

Güneş Aktivitesi Tahmini Üzerine Makine Öğrenmesi Uygulamaları

Mirkan Yusuf Kalkan¹  , Dıaa E. Fawzy² , Metin Yavuz³ 

¹ Ondokuz Mayıs Üniversitesi, Lisansüstü Eğitim Enstitüsü, Fizik Anabilim Dalı, Samsun 55200, Türkiye

² İzmir Ekonomi Üniversitesi, Mühendislik Fakültesi, Havacılık ve Uzay Mühendisliği Bölümü, İzmir 35330, Türkiye

³ Ondokuz Mayıs Üniversitesi, Fen Fakültesi, Fizik Bölümü, Samsun 55200, Türkiye

Accepted: May 12, 2025. Revised: May 12, 2025. Received: November 30, 2024.

Özet

Son zamanlarda makine öğrenmesi (ML) birçok alanda uygulanmaya başlanmıştır. Bu yöntemle, problemlere karşı daha hızlı ve daha kapsamlı çözümler üretilmesi amaçlanmıştır. Güneş Aktivitesinin (GA) 1700 yılından beri çevrimsel değişiminin takibiyle, farklı tahmin senaryoları oluşturabilmek için çeşitli ML algoritmaları kullanılmaktadır. Bu çalışmada, çeşitli ML algoritmalarıyla 25. Güneş Çevrimi'nin tahmin edilmesi ve literatürdeki diğer çalışmalarla birlikte, GA'nın ML uygulamalarıyla tahmin çalışmalarına yönelik farkındalık yaratılması amaçlanmıştır. Güneş Aktivitesinin en önemli göstergelerinden biri olan Güneş Lekesi Sayısı (SSN) temel alınarak, NAR (Doğrusal Olmayan Otoregresif) modeli ile 25. Güneş Çevrimi'nin aylık değerlerinin tahmini gerçekleştirilmiştir. Bu tahminler, ML algoritmalarından Destek Vektör Makinesi (SVM), Rastgele Orman (RF), k-En Yakın Komşular (kNN) ve Gradyan Arttırma (GB) kullanılarak yapılmıştır. Sonuçlar, Ortalama Kareköt Hata (RMSE), Ortalama Mutlak Hata (MAE), Pearson Korelasyon Katsayısı (PCC) ve Nash-Sutcliffe Etkililik Katsayısı (NSE) ile değerlendirilerek, çevrim değerlerinin tahmin sonuçlarıyla karşılaştırılmıştır. Oluşturulan modeller, ilk adımda NAR ve sonraki adımlarda SVM, kNN, GB olmak üzere sırasıyla NAR-SVM, NAR-RF, NAR-kNN ve NAR-GB şeklinde adlandırılmıştır. Modellerin tahmin ettiği maksimum değerler 117.43-117.99 aralığında değişmektedir (NAR ile maksimum değer 118.13 olarak üretilmiştir). RMSE değerleri 0.08–0.72, MAE değerleri 0.07–0.47 arasındadır. PCC değerleri 0.99 ve NSE değerleri de 0.99 olarak bulunmuştur. Sonuçlar, literatürdeki diğer çalışmaların sonuçlarıyla karşılaştırılmıştır.

Abstract

Recently, machine learning (ML) has been applied in many areas. With this method, it is aimed to produce faster and more comprehensive solutions to problems. Various ML algorithms are used to create different prediction scenarios by following the cyclical changes of Solar Activity (SA) since 1700. In this study, it is aimed to predict the 25th Solar Cycle with various ML algorithms and to create awareness about prediction studies with ML applications of SA together with other studies in the literature. Based on Sunspot Number (SSN), one of the most important indicators of Solar Activity, the monthly values of the 25th Solar Cycle were estimated with the NAR (Nonlinear Autoregressive) Model. These estimates were made using the ML algorithms Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbors (kNN) and Gradient Boosting (GB). The results were evaluated with Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Pearson Correlation Coefficient (PCC) and Nash-Sutcliffe Efficiency Coefficient (NSE) and compared with the predicted results of the conversion values. The created models were named as NAR in the first step and SVM, kNN, GB in the following steps, respectively as NAR-SVM, NAR-RF, NAR-kNN and NAR-GB. The maximum values predicted by the models range from 117.43 to 117.99 (the maximum value was produced as 118.13 with NAR). RMSE values range from 0.08 to 0.72, MAE values range from 0.07 to 0.47. PCC values were found as 0.99 and NSE values were found as 0.99. The results were compared with the results of other studies in the literature.

Anahtar Kelimeler: solar cycle – time series prediction – machine learning

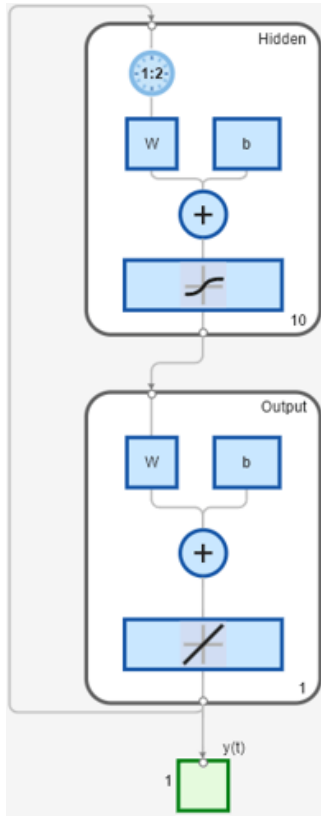
1 Güneş Aktivitesinin (GA) Çevrimsel Tahmini

İlk defa Carrington (1859) ve Hodgson (1859) tarafından gözlemlenen X45 sınıfı Güneş Parlamasının (Cliver ve Dietrich 2013) dikkat çekmesi sonucu, günümüze doğru GA'nın bileşenleri gözlemler geliştikçe sayıca artmış ve bu bileşenler takip edildiğinde çevrimsel periyodik değişimler tespit edilmiştir. 9-11 yıl arasında bir periyoda sahip Güneş Çevrimleri'nin, hem Dünya'nın manyetik alanındaki (jeomanyetizma) (Pesnell

2020) değişimler hem de Dünya iklimindeki (Haigh 2007) değişimlerle ilişkili olduğu saptanmıştır. Bu sayede, Güneş Çevrimlerinin devamlı incelenmesi ve tahmin edilmesi, Dünya'yı manyetik ve iklimsel açıdan gerçekleştirebilecek etkilerine karşı tedbir alınmasında önemli bir rol oynamaktadır.

GA'ndeki çevrimsel değişimlerinin en iyi temsilcilerinden biri olan SSN, 17. yüzyıldan beri kayıt altına alınmaktadır; ancak Güneş Çevrimleri'nin miladı resmi olarak 1755 yılı belirlenmiştir (Kane 2002). Ocak 2020'den itibaren Güneş Çevrimlerinden 25.si gerçekleştirilmektedir ve günümüzde çevrimlerin sadece kaydı tutulmayıp, çeşitli tahmin yöntemleri geliştirilmiştir ve bu yöntemler birçok çalışmada kullanılmıştır

* mirkanyusufkalkan@gmail.com



Şekil 1. MATLAB yazılımından varsayılan ayarlarda oluşturulmuş NAR sinir ağı örneği. 1:2; Gecikme kutusu, w ağırlıklar, b gürültü, gizli ve çıkış katmanlarının alt bölümündeki blok aktivasyon fonksiyonu, gizli ve çıkış katmanları çerçevelerinin sağ alt köşesindeki sayılar katman sayısını ve $y(t)$ tahmin sonuçları bloğunu ifade eder.

(Du 2020; Svalgaard 2020; Wu ve Qin 2021; Kalkan ve diğ. 2022, 2023; Penza ve diğ. 2021; Aparicio ve diğ. 2023; Cao ve diğ. 2024).

GA'nin çevrimsel tahmin yöntemleri üç ana başlık altındadır: öncü, model temelli ve ekstrapolasyon (Petrovay 2020). Bu çalışmada model temelli yöntemlerden ML yöntemi kullanılmıştır. Bir sonraki bölümde ML yöntemleri ele alınmıştır.

2 Makine Öğrenmesi (ML) Uygulamaları

2.1 Doğrusal Olmayan Otoregresif (NAR) Sinir Ağları

NAR, ağıın birkaç katmanını içeren geri besleme (tahmin hatasını minimum değere yaklaştırmak için kullanılan yöntem) (Munro 2010) bağlantılarına sahip dinamik bir tekrarlayan ağıdır. Çıktı değeri önceki değerlerin regresyonundan çıkartılır. NAR sinir ağları birden fazla döneme ait sonuçları elde etmek için kapalı döngü kullanır (Benmouiza ve Cheknane 2016):

$$\hat{y}(t+p) = f(y(t-1) + y(t-2) + \dots + y(t-d)) \quad (1)$$

Denklemden $\hat{y}(t)$ ifadesi bir y değerinin t zamanındaki tahmin sonucunu, eşitliğin sağındaki ifade herhangi bir doğrusal olmayan fonksiyonu, $t-d$ ise t zamanından ne kadar önce verilerin başladığını temsil eder ve diyagram olarak da Şekil 1'de gösterilmiştir.

Bir sinir ağıının çalışmasının önemli bir parçası eğitim adımıdır; eğitim için kullanılan giriş verileri ağıın tahmin süreci uygulandığında sonuçlar ve bu sonuçlar giriş verileriyle

karşılaştırılıp ağırlıklar yaratılır. Ağırlıklar, çıktı verilerinin örnek verilerden ne kadar farklı değerde diğer bir değişle ne kadar "iyi" tahmin ettiğini temsil etmektedir. Ağırlık değerleri eksi veya artı değerleri farketmeksizin sıfıra olabildiğince yakınlaştırarak optimum tahmin sonucu için tekrar tekrar üretilir.

2.2 Destek Vektör Makineleri (SVM)

Cortes ve Vapnik (1995) tarafından geliştirilen bir ML yöntemi olan SVM, verileri en uyumlu bir hiperdüzlem üzerinde konumlandırıp bu düzenleme işlemi ile hem sınıflandırma hem de regresyon problemlerinde uygulanan denetimli öğrenmeye yönelik bir ML yöntemidir (Shmilovici 2010). SVM yöntemi, istatistiksel öğrenme teorisinde, deneysel risk adı verilen bilinen ve sabit bir veri kümesi üzerindeki performansını değerlendirmeye dayalı öğrenme algoritmaları ailesini tanımlayan bir ilke olan "deneysel risk" (Cherkassky 1997) ve "güvenirlik riski" (Bansal ve Shaliastovich 2010) adı verilen veri dağılımındaki düzensizliği en aza indiren işlemlerin toplamıdır.

$$I_z = \sum_{i=1}^N \alpha_i z_i z + b_0 \quad (2)$$

Denklemden I_z , algoritmanın tahmin ettiği sonuçları, $\alpha_i z_i$ bir z düzlemindeki verilerin ($i=1, 2, \dots, N$) doğrusal kombinasyonunu ifade eder ve b_0 destek vektörlerinin hata değerini ifade etmektedir (Cortes ve Vapnik 1995).

2.3 Rasgele Orman (RF)

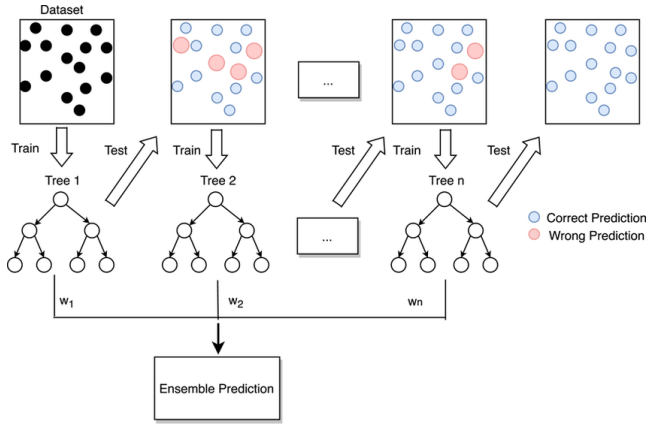
RF, karar ağaçları tabanlı bir "Toplu Denetlenen ML" tekniğidir. Betimleyici ve öngörücü kategorilere ayrılan veri madenciliğinde uygulamalar bulur. Betimleyici veri madenciliği, verileri kategorize etmeye ve özetlemeye odaklanırken, öngörücü veri madenciliği geçmiş verilere dayanarak gelecekteki eğilimleri tahmin eder. İstatistiksel model oluşturmaya, hipotezler üretmek için öngörücü değişkenleri analiz etmeye dayanır. Bu hipotezler doğrulama için test edilir. Doğruluk değerlendirmesi çeşitli hata tahmin tekniklerinin kullanılmasını içerir. Betimleyici veri madenciliği genellikle denetlenmeyen makine öğrenimini kullanırken, öngörücü veri madenciliği denetlenen teknikleri kullanır (Kulkarni ve Pradeep 2012).

2.4 k-En Yakın Komşu (kNN)

kNN en basit ML algoritmalarından biri olarak kabul edilir (Cover ve Hart 1967). kNN algoritması, aynı sınıfa ait benzer örneklerin her test örneği için önce kendisine en yakın komşuyu (veriyi) seçmek, ardından algoritmanın öğrendiği en yakın komşuyu kullanarak bu test örneğini tahmin etmektir (Zhang ve diğ. 2018). kNN algoritmasında, her bir örneğe ait optimum yakın komşu sayısını belirlemek optimum tahmin sonuçları gösterecektir.

2.5 Gradyan Artırımı (GB)

Herhangi bir uyum kriterine dayalı eklemeli genişlemeler için genel bir GB paradigması geliştirilmiştir (Friedman 2001). GB yöntemi, karar ağaçlarının en uygun kombinasyonunu stratejik olarak bulmak için eğitim verilerinin ağırlıklı bir sürümünden temel modelleri ardışık olarak üretir yani GB bir karar ağacı türüdür (Breiman 2001; Zhang ve Haghani 2015). GB, ağaçların yani zayıf tahmin edicilerin birleştirilerek güçlü bir tahmin edici oluşturulması stratejisi ile çalışmaktadır. Veri



Şekil 2. Bir GB karar ağacının tahmin akışı diyagramı (Zhang ve Haghani 2015).

seti tekrar tekrar karar ağaçlarına öğretilir ve ağırlık değeri (w) çıkartılır. Ağırlık değerleri doğru tahmin değerleri sayısını arttırana kadar yani w değerleri sıfıra yaklaşıncaya kadar üretilir (Şekil 2). Karar ağaçları ML algoritmalarının ortak çalışma prensibinde olduğu gibi optimum ağaç sayısı ya da ormanın büyüklüğü seçildiğinde amaca göre duyarlı sonuçlar elde etmeyi sağlamaktadır.

GB, RF'den farklı olarak, karar ağaçlarını bağımsız olarak eğiten "bagging" (*bootstrap aggregating*) yöntemi yerine, önceki karar ağacının sonucuna göre bir sonraki ağacı eğiten "boosting" yöntemini kullanır (yani sıralı öğrenme) (Cha ve diğ. 2021).

3 Doğrulama Yöntemleri

ML algoritmalarında tahmin performansını ölçmek için istatistiksel hata analizi doğrulama yöntemleri kullanılmalıdır. Bu şekilde modelin sonuçlarının giriş verileri ile ne kadar tutarlılık gösterdiği gözlenir. Doğrulama yöntemleri ML algoritmalarının çözülmesi amaçlanan probleme göre (regresyon, sınıflandırma veya kümeleme) değişiklik gösterir. Bunun yanı sıra ortalama karekök hata (RMSE), ortalama mutlak hata (MAE) gibi doğrulama yöntemleri bu problemler için ortak araç olarak kullanılabilir. Bu çalışmada, RMSE (Denklem 3), MAE (Denklem 4), Pearson Korelasyon Katsayısı (PCC) (Denklem 5) ve Nash-Sutcliffe Etkililik Katsayısı (NSE) (Denklem 6; Nash ve Sutcliffe (1970)) tahmin modellerimizin performansını teyit edilmesi için kullanılmıştır. RMSE ve MAE doğrulama tekniklerinin 0'a yakın, PCC ve NSE doğrulama tekniklerinin 1'e yakın olması tahmin modellerinin "doğru"ya yakın olması ve modellerin öğrenme ve test etme parametrelerinin "optimum" olarak seçildiği anlamını taşımaktadır.

$$RMSE = \frac{\sqrt{\sum_{t=1}^n (y_t - o_t)^2}}{n} \quad (3)$$

$$MAE = \frac{\sum_{i=0}^{N-1} |y_i - o_i|}{n} \quad (4)$$

$$r = \frac{\sum_{t=1}^n (y - \bar{y})(o_t - \bar{o})}{\sqrt{\sum_{t=1}^n (y - \bar{y})^2} \sqrt{\sum_{t=1}^n (o_t - \bar{o})^2}} \quad (5)$$

$$NSE = 1 - \frac{\sum_{t=1}^n (y_t - o_t)^2}{\sum_{t=1}^n (o_t - \bar{o})^2} \quad (6)$$

Çizelge 1. ML Modellerinin tahmin sonuçları (RMSE, MAE, PCC, NSE) ve tahmin performansları (G/N: "/" sonrasındaki değer). Tüm değerler için TEMT: Temmuz 2024.

Model	TEMD	RMSE	MAE	PCC	NSE
NAR	118.13	24.45/-	19.14/-	0.98/-	0.75/-
NAR-SVM	117.99	24.71/0.08	19.51/0.07	0.98/0.99	0.74/0.99
NAR-RF	117.43	24.69/0.24	19.51/0.16	0.98/0.99	0.74/0.99
NAR-kNN	117.79	24.70/0.40	19.51/0.25	0.98/0.99	0.74/0.99
NAR-GB	117.61	24.63/0.72	19.45/0.47	0.98/0.99	0.74/0.99

Denklemlerde, y , tahmin sonuçlarını o ise giriş verilerini temsil etmektedir.

4 Sonuçlar ve Yorum

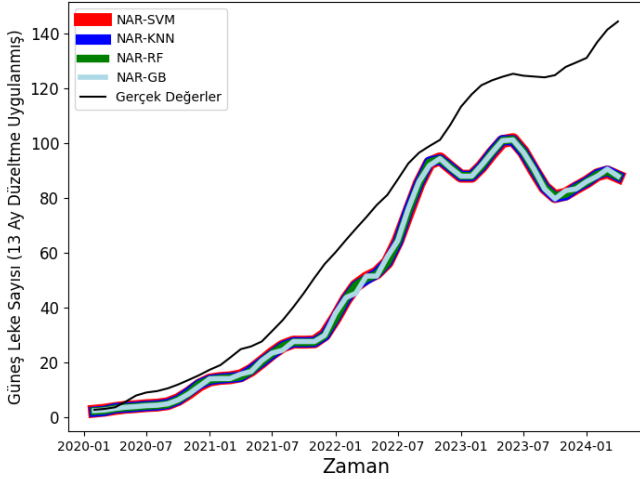
SIDC ("Solar Influences Data Center") tarafından sağlanan SILSO ("Sunspot Index and Long-term Solar Observations") veri paketi kullanılarak, 1759-2019 yıllarına ait SSN verileri analiz edilmiştir. Bu veriler, 25. Güneş Çevrimi periyodunda, NAR, SVM, RF, kNN ve GB gibi makine öğrenimi algoritmalarıyla tahmin edilmiştir. SSN₂₅, ilk olarak NAR algoritmasıyla 1-10-1 (1 giriş katmanı, 10 sinir hücresi, 1 çıkış katmanı) mimarisiyle tahmin edilerek, bu tahminlerden elde edilen sonuç verileri de SVM, RF, kNN ve GB algoritmalarına giriş verisi olarak tanımlanmıştır. Bu şekilde Çizelge 1'deki gibi NAR-SVM, NAR-RF, NAR-kNN ve NAR-GB algoritma kombinasyonları oluşturulmuştur. NAR ve NAR kombinasyonlarının tahmin sonuçları RMSE, MAE, r ve NSE ile hata analizi işlemleri yapılarak tahmin performansları denetlenmiştir.

Çizelge 1'de ML modellerinin "Tahmin Edilen Maksimum Değer" (TEMD), "Tahmin Edilen Maksimumun Tarihi" (TEMT) ve RMSE, MAE, r ve NSE hata analizi doğrulama tekniklerinin gerçek değerlere (G) ve NAR'ın tahmin ettiği değerlere (N) karşılaştırmalı olarak verilmiştir (G/N). Bütün doğrulama tekniği karşılaştırmaları Ocak 2020 ile Nisan 2023 tarihleri arasındaki SSN değerleri üzerinden yapılmıştır ve tüm tahminler sabit Güneş çevrimi periyodu olup ortalama 132 aylık ya da 11 yıllık hesaplanmıştır.

İlk olarak NAR'ın performansı gerçek verilere Temmuz 2022'ye kadar yakın bir trend yakalamıştır. Fakat bu tarihten sonra gerçek verilerin yükseliş eğimi ile tahmin sonuçları eğrileri ayrılmıştır (Şekil 3). Bu fark RMSE (G) ve MAE (G) hata değerlerinde gözükse de PCC (G) ve NSE (G) uyumluluk değerlerinde bu uzaklık gözlenmemektedir. NAR kombinasyonları için sonuçlar hem gerçek değerlere göre hem de NAR'ın tahmin sonuçlarına göre karşılaştırıldığında NAR sonuçlarıyla benzer ve NAR tahmin sonuçlarına "sadık" olduğu gözlenmiştir. NAR kombinasyonlarının NAR'a göre hata değerlerine (RMSE ve MAE) bakıldığında oldukça düşük hata değerine, değer uyumluluklarına (r ve NSE) bakıldığında ise oldukça yüksek uyumluluk gözlenmiştir (Çizelge 1).

Kombinasyonlarından NAR-SVM, NAR'ın tahmin ettiği SSN₂₅(NAR)=118.13 (Temmuz 2024) değerine en yakın değer tahminini SSN₂₅(NAR-SVM)=117.99 (Temmuz 2024) olarak bulmasından, hata değerlerinin ve uyumluluk değerlerinden dolayı en başarılı tahmin kombinasyonu seçilmiştir.

Şekil 3'teki ayrılık 25. Güneş Çevriminin 132 aylık (11 yıl) bir periyottan daha az olabileceği ihtimalini oluşturabilir. Bunun için de NAR-NAR kombinasyonu tahmin zincirinde tahmin



Şekil 3. NAR ve NAR kombinasyonlarının gerçek SSN değerleri ile karşılaştırılması.

edicilerinde parametre değişikliği yapılarak güncel değerleri takip ederek bu değerlere yakın tahmin sonuçları elde edilmesi amaçlanmaktadır. Modeller halen geliştirilmekte ve takip edilen Güneş Leke Sayılarına göre güncellenmektedir.

Kaynaklar

- Apario, A. J. P., Carrasco ve diğ., 2023, *Solar Physics*, 298(8), 100. <https://doi.org/10.1007/s11207-023-02194-7>
- Bansal, R, ve Shaliastovich, I.A, 2010, *The American Economic Review*, 100, 537-541.
- Benmouiza, K, and Cheknane, A., 2016, *Theoretical and Applied Climatology*, 124, 945-958. doi:10.1007/s00704-015-1469-z.
- Breiman, Leo. *Machine Learning*, 2001, 45, 5-32. doi:10.1023/A:1010933404324.
- Cao, J., Xu, T., Deng, L., Zhou, X., Li, S., Liu, Y., Wang, W., & Zhou, W., 2024, *The Astrophysical Journal*, 969(2), 120. <https://doi.org/10.3847/1538-4357/ad4551>
- Carrington, R. C., 1859, *Monthly Notices of the Royal Astronomical Society*, 20, 13-15. doi:10.1093/mnras/20.1.13.
- Cha, G.W., Moon, H.J., Kim, Y.C., 2021, *Int J Environ Res Public Health*, 18(16), 8530. doi:10.3390/ijerph18168530.
- Cherkassky, V., 1997, *IEEE Transactions on Neural Networks*, 8(6), 1564.
- Clover, E. W. ve Dietrich, W. F., 2013, *Journal of Space Weather and Space Climate*, 3, A31. doi:10.1051/swsc/2013053.
- Cover, T. ve Hart, P., 1967, *IEEE Transactions on Information Theory*, 13(1), 21-27.
- Cortes, C ve Vapnik, V., 1995, *Machine Learning*, 20(3), 273-295.
- Du, Z. L., 2020, *Astrophysics and Space Science*, 365(6), 134. doi:10.1007/s10509-020-03818-1.
- Friedman, J. H., 2001, *The Annals of Statistics*, 29(5), 1189-1232. doi:10.1214/aos/1013203451.
- Haigh, J. D., 2007, *Living Reviews in Solar Physics*, 4(1), article 2. doi:10.12942/lrsp-2007-2.
- Hodgson, R., 1859, *Monthly Notices of the Royal Astronomical Society*, 20(1), 15-16. doi:10.1093/mnras/20.1.15a.
- Kalkan, M. Y., Saygac, A. T., & Fawzy, D. E., 2022, *Turkish Journal of Astronomy and Astrophysics*, 3(3), 57-60. doi:10.55064/tjaa.1037256.
- Kalkan, M. Y., Fawzy, D. E., & Saygac, A. T., 2023, *Monthly Notices of the Royal Astronomical Society*, 523(1), 1175-1181. doi:10.1093/mnras/stad1460.
- Kane, R. P., 2002, *Solar Physics*, 205(2), 383-401, doi:10.1023/A:1014296529097.

- Kulkarni, V. Y. ve Pradeep K. S., 2012, 2012 *International Conference on Data Science & Engineering (ICDSE)*, IEEE, 64-68. doi:10.1109/ICDSE.2012.6282329.
- Munro, P., 2010, In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning*, 73-73. Springer US, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_51
- Nash, J.E., ve J.V. Sutcliffe, 1970, *Journal of Hydrology*, 10(3), 282-290. doi:10.1016/0022-1694(70)90255-6.
- Penza, V., Berrilli, F., Bertello, L., Cantoresi, M., & Criscuoli, S., 2021, *The Astrophysical Journal Letters*, 922, L12. <https://doi.org/10.3847/2041-8213/ac3663>
- Pesnell, W., 2020, *Journal of Space Weather and Space Climate*, 10. doi:10.1051/swsc/2020060.
- Petrovay, K., 2020, *Living Reviews in Solar Physics*, 17(1), 2.
- Shmilovici, A., 2010, *Data mining and knowledge discovery handbook*, 231-247. Springer.
- Svalgaard, L., 2020, arXiv. doi:10.48550/arXiv.2010.02370.
- Wu, S.-S. ve Qin, G., 2021, arXiv. doi:10.48550/arXiv.2102.06001.
- Zhang, S, ve diğ., 2018, *Pattern Recognition Letters*, 109, 44-54. doi:10.1016/j.patrec.2017.09.036.
- Zhang, Y. ve Haghani, A., 2015, *Transportation Research Part C: Emerging Technologies*, 58, 308-324. doi:10.1016/j.trc.2015.02.019.

Access:

M25-0335: *Turkish J.A&A* — Vol.6, Issue 3.