon algoritması seçilmiştir. Regresyon analizi ayarlarından; (weka.classifiers.functions.LinearRegression-S 2-R 1.0E-8-num-decimal-places4) seçilip, özellik seçiminde GreedyMethod kullanıma eklenip, çalıştırılmıştır.



Şekil 7. LR Algoritması Seçim ve Ayar Ekranı

3. SONUÇLAR

Gelişmiş Çok Yüksek Çözünürlüklü Radyometre (AVHRR) sensöründen elde edilen beş uydu görüntüsü değeri, bant 4 ve bant 5 arasındaki fark değerleri, kara yüzey sıcaklığı (LST) değerleri, yoğunlaşabilir su buharı (PWAT) ve sıcaklık metrikleri giriş parametreleri olarak kullanılırken, karasal yüzeye gelen güneş ışınımı değerleri çıktı parametreleri olarak kullanılmıştır. Bu araştırmaya dahil edilen tüm veri kümeleri sekiz ana sınıflandırmaya ayrılmıştır. Bu sınıflandırmanın amacı, en uygun sonuçları veren optimal veri kümelerini belirlemek ve Dünya yüzeyine ulaşan güneş radyasyonu üzerinde etkili olan değişkenleri belirlemektir.

Çizelge 3'te açıklanan sonuçlar, korelasyon katsayıları (R) ve kök ortalama kare hatası (RMSE) değerleri ile değerlendirilen en uygun sonuçların, yalnızca uydu kaynaklı verilerden veya yalnızca meteorolojik verilerden kaynaklanmadığını göstermektedir. Tüm giriş parametrelerini tek bir kategori olarak kapsayan A grubu veri kümesidir. Destek Vektör Makinesi (SVM) algoritması $R = 0.889$ ve $RMSE = 113.16$ korelasyon değerleriyle üstün performans elde ederken, Çok Katmanlı Algılayıcı (MLP) algoritması da iyi sonuçlar vermiştir. Tersine, Doğrusal Regresyon (LR) algoritması, yukarıda belirtilen algoritmalara göre belirgin şekilde düşük performans göstermiştir.

Çizelge 3. Veri setlerine göre algoritmaların ilişki ve hata değerleri

| VERİ | | MLP | | SVR | | LR | |
|--------------------------------------------------------------------|------------|--------------|--------|--------------|--------|--------------|--------|
| Alanları | Grubu | R | RMSE | R | RMSE | R | RMSE |
| JDATE BAND1 BAND2 BAND3 BAND4 BAND5 T4-T5 LST PWAT TEMP SOLR | A Grubu | 0.860 | 153.94 | 0.889 | 113.16 | 0.799 | 148.96 |
| JDATE BAND1 BAND2 BAND3 BAND4 BAND5 T4-T5 LST SOLR | B Grubu | 0.628 | 226.10 | 0.753 | 166.67 | 0.591 | 200.92 |
| JDATE BAND1 BAND2 BAND3 BAND4 BAND5 SOLR | C Grubu | 0.732 | 273.89 | 0.753 | 166.98 | 0.496 | 219.74 |
| JDATE T4-T5 LST PWAT TEMP SOLR | D Grubu | 0.817 | 187.54 | 0.892 | 111.97 | 0.768 | 157.68 |
| JDATE BAND1 BAND2 BAND3 BAND4 BAND5 PWAT TEMP SOLR | E Grubu | 0.884 | 116.27 | 0.888 | 113.70 | 0.781 | 154.76 |

| | | | | | | | |
|-----------------------------------------------------|---------|-------|--------|-------|--------|-------|--------|
| JDATE BAND1 BAND2 T4-T5 LST PWAT TEMP SOLR | F Grubu | 0.836 | 136.05 | 0.890 | 113.06 | 0.782 | 153.37 |
| JDATE BAND3 BAND4 BAND5 T4-T5 LST PWAT TEMP SOLR | G Grubu | 0.857 | 138.13 | 0.890 | 112.52 | 0.798 | 148.58 |
| JDATE PWAT TEMP SOLR | H Grubu | 0.869 | 140.11 | 0.887 | 114.56 | 0.797 | 148.70 |

Bu çalışmada bulunan en kötü tahmin değerleri, B ve C gruplarındaki veri kümeleriyle ilişkilendirilmiştir. Bu bulgular, bu araştırma alanında meteorolojik parametrelerin gerekliliğinin altını çizmektedir. PWAT, bant 4 ve bant 5 (T4-T5) arasındaki fark ve LST dahil olmak üzere ham uydu bant verilerinden çıkarılan meteorolojik ürünlerden oluşan veri kümesinde, meteorolojik verilerin yanı sıra SVM algoritması $R = 0.892$ ve $RMSE = 111.97$ değerleri üretirken, MLP algoritması $R = 0.817$ ve $RMSE = 187.54$ ile nispeten karşılaştırılabilir başarılı sonuçlar vermiştir. Ham uydu verisi yerine ham verilerden hesaplanmış ürünlerin kullanılmasının daha kullanışlı olduğu ortaya çıkmıştır.

Yapılan tüm analizlerde, Julian Date (JDATE) verilerinin veri kümelerinden hariç tutulması, tahmini performansın azalmasına neden olmuştur. Bu fenomen, Jülyen tarihinin dünya yüzeyine gelen güneş radyasyonunun hesaplanmasında çok önemli bir değişken görevi görmesi gerçeğine bağlanabilir. Sonuç olarak, JDATE verileri tüm veri kümelerinde bir giriş parametresi olarak dahil edilmelidir. (Ersin, M.,2021). Sonuçların incelenmesi ile, SVM algoritmasının D grubu veri kümesi içinde en yüksek tahmin değerini verdiği açıktır, kızılötesi uydu verilerinin ise MLP ve LR algoritmalarına girdi olarak entegrasyonu, başarı katsayılarının mütevazı bir şekilde artmasıyla sonuçlanmıştır.

Başarı oranlarının artırılması, veri kümesinin büyüklüğü ve değişkenliği ile ilgili model parametrelerine bağlıdır. Bu tür zorlukların ele alınmasında, veri kümesi boyutu, duyarlılık ve doğruluk metriklerini artırabilen, binlerce kaydı kapsayan, kapsamlı boylamsal gözlemlerden elde edilen veriler kullanılarak desteklenebilir. Bu bağlamlarda geliştirilen ve daha kapsamlı veri kümelerini kullanan makine öğrenimi algoritmalarının uygulanmasının, gelecekteki bilimsel araştırmaların doğruluk metriklerini artırmasının muhtemel olduğu açıktır.

Ayrıca bu çalışmada kullanılan AVHRR sensörü tarafından yakalanan dalga boylarından farklı algılama aralıklarını kullanan çalışmalarda farklı sonuçlar gözlemlenebilir. Gelecekteki çalışmalar çeşitli veri kümelerinin ve alternatif algoritmaların dahil edilmesiyle desteklenebilir ve bu çalışmada elde edilen sonuçların iyi bir altlık sunacağı düşünülmektedir.

KAYNAKLAR

- Ackerman, S.A.,** Knox, J. A., 2015, *Meteoroloji Atmosferimizi Anlamak*, Nobel Yayınevi, 3. Baskıdan çeviri / 602 Syf.
- Alpaydın, E.,** 2012. *Yapay Öğrenme*, Boğaziçi Üniversitesi Yayınevi, İstanbul, 496s.
- Aydemir, E.,** 2018. *Weka ile Yapay Zekâ*, Seçkin Yayınevi, Ankara, s216.
- Cortes, C., Vapnik, V.,** 1995. Support-vector networks, *Mach. Learn.*, 20 (3) 82
- Ersin, M.,** 2021. Uydu ve Meteoroloji Verilerinden Makine Öğrenmesi Yöntemleri Kullanılarak Adana Bölgesi İçin Güneş Enerjisi Potansiyelinin Tahmini, Ç.Ü. Fen Bilimleri Enstitüsü Uzaktan Algılama ve CBS Ana Bilim Dalı Yüksek Lisans Tezi, Adana.
- Kaba, K., Sarıgül, M., Avcı, M., Kandırılmaz, H.M.,** 2018. Estimation Of Daily Global Solar Radiation Using Deep Learning Model. *Energy Volume 162*, 1, Pages 126-135.
- Karaman, Ö.A., Ağır, T.T., Arsel, İ.,** 2021. Estimation Of Solar Radiation Using Modern Methods, *Alexandria Engineering Journal*, 60(2):. 2447-2455.
- Şeker, Ş.E.,** 2013. *Weka İle İş Zekası ve Veri Madenciliği*, Cinius Yayınları, İstanbul, s206 Syf.
- Senkal, O.,** 2018. Comparison Of Incoming Solar Radiation At Different Air Density Regimes Using Neural Network Model, *Russ. Meteorol. Hydrol.* 43, 49–55.

Şenkal, O., 2010. Modeling Of Solar Radiation Using Remote Sensing And Artificial Neural Network İn Turkey, *Energy*, Volume 35, Issue 12(2010) 4795:4801, ISSN 0360-5442 85

Şenkal, O., Kuleli, T., 2009. Estimation Of Solar Radiation Over Turkey Using Artificial Neural Network And Satellite Data, *Applied Energy*, Volume 86, Issues 7(8): 1222-1228, ISSN 0306-2619.

Yıldız, B.Y., 2009. Türkiye'nin Bölgesel Atmosferik Değişkenlere Bağlı LST Algoritmalarının Oluşturulması, *Doktora Tezi*, Ç.Ü. Fen Bilimleri Enstitüsü Fizik Ana Bilim Dalı, Adana.

Zhou, Y., Liu, Y., Wang, D., Liu, X., Wangb, Y., 2021 A Review On Global Solar Radiation Prediction With Machine Learning Models In A Comprehensive Perspective, *Energy Conversion And Management*, Volume 235, 113960.

<http://www.cs.waikato.ac.nz/~ml/weka/> [Erişim tarihi: 01.10.2021].