



Machine learning-based weather prediction with radiosonde observations

Eralp Göğen*^{ID}, Selda Güney^{ID}

Department of Electrical and Electronics Engineering, Faculty of Engineering, Başkent University, 06790, Etimesgut, Ankara, Türkiye

Highlights:

- Successful Temperature Estimation by using Radiosonde device
- Analysis of Machine Learning Algorithms for Radiosonde data
- Effects of normalization, feature extraction and feature selection for Radiosonde data

Keywords:

- Radiosonde
- Weather Forecast
- Regression

Article Info:

Research Article
Received: 16.08.2022
Accepted: 04.09.2023

DOI:

10.17341/gazimmfd.1163079

Correspondence:

Author: Eralp Göğen
e-mail: eralpgogen@
magnetrontr.com
phone: +90 312 2466666 /
4014

Graphical/Tabular Abstract

Weather forecasting from past to present is important for humanity. The precise realization of the weather forecast can ensure that the negative effects that will occur are minimized by taking precautions against natural disasters such as floods, tsunamis, etc. Unlike previous temperature prediction studies in the literature, a three-year dataset of Radiosonde observations is utilized in this study. This dataset, measured at intervals of 1mbar up to an altitude of 40 km from the ground, allows for a more accurate modeling of the atmosphere compared to other studies in the literature. In this model, predictions for the highest and lowest temperatures for the next day are made. In this stage, the effects of normalization, feature extraction, and selection on the results are analyzed, and the most suitable model for prediction is determined. Block diagram of methods applied to Radiosonde data in this study is given in Figure A.

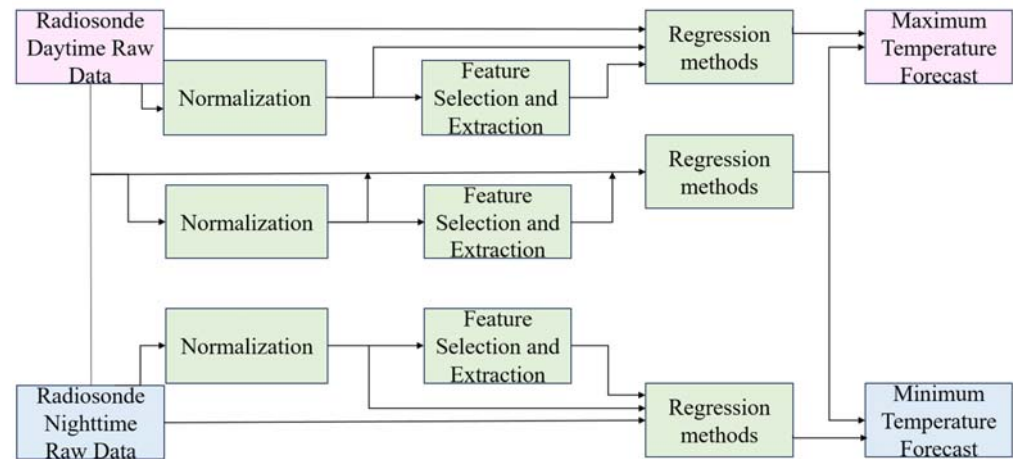


Figure A. Block diagram of methods applied to Radiosonde data

Purpose: Within the scope of this study, weather forecasting is made using radiosonde data. In this estimation, the highest and lowest temperatures are estimated

Theory and Methods: Estimation was made using Machine Learning Algorithms. Unlike the temperature estimation studies previously in the literature, 3-year radiosonde observation data were used. In this way, the atmosphere was modeled much more precisely than other studies in the literature with the data measured at 1mbar intervals up to 40 km above the ground. In this model, the highest and lowest temperature values for the next day are estimated. Different regression methods were compared with the software performed in MATLAB environment.

Results: As a result of these analyses, the highest temperature estimate for the next day was obtained with the highest accuracy using the Gaussian Process Regression method, with an RMSE (Root Mean Square Error) value of 1,2. Using the same method, the lowest temperature estimate was made with an average Root Mean Square Deviation of 2.4.

Conclusion: The results show that more successful temperature estimation is made than the studies in the literature. In this study, the effect of normalization for Radiosonde data was also examined and it was observed that normalization reduced the success. This reveals that some features in the data are more important. The pressure value is higher than other features and is observed to be more effective in estimating temperature.



Radyosonde rasatları ile makine öğrenmesi tabanlı hava durumu kestirimi

Eralp Göğen*^{ID}, Selda Güney^{ID}

Başkent Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, 06790, Etimesgut, Ankara, Türkiye

ÖNEÇIKANLAR

- Radyosonde cihazı kullanılarak Başarılı Sıcaklık Tahmini
- Radyosonde verileri için Makine Öğrenimi Algoritmalarının Analizi
- Radyosonde verileri için normalleştirme, öznetelik çıkarma ve öznetelik seçiminin etkileri

Makale Bilgileri

Araştırma Makalesi

Geliş: 16.08.2022

Kabul: 04.09.2023

DOI:

10.17341/gazimmfd.1163079

Anahtar Kelimeler:

Radyosonde,
hava durumu tahmini,
regresyon

ÖZ

Geçmişten günümüze hava tahmini insanlık için önem arz etmektedir. Hava tahmininin hassas gerçekleştirilebilmesi sel, tsunami vb. doğal afetlere karşı önlemler alınarak oluşacak olumsuz etkileri en düşük seviyeye indirmeyi sağlayabilmektedir. Bu çalışma kapsamında Radyosonde verileri kullanılarak hava durumu kestirimi yapılmaktadır. Bu kestirimde en yüksek ve en düşük sıcaklık tahmini yapılmaktadır. Makine öğrenmesi algoritmaları kullanarak kestirim gerçekleştirilmiştir. Daha önce literatürde bulunan sıcaklık tahmini çalışmalardan farklı olarak 3 yıllık Radyosonde rasat verileri kullanılmıştır. Bu sayede yerden 40 km yüksekliğe kadar 1mbar aralıklarla ölçülmüş veriler ile atmosfer, literatürdeki diğer çalışmalara göre çok daha hassas olarak modellenmiştir. Bu modelde ertesi güne ait en yüksek ve en düşük sıcaklık değerleri kestirilmiştir. Bu aşamada normalizasyon ve öznetelik çıkarma veya seçiminin sonuçlara etkileri analiz edilerek tahmin için en uygun model belirlenmiştir. MATLAB ortamında gerçekleştirilen yazılım ile farklı regresyon yöntemleri karşılaştırılmıştır. Bu analizler sonucunda Gauss Süreci Regresyonu (GSR) yöntemini kullanarak 1,2 ortalama karekök sapması ile ertesi güne ait en yüksek sıcaklık tahmini en yüksek doğrulukla elde edilmiştir. Aynı yöntem kullanarak 2,4 ortalama karekök sapması oranı ile en düşük sıcaklık tahmini yapılmıştır. Sonuçlar literatürdeki çalışmalardan daha başarılı sıcaklık tahmini yapıldığını göstermektedir.

Machine learning-based weather prediction with radiosonde observations

HIGHLIGHTS

- Successful temperature estimation by using Radiosonde device
- Analysis of machine learning algorithms for Radiosonde data
- Effects of normalization, feature extraction and feature selection for Radiosonde data

Article Info

Research Article

Received: 16.08.2022

Accepted: 04.09.2023

DOI:

10.17341/gazimmfd.1163079

Keywords:

Radyosonde,
weather forecast,
regression

ABSTRACT

From the past to the present, weather forecasting holds significant importance for humanity. The precise execution of weather forecasting enables the implementation of precautions against natural disasters such as floods, tsunamis, etc., thereby minimizing the adverse effects that may arise. In this study, weather prediction is conducted using Radiosonde data. Within this prediction, estimations for both the highest and lowest temperatures are made employing machine learning algorithms. Unlike previous temperature prediction studies in the literature, a three-year dataset of Radiosonde observations is utilized. This dataset, measured at intervals of 1mbar up to an altitude of 40 km from the ground, allows for a more accurate modeling of the atmosphere compared to other studies in the literature. In this model, predictions for the highest and lowest temperatures for the next day are made. In this stage, the effects of normalization, feature extraction, or selection on the results are analyzed, and the most suitable model for prediction is determined. The software, implemented in the MATLAB environment, compares different regression methods. As a result of these analyses, utilizing the Gaussian Process Regression (GPR) method, the highest temperature prediction for the next day is achieved with the highest accuracy, with a mean square root deviation of 1.2. Using the same method, the lowest temperature prediction is made with a mean square root deviation ratio of 2.4. The results indicate more successful temperature predictions compared to studies in the literature.

*Sorumlu Yazar/Yazarlar / Corresponding Author/Authors : *eralpgogen@magnetrontr.com, seldaguney@baskent.edu.tr /
Tel: +90 312 2466666 / 4014

1. Giriş (Introduction)

Belirli bir ülke, bölge veya merkezde, bir zaman dilimi içinde görülebilecek meteorolojik olayların gözlem ve analizlere dayanarak sübjektif veya objektif yöntemler kullanılarak önceden öngörülme çalışmaları hava durumu tahmini olarak adlandırılır. Hava tahmini, gözlemler, analiz ve tahmin olmak üzere üç aşamada belirlenir. Hava durumu atmosferdeki belirli şartların özel yerler ve zamanlarda bir araya gelmesinden oluşur. Hava durumunu sıcaklık, nem, rüzgâr hızı, bulutluluk, yağmur, atmosfer basıncı ve bunun gibi birçok faktör etkilemektedir. Mümkün olduğunca doğru hava durumunu tahmin etmek her zaman meteorologların bir numaralı önceliği olmuştur. Hava durumunun hatasız tahmin edilmesi dünya genelinde çok önemlidir. Hatalı gerçekleştirilen tahminler ülkelere ekonomik olarak milyarlarca liralık zararlar oluşturabilir. Bunun yanı sıra can ve mal güvenliği içinde tehlike teşkil etmektedir. Zamanında doğru tahmin edilen hava durumu ile can ve mal kayıplarının önüne geçilebilir. Bu nedenle hava durumunu tahmin etmek için matematiksel ve istatistiksel tabanlı denklemler kullanılır. Ancak, küresel ısınmadan dolayı artık hava durumu daha da karışık ve kaotik olmaya başlamıştır. Bu sebepten dolayı matematiksel yöntemler yeterli olmamaktadır. Hâlâ istatistiksel yöntemler kullanılsa da teknolojinin ve yapay zekânın gelişmesiyle bu yöntemler yerini ya hibrit yöntemlere ya da yapay zekâyâ dayanan yöntemlere bırakmıştır. Bu nedenle makine öğrenme algoritmaları ile hava durumu kestirimi önemli bir çalışma konusudur [1].

Dünya genelinde farklı ülkelere ait meteoroloji birimleri dünya ile eş zamanlı olarak Radyosonde cihazlarını meteoroloji balonları ile kullanarak rasatlar gerçekleştirirler. Bu rasatlar gündüz ve gece rasatları olarak adlandırılırlar. Gerçekleştirilen bu rasatlarda yerden 40 km irtifaya kadar 1mbar aralıklarla her bir yükseklik noktasına ait sıcaklık, işba sıcaklığı, nem, rüzgâr hızı ve rüzgâr yönü değerleri ölçülür. Ardından gerçekleştirilen bütün bu rasatlar Dünya Meteoroloji Organizasyonu'na (DMO) gönderilerek 3-boyutlu atmosfer modeli elde edilir. Bu model kullanılarak istatistiksel hava durumu tahmini gerçekleştirilir.

Radyosonde hava durumu tahmininde kullanılan, balon ile atmosfere yollanarak atmosferin farklı katmanlarındaki meteorolojik verileri toplayan sensörlerden oluşan bir cihazdır. Radyosonde rasatlarının yanı sıra uydu, radar ve otomatik yer gözlem istasyonu ölçümleri de gerçekleştirilmektedir. Ancak, hava durumu kestirimi için farklı irtifalara ait ölçümleri barındırdığından dolayı en baskın rasat türü Radyosonde rasatlarıdır. Bu çalışmada da Radyosonde verileri kullanılmıştır.

Bu çalışmada yer gözlem istasyonundan alınan sabit irtifalı veriler yerine yüksek irtifa meteoroloji ölçüm cihazı olan Radyosonde ile gerçekleştirilen 3-boyutlu rasat sonuçları kullanılarak geleneksel makine öğrenmesi algoritmaları ile eğitilmiştir. 2. Bölümde hava durumu tahmini ile ilişkili çalışmalar incelenmiştir. 3. Bölümde ise Radyosonde cihazı ve bu cihazdan elde edilen veriler anlatılmaktadır. Yine aynı bölümde tahmin için kullanılan yöntemler açıklanmıştır. Gerçekleştirilen eğitimler sonucunda elde edilen modeller ile ertesi güne ait en yüksek ve en düşük hava sıcaklığı değerleri tahmin edilmiştir. 4. Bölümde bu sonuçların analizi yapılmaktadır. Son bölümde bu çalışma kapsamında elde edilen veriler verilmektedir.

2. Literatür Taraması (Related Works)

Hava durumu kaotik bir durumdur ve bunu basit yöntemlerle kestirmek yanlış sonuçlara neden olabilir. Bu sebepten ötürü meteorologlar bu kaotik olayı anlamlandırıp tahmin yürütmek için istatistiksel tabanlı ve yapay zekâ tabanlı yöntemler ile çözmeye çalışmışlardır. Günümüzde yapay zekâ algoritmalarının gelişmesiyle yapay zekâ tabanlı kestirimler istatistiksel tabanlı kestirimlere göre daha keskin cevap vermeye başlamıştır [2].

Yapay Sinir Ağları (YSA) hava tahmini benzeri kaotik sorunları çözmek için uygun yöntemlerden biridir [3]. Bunun sebebi YSA insan beyni çalışmasından esinlenerek gerçekleştirilmiş ve doğrusal olmayan problemlerde de etkin sonuçlar vermektedir [1]. Literatür araştırmasında; öncelikle istatistiksel tabanlı çalışmalar, ardından da makine öğrenmesi tabanlı çalışmalar incelenmiştir.

Fente çalışmasında hava durumu verilerini Ulusal Hava Durumu Merkezi'nden (UHDM) toplamıştır [4]. Uzun Kısa Vadeli Bellek (Long Short Term Memory, LSTM) algoritması kullanılarak farklı kombinasyonlarla eğitilip test edilmiştir. Kullandığı veriler UHDM'den alınmış 10 yıllık verilerdir. Bu veriler; sırası ile sıcaklık, yağış, rüzgâr hızı, basınç, sis yoğunluğu ve nem verisinden oluşmuş ve otomatik istasyonlar tarafından toplanmıştır. Fente'nin çalışmasında kestirmeye çalıştığı günlük ortalama sıcaklık değeri ve başarısı %95'tir.

Sobrevilla çalışmasında Tiwi adasındaki hava durumunu tahmin etmek için esnek yayılma algoritması ile eğitim yapmıştır [1]. Bu çalışma günlük hava tahminini bulmak içindir. Bu çalışmada kullanılan veriler Gelişmiş Bilim ve Teknoloji Enstitüsü'nden toplanan 2012-2015 yılları arasındaki ölçümlerdir. Kullandığı verilerde bozulmalar ve boş alanlar olduğundan ilk olarak bu alanları Fourier uyumlandırma ve çoklu atama yöntemleri ile doldurarak eğitim gerçekleştirmiştir. İkinci eğitimde ise bu bozuk satırları tamamen eğitim verisinden çıkararak eğitim gerçekleştirmiştir. Yaptığı bu iki çalışma sayesinde düzeltmek yerine tamamen çıkarmanın daha etkili olduğu sonucuna ulaşmıştır. Bu çalışma 17000 yinelemede 10 gizli nöronla optimal modele ulaşmıştır. Bu çalışmaya göre günlük hava tahmininde, ortalama tahmin doğruluğu %98,96743 olarak elde etmiştir.

Nandar çalışmasında Mynmar'daki aylık yağmur oranlarını tahmin etmek için uğraşmıştır. Kullandığı öğrenme algoritması Bayesian Ağı (Likelihood weight) algoritmasıdır [5]. Algoritmanın başarısını da karışık ortalama kare hata (Root Mean Square Error, RMSE) başarı kriterini kullanarak hesaplamıştır. Kullandığı veri seti Mynmar'ın 1990 ile 2006 yıllarındaki yağmur yağış verileridir. Veri setini parçalama şekli 1990 ile 2005 arasını eğitim için kullanıp, 2006 yılındaki veriler ile oluşturulan modeli test etmiştir. Yağmur oranlarını kestirmede RMSE oranı %3 ile tahmin etmiştir.

Rahul çalışmasında Ulusal İklim Veri Merkezi'nden yer gözlem istasyonları tarafından ölçülen Kasım 2007'den Ekim 2017'ye kadarki hava durumu verileri ile eğitim yapmıştır [6]. Veri seti; sıcaklık, nem, çiy noktası, basınç, görünürlük ve yağış gibi birçok hava durumu özelliğinden oluşmaktadır. LSTM tekniği kullanılarak uygun bir doğruluk ile hava durumu tespit edilmek istenmiştir. Hava durumu %84 başarı ile doğru hesaplanmıştır.

3. Gereç ve Yöntemler (Material and Methods)

Hava durumu tahmini konusunda daha önce yapılan çalışmalar incelenmiştir. Bu çalışmalar içerisinde bazı çalışmalar hava durumunu kestirmek için matematiksel yöntemler kullanırken bazı araştırmacılar ise bu kestirimi makine öğrenmesi yöntemleri ile yapmıştır. Bu çalışmada ise Radyosonde verileri geleneksel makine öğrenmesi algoritmaları ile eğitilmiştir.

3.1. Radyosonde (Radyosonde)

Hava tahmininin gerçekleştirilebilmesi için atmosferin 3 boyutlu olarak modellenmesi gerekmektedir. Atmosferin modellenmesi için havanın yerden 40 km yüksekliğe kadar her bir yükseklik değeri karşılığında; sıcaklık, bağıl nem, rüzgâr hızı ve rüzgâr yönü bilgilerinin ölçülmesi gerekmektedir. Ancak bu bilgiler kullanılarak atmosferin modellenmesi ve bu modelin anlamlı bir şekilde kullanılabilmesi için en az 300 km aralıkla bütün dünyanın hava

durumunun aynı anda ölçülmesi gerekmektedir. Ayrıca, bu ölçümlerin günde en az iki kere tekrarlanması gereklidir.

Meteoroloji Genel Müdürlüğü (MGM), bütün dünya ile aynı anda Türkiye’de 9 (Ankara, İstanbul, Diyarbakır, Samsun, Adana, İzmir, Isparta, Erzurum ve Kayseri) farklı noktadan bu ölçümleri gerçekleştirmektedir. Ölçümler öğlen 14:30 ve gece 02:30’da dünya ülkeleri ile eşzamanlı olarak yapılmaktadır. Şekil 1’de MGM tarafından kullanılan bu cihazlar görülmektedir.

Radyosonde içerisinde bulundurduğu Küresel Konum Belirleme Sistemi (Global Navigation Satellite Systems, GNSS) sistemi sayesinde 1559 Mhz – 1610 Mhz aralığında yayın yaparak bütün küresel konumlandırma uydularına aynı anda bağlanabilmektedir [11].

Yükseklik ölçümü GNSS algılayıcısı tarafından ayrı ayrı veya beraber yapılabilmektedir. Bu sayede sinyal takibinin zor olduğu ortamlarda yükseklik bilgisi kopmadan rasata devam edilebilmektedir. Dahili ısıtıcı nem sensörü sayesinde bulut içi nemden kaynaklı buzlanmaların önüne geçilebilmekte ve ölçüm doğruluğu artırılmaktadır.

Bilgi Teknolojileri ve İletişim Kurumu (BTK) tarafından meteoroloji uygulamaları için tahsil edilmiş 400 MHz ile 406 MHz frekans aralığında yayın yapılabilen ve çıkış gücü 27 dBm’e kadar ayarlanabilmektedir. Şekil 2’de Radyosonde blok şeması görülmektedir. Radyosonde elektronik olarak; elektronik baskı devre kartı, GNSS alıcısı, nem sensörü, sıcaklık sensörü, işlemci, güç kontrol birimi, batarya, göndermeç ve opsiyonel basınç sensöründen mekanik olarak ise köpük kutusu, ip sargısı 30 metre ip ve antenden oluşmaktadır.

Tablo 1’de Radyosonde üzerinde bulunan sensörler ve bu sensörlere ait DMO tarafından belirlenmiş sahip olması gereken ölçüm aralıkları görülmektedir.

3.2. Veri Seti (Data Set)

Radyosonde rasat verisi TEMP adı verilen özel bir format ile toplanır. Rasat verisinde verileri ayırmak için ayrıç olarak “|” işareti kullanılır. Herhangi bir nedenden dolayı anlık veri doğru alınmadı ise o bölge boş bırakılmaz “/” işareti konulur. Veri toplama işlemi balon yükselirken her 1mb aralıklı olacak şekilde toplanır ve Şekil 3’de görülmektedir. Her bir satırda bulunan veriler ve örnek bir satır verisi aşağıda verilmektedir:

- İstasyon Numarası
- Yıl
- Ay
- Gün
- Rasat Saati
- Referans Saati
- Basınç Seviyesi
- Yükseklik
- Sıcaklık
- Nem
- İşba Sıcaklığı
- Rüzgâr Yönü
- Rüzgâr Hızı

istno|yıl|ay|gün|rasat_saati|referans_saati|
|basınç_seviyesi|yükseklik|sıcaklık|nem|işba_sıcaklığı|rüzgâr_yönü|r
üzgâr_hızı

Veri setinde kullanılan etiket verisi (ground truth) ise gerçek o günkü maksimum ve minimum sıcaklık verileri olarak kullanılmıştır. Gündüz rasatlarından alınan verilere o günkü maksimum sıcaklık, gece rasatlarında ise o günkü minimum sıcaklık etiketi konulmuştur.

Rasat verisi bir balonun ilgili yerden serbest bırakılıp patlayana kadar süredeki her bir 1mb artışında yukarıdaki gibi paylaşılan veri formatında topladığı hâlden oluşur. Bunu daha iyi açıklamak için örnek bir satır ve açıklaması aşağıda verilmektedir:



Vaisala (Finlandiya) [7]



Meteo Modem (Fransa) [8]

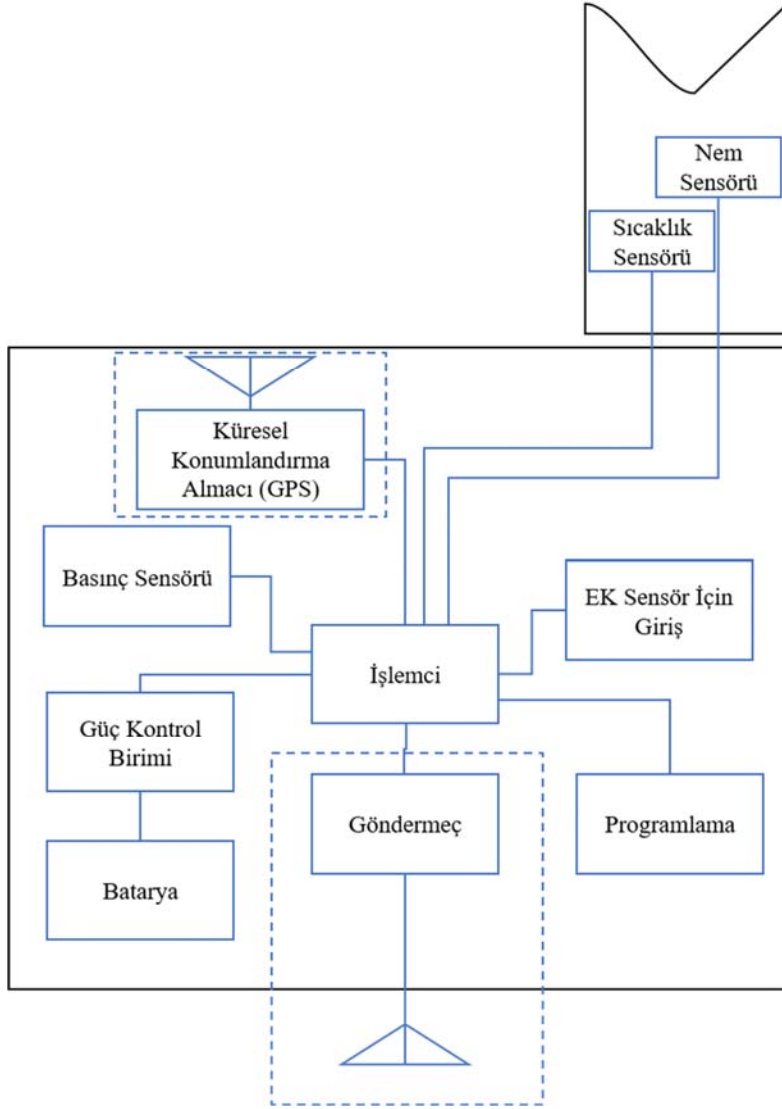


Meisei (Japonya) [9]



Graw (Almanya) [10]

Şekil 1. Meteoroloji Genel Müdürlüğü Tarafından Kullanılmakta Olan Yabancı Radyosonde Meteorolojik Ölçüm Cihazları (Foreign Radiosonde Meteorological Measurement Devices Used by the General Directorate of Meteorology)



Şekil 2. Radyosonde Blok Şeması (Radiosonde Block Diagram)

Örnek:

17130|2016|7|15|12|11:34:33|909|890|346|7|-62|300|52

17130 : istasyon no
 2016 : rasadın ait olduğu yıl
 7 : rasadın ait olduğu ay
 15 : rasadın ait olduğu gün
 12 : rasadın saati (G-T - yalnızca 00 veya 12 olabilir)
 11:34:33 : referans zamanı (balondan kaydın alındığı zaman)
 909 : ölçüm alınan basınç seviyesi (mb) (Sadece Yer seviyesinde basınç bir ondalıklı olmalı, üst seviyelerde tamsayı olmalıdır.
 890 : yükseklik (jpm) jeo potansiyel metre
 34,6 : sıcaklık (°C cinsinden sıcaklığın bir ondalıklı değeri.)
 7 : nem (% ve tamsayı olarak. Gerçek değer %'73'tür.)
 -6.2 : işba sıcaklığı (°C cinsinden, bir ondalıklı değeri)
 300 : rüzgâr yönü (1-360° üzerinden)
 5.2 : rüzgâr hızı (knot olarak, bir ondalıklı değeri)

Örnek kayıtlar:

17130|2016|7|15|12|11:34:33|909|890|346|7|-62|300|52

17130|2016|7|15|12|11:34:36|908|902|331|6|-95|281|95
 17130|2016|7|15|12|11:34:38|907|914|323|6|-101|295|70
 17130|2016|7|15|12|11:34:40|906|925|318|6|-96|290|97

Bu kayıtlardan elde edilen bu çalışmada kullanılan verilerin ham ve normalize edilmiş haline birkaç örnek Tablo 2'de verilmektedir.

3.2.1. İstasyon numarası (Station number)

DMO'nun dünyadaki rasat yapılan belli bölgelere verdiği bir numaradır. Türkiye'de 9 bölge bulunmaktadır. Bunlar aşağıda belirtilen bölgelerdir.

- SAMSUN, 17030
- İSTANBUL, 17064
- ANKARA, 17130
- İZMİR, 17220
- İSPARTA, 17240
- DİYARBAKIR, 17281
- ADANA, 17351
- ERZURUM, 17095
- KAYSERİ, 17196

Tablo 1. Radyosonde Özellikleri [12] (Radiosonde Features)

Sıcaklık	Ölçüm Aralığı	-95° C ile +60°C
	Doğruluk	< 0,4°C
	Çözünürlük	0,1°C
	Bildirim Süresi	<0,4s
Nem	Ölçüm Aralığı	1% RH ile 100% RH
	Doğruluk	< 5% RH
	Çözünürlük	0,1 % RH
	Bildirim Süresi	<0,2s
Küresel Konumlandırma Almacı (GPS)	Frekans	1559 MHz ile 1610 MHz
	Bağlanılabilir Uydu Sayısı	72 Kanal
Yükseklik	Ölçüm Aralığı	-500 m ile 50.000 m
	Çözünürlük	0,1 m
Yükseklik Ölçümü	GPS veya Basınç Sensörü (opsiyonel)	
Rüzgâr Yönü	Ölçüm Aralığı	0° ile 360°C
	Çözünürlük	0,01°
Rüzgâr Hızı	Ölçüm Aralığı	0 m/s ile 200 m/s
	Çözünürlük	0,01 m/s
Göndermeç	Frekans Aralığı	400 MHz ile 406 MHz
	Bant Genişliği	1 kHz
	Çıkış Gücü	27 dBm (Ayarlanabilir)
	Tipleri	GFSK/OOK/MOK/FSK
Modülasyon	Baud Rate	9.600 / 4.800
	Mesafe	250 km
	Örnekleme	0,5 s
Batarya	Pil Tipi	2 x CR123
Ağırlık	110 gr	
Süre	Min. 6 Saat	
Boyut	115(E)x80(B)x55(Y) mm	

Tablo 2. Kullanılan veri setinden örnek ham ve normalize edilmiş veriler
(Examples for raw and normalized data from the used dataset)

Öznitelikler	Yükseklik	Sıcaklık	Nem	İşba Sıcaklığı	Rüzgâr Yönü	Rüzgâr Hızı
Ham Veri 1	890	346	7	-62	300	52
Normalize Edilmiş Veri 1	0	1	0,4166	1	0,8328	0,0554
Ham Veri 2	902	331	6	-95	281	95
Normalize Edilmiş Veri 2	0,0004	0,986	0,3333	0,96	0,7799	0,1507
Ham Veri 3	914	323	6	-101	295	70
Normalize Edilmiş Veri 3	0,00088	0,9786	0,3333	0,952727	0,8189	0,0953
Ham Veri 4	925	318	6	-96	290	97
Normalize Edilmiş Veri 4	0,0011	0,9739	0,3333	0,9588	0,8050	0,1552

Bu bölgeler, DMO tarafından Türkiye'yi etkileyen hava kütleleri ve standartlara uygun olarak belirlenmiştir. Ayrıca bu bölgelerin seçilme sebebi Dünya'nın geoit şeklinden dolayı aralarındaki mesafenin belli bir seviyede olmasıdır. Şekil 3'te de görüldüğü gibi Dünya yuvarlak olduğu için anten, anteni görerek haberleşme en yüksek 250 km'dir. Bu yüzden Radyosonde atış bölgeleri arası en yüksek mesafenin 500 km olarak belirlenmesi DMO tarafından şart koşulmuştur.

DMO bu istasyonların günde 4 defa rasat yapmasını tavsiye etmektedir. Fakat ülkemizdeki istasyonlar bütün dünya genelinde uygulandığı gibi, 00 UTC ve 12 UTC olmak üzere günde iki defa rasat yapmaktadır. Rasat yapılan bölgeler Türkiye'de 9 adettir ve bu bölgeler Şekil 4'te görülebilmektedir.

3.3. Kullanılan Yöntemler (Methods)

Radyosonde verileri ile modeli eğitmek için literatürde en çok kullanılan makine öğrenmesi yöntemleri uygulanmıştır. Bu makine

öğrenmesi yöntemlerinin kaotik olan Radyosonde rasatları üzerindeki başarıları incelenmiştir. Bunlar regresyon, güçlendirilmiş ağaç algoritması (Boosted Tree), Destek Vektör Makinesi (DVM) ve Gauss Süreci Regresyonu (GSR) yöntemleridir. Şekil 5'te bu çalışmada uygulanan yöntemlerin blok diyagramı verilmektedir.

3.3.1. Regresyon (Regression)

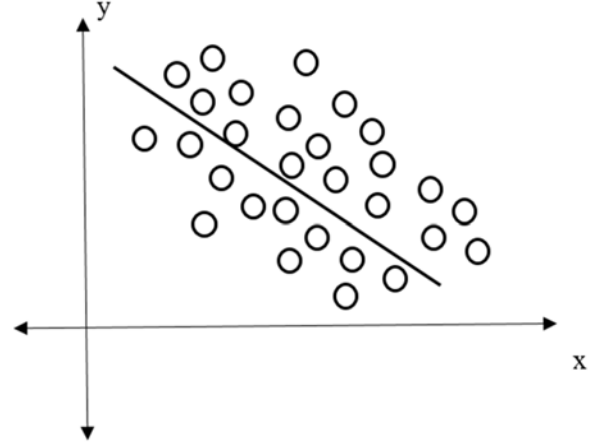
Regresyon, değişkenler arasındaki ilişkiyi tanımlamak için kullanılan istatistiksel bir analiz yöntemidir. İlişki bağımlı ve bağımsız değişkenler arasında tanımlanabilir.

$$Y = f(X, \beta) \quad (1)$$

Eş. 1'de Y: bağımlı bir değişken, X: bağımsız bir değişkendir. B: bilinmeyen bir parametredir. Şekil 6'da regresyon örneği görülmektedir.



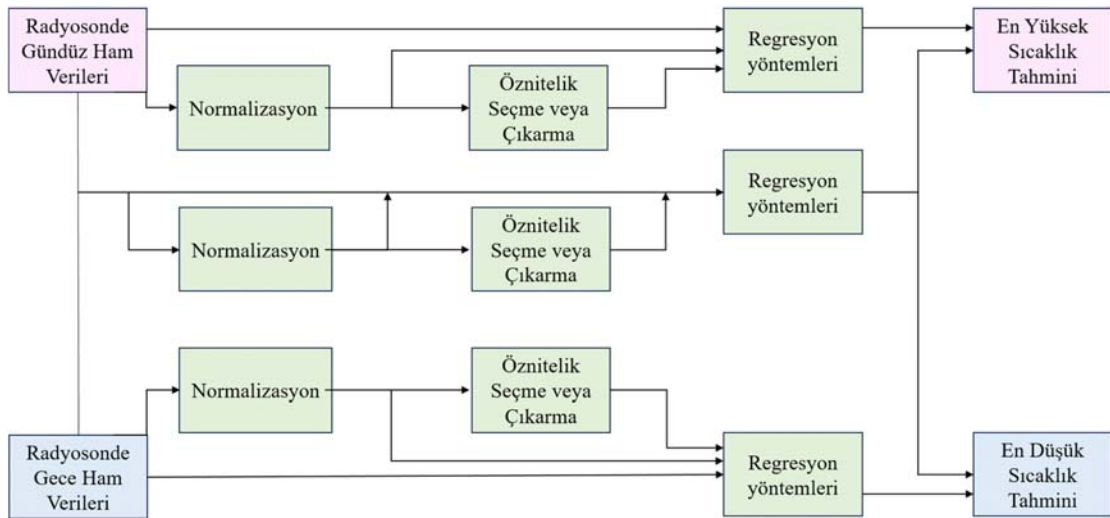
Şekil 3. Dünya'nın Yuvarlaklığının Yer İstasyonu Antenine Olan Etkisi [13]
(The Effect of Earth's Roundness on the Ground Station Antenna)



Şekil 6. Örnek Bir Regresyon Yöntemi
(An Example of a Regression Method)



Şekil 4. MGM Bünyesinde Kurulu Bulunan Radyosonde İstasyonları ve İstasyon Numaraları [14]
(MGM Radiosonde Stations and Station Numbers)



Şekil 5. Radyosonde Verilerine Uygulanan Yöntemlerin Blok Diyagramı (Block Diagram of Methods Applied to Radiosonde Data)

Değişken bağımlılık, tek değişkenli veya çok değişkenli regresyon şeklinde olabilir. Tek değişkenli regresyon Eş. 2'deki gibidir.

$$y = a + bx + \varepsilon \quad (2)$$

Eş. 2'de y: bağımlı bir değişken iken x ise b katsayısının katı olan bağımsız değişkendir, a ise bir sabit sayıdır. Çok değişkenli regresyon ise aynı anda birkaç değişken arasındaki bağımlılığı belirlemek için de kullanılır. Eş. 2 ifadesi çok değişkenli olunca Eş. 3'e döner.

$$y = a + b_1x + b_2x + b_3x + \dots + b_nx + \varepsilon \quad (3)$$

Regresyon modelleri, bağımlı değişkenlerin sonucunu bağımsız değişkenlerden tahmin eder. Karmaşık problemlerin üstesinden gelmek için regresyon analizi kullanılan bir yöntemdir [15].

3.3.2. Karar ağacı algoritması (Decision tree algorithm)

Karar ağaçları algoritmaları genelde veri madenciliğinde kullanılır [16]. Amaç girdi değişkeni ile uygun bir hedef değerini elde eden bir model çıkararak kestirim yapmaktır. Algoritma bir ağaç gibidir ve bölgelerden oluşur bu bölgeler; kök, dallar ve yapraklardır.

Ağacın her yaprağı bir sınıfla veya sınıflar üzerinde bir olasılık dağılımıyla hesaplanır. Bunun için entropi kullanılır. En belirgin özelliği belirlemek için entropi hesaplanır ve hesaplar arasında en yüksek olan bir alt dal olarak seçilir.

$$E = \sum_{i=1}^n p(x_i) \log_2\left(\frac{1}{p(x_i)}\right) - \sum_{i=1}^n p(x_i) \log_2(p(x_i)) \quad (4)$$

Eş. 4'de E: entropiyi, $p(x_i)$: olasılık ağırlık fonksiyonunu belirtir. Karar Ağacı Algoritmaları Gürültüye karşı dayanıklı algoritmalar.

3.3.2.1. Güçlendirilmiş ağaç (Boosted trees)

Güçlendirilmiş ağaç yapısı, makine öğrenmesi alanında yaygın olarak kullanılan bir topluluk (ensemble) öğrenme yöntemidir. Güçlendirilmiş ağaçlar, ağaç tabanlı modellerin bir araya getirilerek daha güçlü bir model oluşturulmasını sağlar. Güçlendirilmiş ağaçlar, genellikle gradyant güçlendime (Gradient Boosting) adı verilen bir tekniği kullanır. Gradyant güçlendirme, zayıf tahmin modellerini (genellikle karar ağaçları) bir araya getirerek güçlü bir tahmin modeli oluşturur. Her bir zayıf model, hatalar üzerine odaklanarak ardışık olarak eğitilir ve hataların azaltılmasına odaklanır [17]. Güçlendirilmiş ağaçların avantajları güçlü tahmin yeteneği, öznelik sıralaması yaparak önemli öznelikleri vurgulaması, dayanıklılık ve aşırı öğrenmeye karşı direncidir. Güçlendirilmiş ağaçlar, sınıflandırma ve regresyon gibi çeşitli makine öğrenmesi problemlerinde etkili bir şekilde kullanılabilir. Güçlendirilmiş ağaçlar ile karar ağaçları arasındaki önemli farklar mevcuttur. Karar ağacı, tek bir ağaç modeli olarak çalışırken, güçlendirilmiş ağaç bir topluluk yöntemidir.

Her iki modelin avantajları ve dezavantajları vardır ve kullanım senaryoları bağlamına bağlı olarak tercih edilebilir. Karar ağacı, basit ve yorumlanabilir bir model istendiğinde tercih edilebilirken, güçlendirilmiş ağaç daha yüksek tahmin gücü gerektiren karmaşık problemlerde tercih edilebilir.

3.3.3. Destek vektör makinesi (Support vector machine)

DVM, Bell Laboratuvarlarında geliştirilmiş bir makine öğrenme algoritmasıdır. İlk denemeler, test örneklerinin ikili sınıflandırmasına ve örüntü tanıma odaklanmıştır. Genel olarak, DVM'ler, giriş

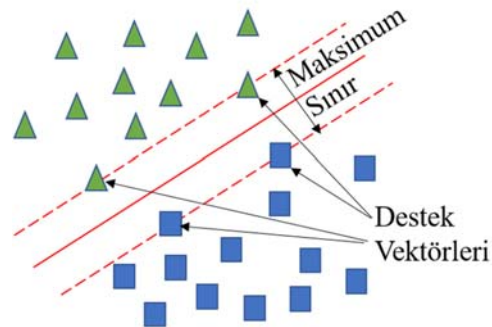
vektörlerini sonsuz boyutlu bir öznelik uzayına eşleyerek bir karar alanı oluşturmak için bir algoritma kullanır. Sonrasında, sonsuz veya yüksek boyutlu özellik uzayında doğrusal bir regresyon yürütülür. Çoğu zaman, çok boyutlu girdi vektörü x ile çıktı y arasındaki ilişki bilinmediği ve çoğunlukla doğrusal olmadığı için bu eşleme işlemi gereklidir. Destek Vektör Makinesi Regresyon (DVR), çok boyutlu girdi vektörlerini çıktı değerlerine uyduran doğrusal bir hiper düzlem bulmayı amaçlar. Sonuç daha sonra bir test setinde yer alan gelecekteki çıktı değerlerini tahmin etmek için kullanılır. X(i) veri noktasının giriş vektörü (i), a(i) gerçek değer ve n veri noktası sayısı ile $P = (x_i, a_i)$, $i = 1, \dots, n$ bir dizi veri noktası tanımlar. Doğrusal fonksiyonlar f için, DVR tarafından oluşturulan hiper düzlem aşağıdaki gibi belirlenir [18]. Eş. 5 tahmin edilen değeri hesaplamak için kullanılır.

$$f(x) = wx + b \quad (5)$$

Eş. 6'da doğrusal bir regresyon modeline benzerlikler gösterir. Tahmin edilen değer, $f(x)$, bir w eğimine ve bir b kesişmesine bağlıdır. Genel olarak, iyi bir genelleme davranışını sürdürürken girdiler ve çıktılar arasındaki ilişkiyi öğrenmek arasında bir denge kurmak istenir. Eğitim hatalarını en aza indirmeye çalışmak aşırı odaklanma ve ezberlemeye neden olabilir. Güven aralığı, öğrenme modelinin karmaşıklığı olarak düşünülebilecek olan DVM'nin Vapnik-Chervonenkis boyutuyla ilgilidir. Bu nedenle, ek eğitim hataları pahasına güven aralığı geliştirilerek iyileştirilmiş genelleme elde edilebilir [19]. Bu takası kontrol etmek için birincil araç önemini açıklayan ceza parametresi C'dir. İyi eğitim ve genelleme davranışları arasındaki denge Eş. 6'da yansıtılır.

$$R = \frac{C}{n} \sum_{i=1}^n L_{\epsilon}(a_i, f(x_i)) + \frac{1}{2} w^2 \quad (6)$$

Burada R, eğitim hatalarından ve model karmaşıklığından kaynaklanan bileşik riski ifade eder. Doğal olarak, R riskinin mümkün olduğu kadar düşük tutulması gerekir. Şekil 7'de görüldüğü üzere DVM'ye ait iki sınıfın birbirinden ayrılmasına örnek gösterilmiştir. DVM her iki sınıfın da bulunduğu en uç noktaları destek vektörü olarak atar. Bu destek vektörleri ile iki sınıf arasındaki uzaklık en büyük olacak şekilde bir hiper düzlem belirlemeye çalışır. Eklenen her yeni veri bu hiper düzleme göre sınıflandırılır.



Şekil 7. DVM Sınıflandırma Örneği (SVM Classification Example)

Verileri doğrusal olmayan bir şekilde ayırmak için ikili bir optimizasyon formu ve Çekirdek Hilesi (Kernel Trick) kullanılmaktadır. Geometrik marjının, karar fonksiyonunun gradyanının normunun tersine eşit olduğundan, fonksiyonel marj, ikinci dereceden fonksiyonun denklemdir. Kernel DVM'nin ikinci dereceden bir optimizasyon ayarına yerleştirildiği kanıtlanmıştır. Bu ayar, ikili formun kullanılmasını veya Çekirdek Hilesi'nin kullanılmasını gerektirmez [20-23].

3.3.4. Gauss süreci regresyonu (Gaussian process regression)

Denetimli öğrenme problemleri sınıflandırma ve regresyon olarak iki ana kategoriye ayrılabilir. Sınıflandırma çıktıları ayrık etiketler iken, regresyon sürekli niceliklerin tahmini ile ilgilidir. Gauss süreci, herhangi bir sonlu alt kümesinin ortak bir Gauss dağılımına sahip olduğu rastgele değişkenlerin bir koleksiyonudur [24]. Gauss süreci, tamamen ortalama ve kovaryans fonksiyonlarıyla belirtilir.

$$m(x) = E[f(x)] \quad (7)$$

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x')))] \quad (8)$$

Eş. 7 ve Eş. 8'de $f(x)$: fonksiyonu $m(x)$: sürecinin ortalama fonksiyonunu, $k(x, x')$: kovaryans fonksiyonunu belirtir. Gauss süreci aşağıdaki şekilde gösterilir:

$$f(X) \sim GP(m(X), k(X, X')) \quad (9)$$

Eş. 9'da GP : gauss süreç fonksiyonu belirtmiştir. Tahmini dağılım, Gauss varsayımına ve Gauss sürecinin sıra dışı özelliğine dayanarak türetilir; bu, ortak dağılımın Gauss olduğunu ima eder.

GSR, bir kovaryans fonksiyonu seçimini gerektirir. Kovaryans fonksiyonu ve bazı parametreleri, öykü hakkındaki ön bilgileri analize yansıtmanın bir yoludur [25].

$$k(X, X') = \sigma_f^2 \exp\left[-\frac{(x-x')^2}{2l^2}\right] \quad (10)$$

Eş. 10'da $k(x, x')$: kovaryans fonksiyonunu, σ : ise varyansı belirtir. Varyansın kare üstel kovaryans fonksiyonudur ve yapısı yakın noktalara büyük önem verir. Böylece düzgün fonksiyon benzer yerel davranışa sahiptir, burada l : karakteristik uzunluk ölçeğini tanımlar.

3.3.5. Temel bileşen analizi (Principal component analysis)

Temel Bileşen Analizi (TBA) veri analizi ve boyut indirgeme yöntemlerinden biridir. TBA, çok boyutlu veri setlerindeki değişkenler arasındaki ilişkileri ve veri setinin içerdiği bilgiyi anlamak için kullanılır [26]. TBA'nın temel amacı, veri setindeki değişkenler arasındaki ilişkileri ifade eden yeni bir değişken seti oluşturmaktır. Bu yeni değişkenler, orijinal veri setindeki değişkenlerin doğrusal bir kombinasyonudur. Bu kombinasyonlar, veri setindeki en fazla değişikliği (varyansı) açıklayan yöne doğru sıralanır. Bu şekilde, veri setinin boyutu azalmaktadır. Ayrıca, yüksek boyutlu veri setlerindeki gürültüyü azaltmak veya veri setini daha kolay yorumlanabilir hale getirmek için de TBA kullanabilmektedir.

TBA'nın çalışma prensibi matematiksel olarak çok boyutlu veri matrisini analiz eder. Temel olarak, veri matrisinin kovaryans matrisini hesaplar. Bu matrisin özdeğerlerini ve özvektörlerini bulur. Özvektörler, orijinal veri setindeki değişkenlere karşılık gelen yeni bileşenlerin yönlendirilmesini sağlar. Özdeğerler, her bileşenin veri setindeki varyansı açıklama yeteneğini gösterir ve bileşenlerin önem sırasını belirler. Boyut indirgemede seçilen varyans veya temel bileşen sayısına göre veri yeni bileşenleri ile temsil edilerek boyut indirgenmiş olmaktadır. Bu çalışmada yeni bileşenlerin %97 varyans ile temsili seçilmiştir. Bu varyans değeri deneysel çalışmalar ile en iyi sonuç veren değer olarak belirlenmiştir.

3.3.6. Karekök ortalama kare hatası (Root mean squared error)

RMSE, bir regresyon modelinin tahminlerinin gerçek değerlerle ne kadar uyumlu olduğunu ölçen bir hata metriğidir. RMSE, tahminlerin gerçek değerlerden ne kadar sapma gösterdiğini ifade eder.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (11)$$

Eş. 11'de her bir veri noktası için gerçek değerler (y_i) ve tahmin edilen değerler (\hat{y}) elde edilmelidir.

RMSE, hata metriği olarak kullanıldığında, daha büyük hataların daha fazla ağırlığa sahip olduğu için büyük hataları vurgular. Dolayısıyla, RMSE'nin değeri ne kadar düşük olursa, modelin tahminleri gerçek değerlere o kadar yakındır ve modelin daha iyi performans gösterdiği söylenebilir. RMSE, genellikle regresyon modellerinin performansını değerlendirmek ve farklı modeller arasında karşılaştırma yapmak için kullanılır.

3.3.7. Ortalama kare hatası (Mean squared error)

Ortalama Kare Hatası (Mean Squared Error, MSE), regresyon modellerinin performansını değerlendirmek için yaygın olarak kullanılan bir diğer hata metriğidir. MSE, tahminlerin gerçek değerlerle ne kadar uyumlu olduğunu ölçer ve tahminlerin gerçek değerlerden ne kadar sapma gösterdiğini hesaplar.

$$MSE = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n} \quad (12)$$

Eş. 12'de her bir veri noktası için gerçek değerler (y_i) ve tahmin edilen değerler (\hat{y}_i) ile ifade edilmektedir. MSE'nin değeri, hata karelerinin ortalama büyüklüğünü gösterir. MSE'nin daha küçük olması, modelin tahminlerinin gerçek değerlere daha yakın olduğunu ve daha iyi bir model performansını gösterdiğini gösterir. MSE'nin değeri sıfıra yaklaştıkça, modelin tahminleri gerçek değerlere tam olarak uyar.

3.3.8. Ortalama mutlak hata (Mean absolute error)

Ortalama Mutlak Hata (Mean Absolute Error, MAE) regresyon modellerinin performansını değerlendirmek için kullanılan hata metriklerinden biridir. MAE, tahminlerin gerçek değerlerle ne kadar uyumlu olduğunu ölçer ve tahminlerin gerçek değerlerden ne kadar sapma gösterdiğini hesaplar.

$$MAE = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{n} \quad (13)$$

Eş. 13'de (y_i) gerçek değerleri ve (\hat{y}_i) tahmin edilen değerleri göstermektedir. Burada, n veri noktalarının sayısını temsil eder. MAE'nin değeri, hataların ortalama büyüklüğünü gösterir. MAE'nin daha küçük olması, modelin tahminlerinin gerçek değerlere daha yakın olduğunu ve daha iyi bir model performansını gösterdiğini gösterir. MAE, regresyon modellerinin performansını değerlendirmek için sıklıkla kullanılır ve hataların büyüklüğüne doğrudan odaklanır. MAE, ölçeklenebilir bir metrik olduğu için farklı ölçeklere sahip veri setlerinde karşılaştırma yapmak için de uygundur. Ancak, hataların büyüklüklerini yansıttığı için aykırı değerlere daha duyarlı olabilir.

4. Sonuçlar ve Tartışma (Results and Discussion)

Bu çalışmada kullanılan veriler 17130 numaralı Ankara bölgesinde bulunan istasyon tarafından ölçülen 3 yıllık (2016-2017-2018) gündüz ve gece atılan rasat verilerinden oluşmaktadır. Bu veriler 40 km irtifaya kadar toplanan basınç seviyesi, yükseklik, sıcaklık, nem, işba sıcaklığı, rüzgâr yönü ve rüzgâr hızından oluşmaktadır. Ancak gerçekleştirilen rasatlarda, rasat balonu farklı irtifalarda atmosferik etkilerden dolayı patladığı için bütün rasatlar farklı irtifalarda sonlanmaktadır. Bu sebepten dolayı 95 mbar'a kadar gerçekleştirilen ölçümler hava tahmininde kullanılmıştır. 95 mbar'ın altında kalan veriler veri setine dâhil edilmemiştir. Ayrıca sinyal karıştırıcı ve farklı sebeplerden dolayı kullanıma uygun olmayan rasatlar da veri setine dâhil edilmemiştir.

Veri seti, 3 yıllık ölçüm sonucunda elde edilen 944 ölçümden oluşmaktadır. Bu veri setini eğitim ve test olarak ayırmak için çapraz doğrulama uygulanmıştır. K çaprazlama doğrulamadaki k değeri 5, 10 ve 15 olarak üç farklı değer seçilip bu üç değerde eğitimler yapılmıştır. Gece ve gündüz sıcaklık tahminleri için iki ayrı model oluşturulmuştur. En düşük sıcaklık tahmini için 00.00 UTC de toplanan veriler kullanılırken, En yüksek sıcaklık tahmini için 12.00 UTC'de toplanan veriler kullanılmıştır. Yöntemlerin başarısını analiz etmek için farklı eğitim setleri kullanarak başarıya etkisi analiz edilmiştir.

Veri setinde yükseklik ölçümleri ve her yükseklikte ölçülen sıcaklık ölçümleri, nem ölçümleri, işba sıcaklık ölçümleri, rüzgâr yönü ölçümleri, rüzgâr hızı ölçümleri bulunmaktadır.

Bu makale kapsamında, başlangıç olarak eğitimlerde üç yıllık gündüz ve gece toplanmış rasat verileri kullanılarak gün içerisindeki en yüksek ve en düşük sıcaklık kestirimi gerçekleştirilmiştir. Ardından rasat verilerine öz nitelik çıkarımı yöntemi uygulanmıştır. Daha sonra veri seti yerden farklı irtifalara kadar ayrıştırılarak eğitimler gerçekleştirilmiştir. Son olarak veri seti gündüz ve gece rasatları olarak ayrıştırılmış ve gündüz veri seti kullanılarak en yüksek hava sıcaklığı, gece veri seti kullanılarak ise en düşük hava sıcaklığı tahmin edilmiştir.

Her eğitimde 16 farklı veri ile makine öğrenme algoritması K=5, K=10 ve K=15 değerleri için ayrı ayrı eğitilmiştir. Aynı zamanda veriye hem TBA uygulanarak hem de uygulanmadan bütün modeller

eğitilmiştir. Eğitimlerden elde edilen sonuçların özeti Tablo 3'de verilmektedir.

Bu çalışmada 4 farklı makine öğrenmesi yöntemi kullanılmış olup, en başarılı sonuçlar Tablo 3'de verilmiştir. Tüm eğitim modellerinde farklı makine öğrenmesi yöntemlerinin sonuçları ise rmse değeri açısından karşılaştırılmış olup Tablo 4'de özetlenmiştir. Eğitim 5, 6 ve 7'de Meteoroloji Genel Müdürlüğünden elde edilen önemli yükseklik bilgilerine göre veri seti 3 farklı yükseklik için hazırlanmıştır [27]. Belirli yüksekliğe kadar verilerin etkileri incelenmiştir. Farklı yükseklikler için elde edilen başarı metrikleri incelendiğinde yükseklik arttıkça başarı değerlerinin de tüm metrikler için arttığı gözlenmektedir. Yüksekliğin artması, o yüksekliğe kadar elde edilen verilerin de artmasına neden olmaktadır. 600 mBar seviyesine kadar elde edilen veri boyutuna göre 400mBar seviyesinde elde edilen veri boyutu daha yüksektir. Buradan makine öğrenmesi yöntemi ile daha fazla veriyle daha yüksek başarılar elde edildiği sonucu çıkartılabilir. Aynı durum Tablo 4'deki Eğitim 5, 6 ve 7 incelendiğinde de gözlenmektedir. Belli bir yüksekliğe kadar elde edilen sonuçlar, tüm yükseklik dikkate alınarak elde edilen sonuçlara göre daha düşük olduğu belirlenmiştir. Bu nedenle bu çalışmadaki Radyosonde verilerinde, tüm yüksekliklerde elde edilen verilerin makine öğrenmesi yöntemlerinde kullanımı önerilmektedir.

Bu çalışma daha önce yapılmış 4 farklı çalışma ile Tablo 5'deki gibi karşılaştırılmıştır. Tablo 5'de önceden gerçekleştirilmiş literatürde bulunan 4 farklı çalışma ile bu çalışmalar süresince oluşan hava durumu kestirim başarı oranları görülmektedir.

Tablo 3. Farklı Eğitim Modellerinin Başarı Sonuçları (K=15) (Success Results of Different Training Models)

	Veriler	Parametreler	Hangi Yöntem	Başarı (RMSE)	Başarı (MSE)	Başarı (MAE)
Eğitim 1	3 yıllık gündüz ve gece rasat verileri. Veriler normalize edilmiştir.	TBA yok.	GSR	3,5101	12,321	2,6833
Eğitim 2	3 yıllık gündüz ve gece rasat verileri. Veriler normalize edilmiştir.	TBA var (varyans % 97)	GSR	5,5601	30,914	4,3935
Eğitim 3	3 yıllık gündüz ve gece rasat verileri. Veriler normalize edilmemiştir.	TBA yok.	Ensemble (Boosted Trees)	2,7496	7,5601	2,0921
Eğitim 4	3 yıllık gündüz ve gece rasat verileri. Veriler normalize edilmemiştir.	TBA var (varyans % 97)	GSR	4,5963	21,126	3,7225
Eğitim 5	3 yıllık gündüz ve gece rasat verileri kullanılmıştır. Veriler normalize edilmiştir. Bu rasat verileri yer seviyesinden ilk 400mbar kadar olan bölge alınmıştır.	TBA yok.	GSR	4,2084	17,711	3,3011
Eğitim 6	3 yıllık gündüz ve gece rasat verileri kullanılmıştır. Veriler normalize edilmiştir. Bu rasat verileri yer seviyesinden ilk 500mbar kadar olan bölge alınmıştır.	TBA yok.	GSR	4,6212	21,356	3,6409
Eğitim 7	3 yıllık gündüz ve gece rasat verileri kullanılmıştır. Veriler normalize edilmiştir. Bu rasat verileri yer seviyesinden ilk 600mbar kadar olan bölge alınmıştır.	TBA yok.	GSR	5,1316	26,333	4,0355
Eğitim 8	3 yıllık gündüz verisi ile en yüksek sıcaklık kestirimi yapılmıştır. Veriler normalize edilmiştir.	TBA yok.	GSR	2,6187	6,8576	2,0455
Eğitim 9	3 yıllık gündüz verisi ile en yüksek sıcaklık kestirimi yapılmıştır. Veriler normalize edilmemiştir.	TBA yok.	GSR	1,2893	1,6623	0,9145
Eğitim 10	3 yıllık gece verisi ile en düşük sıcaklık kestirimi yapılmıştır. Veriler normalize edilmiştir.	TBA yok.	GSR	3,6580	13,381	2,9322
Eğitim 11	3 yıllık gece verisi ile en düşük sıcaklık kestirimi yapılmıştır. Veriler normalize edilmemiştir.	TBA yok.	GSR	2,4801	6,1509	1,9758

Tablo 4. Uygulanan Makine Öğrenmesi Yöntemleri için Elde Edilen Başarım Sonuçları (rmse)
(Performance Results (rmse) for All Machine Learning Methods)

Eğitim No												
	1	2	3	4	5	6	7	8	9	10	11	
Algoritma Adı												
Doğrusal Regresyon	64,17	5,88	40,29	4,64	116,92	155,38	189,87	37,99	9,35	30,69	20,16	
DVM (Doğrusal)	4,53	6,03	3,29	4,72	5,48	5,99	6,46	3,7	1,41	4,2	2,64	
DVM (Küadratik)	4,01	5,67	5,73	5,14	4,86	5,17	5,75	3,35	1,66	3,9	2,95	
DVM (Kübik)	5,1	5,77	6,82	5,68	6,01	7,47	8,31	3,39	1,89	5,74	2,89	
DVM (Fine Gauss)	10,28	5,98	10,06	4,75	10,19	10,17	9,96	10,1	9,94	7,17	7,13	
DVM (Medium Gauss)	4,25	8,74	3,95	8,44	4,99	5,35	5,6	3,81	2,14	3,99	2,71	
DVM (Coarse Gauss)	4,76	10,3	4,14	10,27	5,73	6,08	6,14	4,43	1,84	4,71	2,62	
Topluluk (Güçlendirilmiş Ağaç)	4,62	7,6	2,75	6,14	5,13	5,34	5,39	4,2	1,56	4,28	2,49	
Topluluk (Torbalı(Bagged) Ağaç)	4,69	7,86	2,77	6,1934	5,31	5,4	5,37	4,34	1,37	4,35	2,54	
GSR(Karesel Üstel GPR)	10,28	5,56	10,28	4,6	10,26	10,25	10,25	10,11	10,12	7,18	7,18	
GSR (Üstel)	8,85	8,97	5,79	5,72	6,86	6,34	5,61	7,87	4,93	6,78	5,87	
GSR (Rasyonel Küadratik)	3,51	5,56	3,14	4,6	4,21	4,62	5,13	2,62	1,29	3,66	2,48	

Tablo 5. Hava Durumu Kestirimi Literatür Karşılaştırması(Literature Comparison of Weather Forecasting)

Çalışma	Veri Büyüklüğü	Ölçüm Metodu	Yöntem	Başarı Oranı (%)	RMSE
Nandar [4]	6 yıllık	Yer Gözlem İstasyonu	DVM	-	3
Rahul [5]	10 yıllık	Yer Gözlem İstasyonu	LSTM	84	
Anbarasi [28]	1 yıllık	Yer Gözlem İstasyonu	LSTM	89	
Wang [29]	1 gün	Yer Gözlem İstasyonu	Büyük Veri Kapsül Ağlar (MICaps)		5
Bu Çalışma	3 yıllık	Radyosonde	GSR		1.2

Nandar [4] çalışmasında, Mynmar'daki yağmur oranlarını tahmin etmeye çalışmıştır. Kullandığı veri 1990'dan 2006'ya kadarki hava durumu verileridir. Bu çalışmada hava sıcaklık tahmini yerine Mynmar'da oluşan selleri kestirmek için yağmur yağma miktarını hesaplamaya çalışılmıştır. DVM algoritmasını kullanmış ve RMSE %3 hata ile hesaplayabilmiştir.

Rahul [5] çalışmasında, Yer Gözlem İstasyonlarından topladığı verileri sırası ile sıcaklık, nem, çiy noktası, basınç, görünürlük ve yağış verilerini kullanarak eğitimlerini yapmıştır. Veri seti 2007'den 20017'e kadar olan verilerdir. Çalışmasında ulaştığı başarı oranı %84'tür.

Anbarasi [28] çalışmasında, büyük veri yöntemi ile hava durumu kestirimi yapmaya çalışmıştır. Kestirmeye çalıştığı veri denizciler için denizlerdeki hava durumudur. Büyük veriyle eğitim yapmak için 3B Okyanus verisi kullanılmış. Bu veri ile deniz yüzeyindeki sıcaklık verisini tahmin etmeye çalışmıştır. Çalışmalarında ulaştığı hata oranı %11'dir. Bu çalışmanın asıl amacı büyük verilerde daha hızlı tahmin yapmayı sağlamaktır.

Wang [29] çalışmasında, kendi yaptığı hava durumu veri toplama sistemi ile topladığı sıcaklık, nem, rüzgâr hızı, rüzgâr yönü verilerini kullanarak kendini eğitip bir sonraki gün için tahmin yürütmüştür. RMSE %5 olarak elde edilmiştir.

Gerçekleştirilen bu çalışmada Nandar [4], Rahul [5], Anbarasi [28] ve Wang'dan [29] farklı olarak Radyosonde rasatlarından elde edilen 3-boyutlu 3 yıllık veriler kullanılarak makine öğrenme algoritmaları ile hava durumu kestirimi gerçekleştirilmiştir. Bu sayede daha önce literatürde gerçekleştirilen çalışmalardan daha düşük hata oranı ile hava durumu kestirimi gerçekleştirilmiştir. Literatürde belirtilen çalışmalar RMSE başarı kriterini veya yüzdelik doğruluk değerini belirttiği için Tablo 5'deki karşılaştırma da buna göre yapılmıştır. Diğer başarı ölçütlerini de kullanarak yapılacak bir karşılaştırmanın daha objektif değerlendirme imkânı sunacağı düşünülmektedir. Bu nedenle bu çalışmada Tablo 3'de sunulan farklı başarı ölçütlerine de yer verilerek gelecek çalışmalar için kıyaslama imkânı sunulmuştur.

Bu araştırma, diğer literatür çalışmalarından farklı olarak, 3 yıl boyunca kaydedilen Radyosonde gözlemlerinden elde edilen verilerle oluşturulan üç boyutlu atmosfer modelini kullanarak Makine Öğrenimi Algoritmaları eğitilmiştir. Bu eğitilen modeller kullanılarak bir sonraki gün için en yüksek ve en düşük sıcaklık tahminleri yapılmıştır. Tahmin algoritmalarının performansı değerlendirildiğinde, bu veri yapısına en uygun regresyon yöntemi olarak Gauss Süreci Regresyon yöntemi belirlenmiştir. Ayrıca, Radyosonde verilerinin normalleştirilmesinin etkisi incelenmiş ve normalleştirme işleminin bu veri kümesinde başarıyı azalttığı gözlenmiştir. Bu durum, bazı özneliklerin diğerlerine göre daha önemli olduğunu ortaya koymaktadır. Basınç değeri, diğer özneliklere kıyasla daha yüksek bir değere sahip olduğundan, sıcaklık tahmininde daha etkili olduğu gözlenmiştir. Ayrıca, en yüksek ve en düşük sıcaklık tahmini için ayrı modellerin kullanılmasının başarıyı artırdığı tespit edilmiştir. Yapılan tüm analizler sonucunda, önerilen yöntemle en yüksek sıcaklık tahmini için ortalama karekök sapma değeri 1,2 ve en düşük sıcaklık tahmini için ortalama karekök sapma değeri 2,4 olarak elde edilmiştir. Gündüz sıcaklık tahminindeki hata oranı ile gece sıcaklık tahminindeki hata oranı arasındaki fark, gündüz rasat veri sayısının gece rasat veri sayısına göre daha fazla olmasından kaynaklandığı düşünülmektedir.

5. Sonuçlar (Conclusions)

Dünya genelinde hava tahminlerinin yapılıp, bu verilere göre önlemler alınması çok önemlidir. Bu verilerin ölçülmemesi veya yanlış ölçülmeleri, ülke ekonomilerine çok büyük zararlar verebilmektedir. Bu çalışmada kullanılan Radyosonde Meteoroloji Ölçüm cihazı, içerisinde sıcaklık, basınç, nem ve GPS sensörleri barındırır. Üzerindeki sensörleri kullanarak o anlık sıcaklık, yüzde bağıl nem, rüzgâr hızı ve rüzgâr yönü değerlerini ölçerek, üzerinde bulunan anteni ve göndermeç devresi vasıtasıyla yer istasyonuna gönderir. Yer İstasyonuna iletilen meteorolojik veriler işlenerek, DMO veri formatları olan TEMP35 ve BUFR kod gibi formatlara çevrilir. Ardından elde edilen rasatlar DMO'ya gönderilerek yaklaşık atmosferik model oluşturulur. Bu model kullanılarak bu çalışma kapsamında Makine Öğrenmesi Algoritmaları tabanlı kestirimler

kullanılmıştır. Literatürde daha önce Makine Öğrenmesi Algoritmaları tabanlı kestirimler bulunuyor olsa da gerçekleştirilen çalışmalar yer gözlem istasyonlarından elde edilmiş tek boyutlu ve sabit irtifaya ait ölçümlere dayanmaktadır. Bu çalışmada literatürdeki diğer çalışmalardan farklı olarak 3 yıllık Radyosonde rasatları kullanılarak oluşturulan 3-boyutlu atmosfer modeli ile Makine Öğrenmesi Algoritmaları eğitilmiş ve eğitilen bu modeller ile ertesi güne ait en yüksek ve en düşük sıcaklık kestirilmeye çalışılmıştır. Kestirim algoritmalarının başarıları incelenmiş içlerinden bu veri yapısına en uygun regresyon yöntemi olarak Gauss Süreci Regresyon belirlenmiştir. Ayrıca Radyosonde verilerine, normalizasyonun etkisi incelenmiş olup normalize edilmesinin başarıyı düşürdüğü gözlenmiştir. Bu verideki bazı özneliklerin daha önemli olduğu sonucunu ortaya çıkarmaktadır. Basınç değeri diğer öz niteliklere göre yüksek değerde olup sıcaklık tahmininde daha etkin olduğu gözlemlenmektedir. Bunların yanı sıra en yüksek sıcaklık ve en düşük sıcaklık tahmini için ayrı modellerin kullanılmasının da başarıyı arttırdığı tespit edilmiştir. Tüm analizler sonucunda önerilen yöntem ile 1,2 Ortalama Karekök Sapması değeriyle ertesi güne ait en yüksek sıcaklık 2,4 Ortalama Karekök Sapması değeri ile en düşük sıcaklık için kestirilmiştir. Gündüz sıcaklık tahminindeki hata oranı ile gece sıcaklıktaki hata oranı arasındaki farkın nedeni gündüz rasat veri sayısının gece rasat veri sayısına göre fazla olmasından kaynaklanmaktadır.

Gelecek Çalışmalar (Future Works)

İlerleyen çalışmalarda daha fazla gün içeren Radyosonde verileri kullanılarak derin öğrenme algoritmalarıyla kestirim yapılarak hava tahmini hata oranı azaltılabilir. Ayrıca, bu çalışmada Ankara ilinde gerçekleştirilen 3 yıllık Radyosonde rasatları kullanılmıştır. İlerleyen çalışmalarda tüm ülke genelinde veya dünya genelindeki geçmişe yönelik bütün Radyosonde rasatları birleştirilerek eğitimler gerçekleştirilebilir. Böylelikle hava tahminindeki hata oranı daha da azaltılabilir.

Kaynak (References)

- Sobrevilla K. L. M. D., Quiñones A. G., Lopez K. V. S., Azaña V. T., Daily weather forecast in Tiwi, Albay, Philippines using Artificial Neural Network with missing values imputation, IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON, Singapore, 2981–2985, 22-25 Kasım, 2016.
- Luis F., Moncayo G., Meteorology Today: an introduction to weather, climate, and the environment, Cengage Learning, Boston, Massachusetts, ABD, 2014.
- Rahul G. K., Singh S., Dubey S., Weather forecasting using Artificial Neural Networks, ICRITO- IEEE 8th Int. Conf. Reliab. Infocom Technol. Optim. (Trends Futur. Dir.), Noida, India, 21–26, 04-05 Haziran, 2020.
- Fente D. N., Singh D. K., Weather forecasting using Artificial Neural Network, Proc. Int. Conf. Inven. Commun. Comput. Technol. (ICICCT 2018), Coimbatore, India, 1757–1761, 20-21 Nisan, 2018.
- Nandar A., Bayesian network probability model for weather prediction, Proc. 2009 Int. Conf. Trends Inf. Technol. (CTIT 2009), Dubai, United Arab Emirates, 120–124, 15-16 Aralık, 2009.
- Abhishek K., Singh M.P., Ghosh S., Anand A., Weather forecasting model using Artificial Neural Networks, Procedia Technology, 4, 311-318, 2012.
- Vaisala. Vaisala Radiosondes. <https://www.vaisala.com/en/products/weather-environmental-sensors/upper-air-radiosondes-rs41-rs41-e-models> Güncellenme tarihi Ocak, 2024. Erişim tarihi Şubat 27, 2024.
- Meteomodem. Meteomodem Radiosonde M10. <https://www.meteomodem.com/m10> Güncellenme tarihi 31 Mart, 2023. Erişim tarihi Şubat 27, 2024.
- Meisei Electric. iMS-100 GPS Radiosonde. <https://www.meisei.co.jp/english/products/meteorology/upper-air/p578>. Güncellenme tarihi Ocak, 2024. Erişim tarihi Şubat 27, 2024.
- Graw. Graw Radiosondes DFM-17. <https://www.graw.de/products/radiosondes/dfm-17/>. Güncellenme tarihi Ocak, 2024. Erişim tarihi Şubat 27, 2024.
- Yağan Y., Ravinsonde Rasatları Kurs Notları, 2007. <https://silo.tips/download/ravinsonde-rasatları-kurs-notları>. Güncellenme tarihi Ocak, 2017. Erişim tarihi Şubat 27, 2024.
- Magnetron Savunma ve Uzay A.Ş. <https://magnetrontr.com/>. Güncellenme tarihi Ocak, 2021. Erişim tarihi Şubat 27, 2024.
- Haupt S. E., Cowie J., Linden S., McCandless T., Kosovic B., Alessandrini S., Machine learning for applied weather prediction, Proceeding of IEEE 14th Int. Conf. eScience (e-Science 2018), Amsterdam, Hollanda, 276–277, 29 Ekim-01 Kasım, 2018.
- Meteoroloji Genel Müdürlüğü. Sıkça Sorulan Sorular. <https://www.mgm.gov.tr/genel/sss.aspx?s=havatahmini>. Güncellenme tarihi Ocak, 2024. Erişim tarihi Şubat 27, 2024.
- Vinet L., Zhedanov A., A ‘missing’ family of classical orthogonal polynomials, J. Phys. A Math. Theor., 44 (8), 085201, 2011.
- Gülmez B., Kulluk S., Analysis and price prediction of secondhand vehicles in Türkiye with big data and machine learning techniques, Journal of the Faculty of Engineering and Architecture of Gazi University, 38 (4), 2279-2289, 2023.
- Yazıcı G., Performance evaluation of preprocessing to pca combined machine learning techniques on pharmaceutical and mineral samples by laser-induced breakdown spectroscopy, Yüksek Lisans Tezi, Atılım Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 2023.
- Kavitha S., Varuna S., Ramya R., A comparative analysis on linear regression and support vector regression, Proceeding of 2016 Online Int. Conf. Green Eng. Technol., Coimbatore, 1-5, 19 Kasım, 2016.
- Yang H., Chan L., King I., Support vector machine regression for volatile stock market prediction, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2412, 391–396, 2002.
- Cao L. J., Tay F. E. H., Support vector machine with adaptive parameters in financial time series forecasting, IEEE Trans. Neural Networks, 14 (6), 1506–1518, 2003.
- Jain U., Nathani K., Ruban N., Raj A. N. J., Zhuang Z., Mahesh V. G. V., Cubic SVM classifier based feature extraction and emotion detection from speech signals, Proc. - 2018 Int. Conf. Sens. Networks Signal Process., Xi'an, China, 386–391, 28-31 Ekim, 2018.
- Virdi P., Narayan Y., Kumari P., Mathew L., Discrete Wavelet Packet based Elbow Movement classification using Fine Gaussian SVM, 1st IEEE Int. Conf. Power Electron. Intell. Control Energy Syst. ICPEICES 2016, Delhi, India, 1-5, 4-6 Temmuz, 2016.
- Yan K. W., Study on the forecast of air passenger flow based on SVM regression algorithm, 2009 First International Workshop on Database Technology and Applications, Wuhan, China, 325–328, 25-26 Nisan, 2009.
- Asheri H., Bayati A., Rabiee H.R., Motion Vector Recovery With Gaussian Process Regression, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 953–956, 22-27 Mayıs, 2011.
- Rasmussen C.E., Williams C.K.I., Gaussian Processes for Machine Learning, MIT Press, Massachusetts, A.B.D., 2006.
- Wozniak M., Experiments with boosted decision tree classifiers, 2008 Eighth International Conference on Intelligent Systems Design and Applications, Kaohsiung, Taiwan, 552–557, 26-28 Kasım, 2008.
- Ackerman S.A., Knox J.A., Meteorology: Understanding the Atmosphere, Jones & Bartlett Learning, Burlington, Massachusetts, ABD, 244–270, 2015.
- Anbarasi J. D., Radha V., Review on Marine Weather Forecasting with Big data, International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 783–790, 9-11 Mayıs, 2022.
- Wang Y., Ni K., Wang X., Zhu J., Design of Automatic Weather Monitoring and Forecasting System based on Internet of Things and Big Data, International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 979–982, 7-9 Nisan, 2022.