(Geliş Tarihi / Received Date:13.06.2023)

|  |  |
| --- | --- |
| (**Araştırma Makalesi**) | |
| **Rastgele Orman Modeli kullanarak Enerji Tüketimi Tahmini**    **Ayşe Ayhan1, Furkan Melik Demiray2, Resulberdi Akyev3**    1 Eskişehir Osmangazi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği.Bölümü, 152120201058, Eskişehir  2 Eskişehir Osmangazi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği.Bölümü, 152120191022, Eskişehir  3 Eskişehir Osmangazi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği.Bölümü, 152120181074, Eskişehir | |
|  | |
|  |  |
| **Anahtar Kelimeler:**  *Güç tüketimi, enerji, enerji tüketim tahmini, makine öğrenmesi, yapay zeka, rastgele orman modeli.* | **Özet:** Çalışmamız Tétouan şehrinin 10 dakikada bir veya her saat başı ölçülen sıcaklık, nem, rüzgar hızı, genel difüzyon hızı ve difüzyon hızına göre bölge bir, bölge iki, bölge üç için güç tüketim değerlerinin bulunduğu veri seti üzerinden yürütülmüştür. Bu veri seti ile makine öğrenmesi, model eğitimi, farklı modellerin eğitimi ile eğitilen modeller içerisinde hangi modelin performansı, skor değerinin daha iyi olduğunun incelenip karar verilmesi üzerinde araştırmalar gerçekleştirilmiştir. Değerlendirmeler sonucunda performansı en iyi bulunan algoritma (model) kullanılarak zamana, nem oranına, rüzgar hızına, genel difüzyon hızına, difüzyon hızına bağlı bir şekilde bölge bir, bölge iki ve bölge üç için bir haftalık ve bir aylık güç tüketimi tahmini yapılmıştır. Bu çalışmada, dünya ekonomisinde güç tüketiminin önemli rolü dikkate alınarak, Tétouan Şehri’nin enerji tüketimi tahmini için beş farklı model kullanılmıştır: Lineer Regresyon, Karar Ağaçları, K-En Yakın Komşu, Destek Vektör Modeli ve Rastgele Orman. Modellerin işlevselliği, açık kaynaklı bir veri kümesiyle doğrulanmış ve farklı performans ölçütleri kullanılarak değerlendirilmiştir. Test sonuçları, Rastgele Orman modelinin ölçülen ve tahmin edilen enerji tüketimi arasında en düşük hata oranına sahip olduğunu göstermektedir. Buna ek olarak, Rastgele Orman modeli diğer modellere kıyasla daha kısa eğitim ve test süresi gerektirmektedir. Bu nedenle, Rastgele Orman modeli güç tüketimi tahmini için etkili ve doğru bir çözüm sunmaktadır. |
|  |  |
|  |  |
| **Power Consumption Prediction using Random Forest Model** | |
|  | |
|  | |
| **Keywords:**  Power consumption, energy, energy estimation, machine learning, artificial intelligence, random forest model. | **Abstract:** Our study was conducted using a dataset containing power consumption values for region one, region two, and region three based on temperature, humidity, wind speed, general diffusion rate, and diffusion rate measured every 10 minutes or hourly in the city of Tétouan. Machine learning techniques, model training, and evaluation of different models were performed to determine which model performed better based on their score values. The best-performing algorithm (model) was used to predict weekly and monthly power consumption for region one, region two, and region three, considering time, humidity, wind speed, general diffusion rate, and diffusion rate. In this study, considering the significant role of power consumption in the global economy, five different models were used for predicting energy consumption in Tétouan City: Linear Regression, Decision Trees, K-Nearest Neighbors, Support Vector Model, and Random Forest. The functionality of the models was verified using an open-source dataset and evaluated using different performance metrics. The test results indicate that the Random Forest model has the lowest error rate between the measured and predicted energy consumption. Additionally, the Random Forest model requires shorter training and testing time compared to other models. Therefore, the Random Forest model provides an effective and accurate solution for power consumption prediction. |
|  |  |

# 1. Giriş (Introductıon)

2014 Nüfus Sayımına göre, Tétouan, 11.570 km²'lik bir alana sahip, 380.000 nüfuslu Fas'ta bulunan bir şehirdir. Şehir ülkenin kuzey bölgesinde yer alır ve Akdeniz'e bakar. Hava yaz aylarında özellikle sıcak ve nemlidir.

Devlete ait ONEE (Elektrik ve İçme Suyu Ulusal Ofisi)'ye göre, Fas'ın 2019'daki enerji üretiminin büyük bir kısmı kömürden (38%) oluşmaktaydı. Ardından hidroelektrik (16%), fuel oil (8%), doğalgaz (18%), rüzgar (11%), güneş (7%) ve diğerleri (2%) takip etmektedir. [1]

Yenilenebilir olmayan kaynaklara ( %64) güçlü bir bağımlılığın varlığı göz önüne alındığında, enerji tüketimini tahmin etmek paydaşların satın almalarını ve stoklarını daha iyi yönetmelerine yardımcı olabilir. Bunun üzerine Fas'ın planı yenilenebilir kaynaklardan gelen üretimi artırarak enerji ithalatını azaltmaktır. Rüzgar ve güneş gibi kaynakların yıl boyunca mevcut olmama riski olduğu bilinmektedir.

Ülkenin enerji ihtiyaçlarını anlamak için orta ölçekli bir şehirden başlamak bu kaynakları planlamak için büyük bir adım olabilir.

# 2. Materyal ve Metot (materıals and methods)

## 

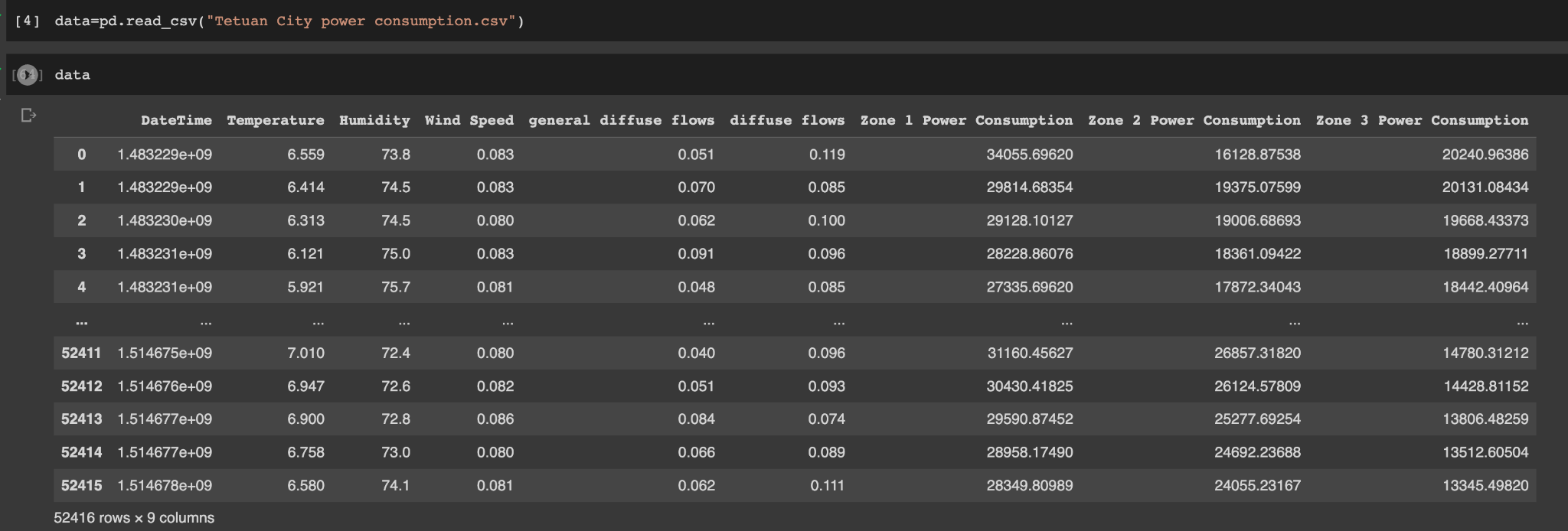
## 2.1. Data Set

Power consumption of Tetouan city Veri Seti (UCI)

# Veri setimiz 2017 yılının başından beri on dakikalık aralıklarla, sensörler aracılığı ile ölçülen:

Sıcaklık, Nem, Rüzgar Hızı, Genel Dağınık Akışlar, Dağınık Akışlar, Zone 1 Güç Tüketimi, Zone 2 Güç Tüketimi, Zone 3 Güç Tüketimi değerlerinden oluşmaktadır.

**Şekil 1.** Veri setinin bir bölümü.



Verilerin %20’si test için, %80’i eğitim için kullanıldı.

## 2.2. Kullanılan Modeller

**Rastgele Orman modeli:**

* Veri Toplama:

Öncelikle, Tetouan Şehri'ndeki güç tüketimi verilerini toplamamız gerekiyor. Bu veriler genellikle belirli bir zaman diliminde (on dakikalık ve bir saatlik aralıklarla) kaydedilmiş olabilir. Güç tüketimi verilerini yanı sıra diğer potansiyel bağımsız değişkenler de (örneğin, hava sıcaklığı, saat dilimi, rüzgar) kaydedilmiş olmalıdır.

* Veri Düzenleme:

Toplanan verileri düzenlemeliyiz. Eksik veya anormallikler içeren verileri temizlemeli ve uygun bir formatta sunmalıyız. Örneğin, tarih ve saat bilgilerini uygun bir zaman formatına dönüştürmeliyiz.

* Bağımlı ve Bağımsız Değişkenlerin Belirlenmesi:

Güç tüketimi, bağımlı değişkenimiz olacak. Diğer potansiyel etkileyici faktörleri belirlemeli ve bunları bağımsız değişkenler olarak kabul etmeliyiz. Örneğin, hava sıcaklığı, saat dilimi, rüzgar gibi faktörler Tetouan Şehri'nin güç tüketimini etkileyebilir.

* Veri Analizi ve Model Oluşturma:

Veri setimizi kullanarak Rastgele Orman modelini oluşturabiliriz. Rastgele Orman, bir ansamble modelidir ve birçok karar ağacının bir araya gelerek oluşturduğu bir yapıdır.

Model eğitimi sırasında, Rastgele Orman, rastgele örneklemelerle (bootstrap) ve rastgele değişken seçimiyle (feature sampling) birden çok karar ağacı oluşturur. Her bir ağaç, veri setinin farklı bir alt kümesiyle eğitilir ve bağımsız değişkenlerin rastgele bir alt kümesini kullanır. [3]

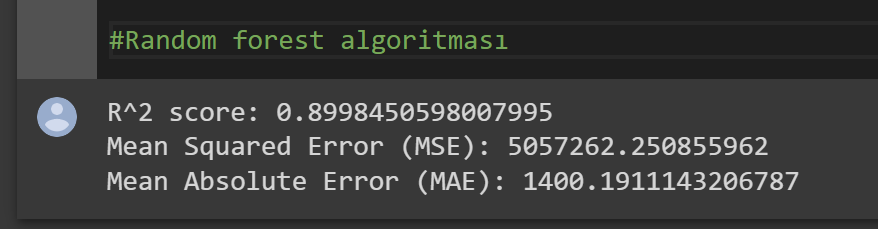
Model, her bir ağacın tahminlerini birleştirerek final tahmini yapar. Bu şekilde, Rastgele Orman modeli, Tetouan Şehri'nin güç tüketimini tahmin etmek için veri setindeki karmaşık ilişkileri tanımlar.

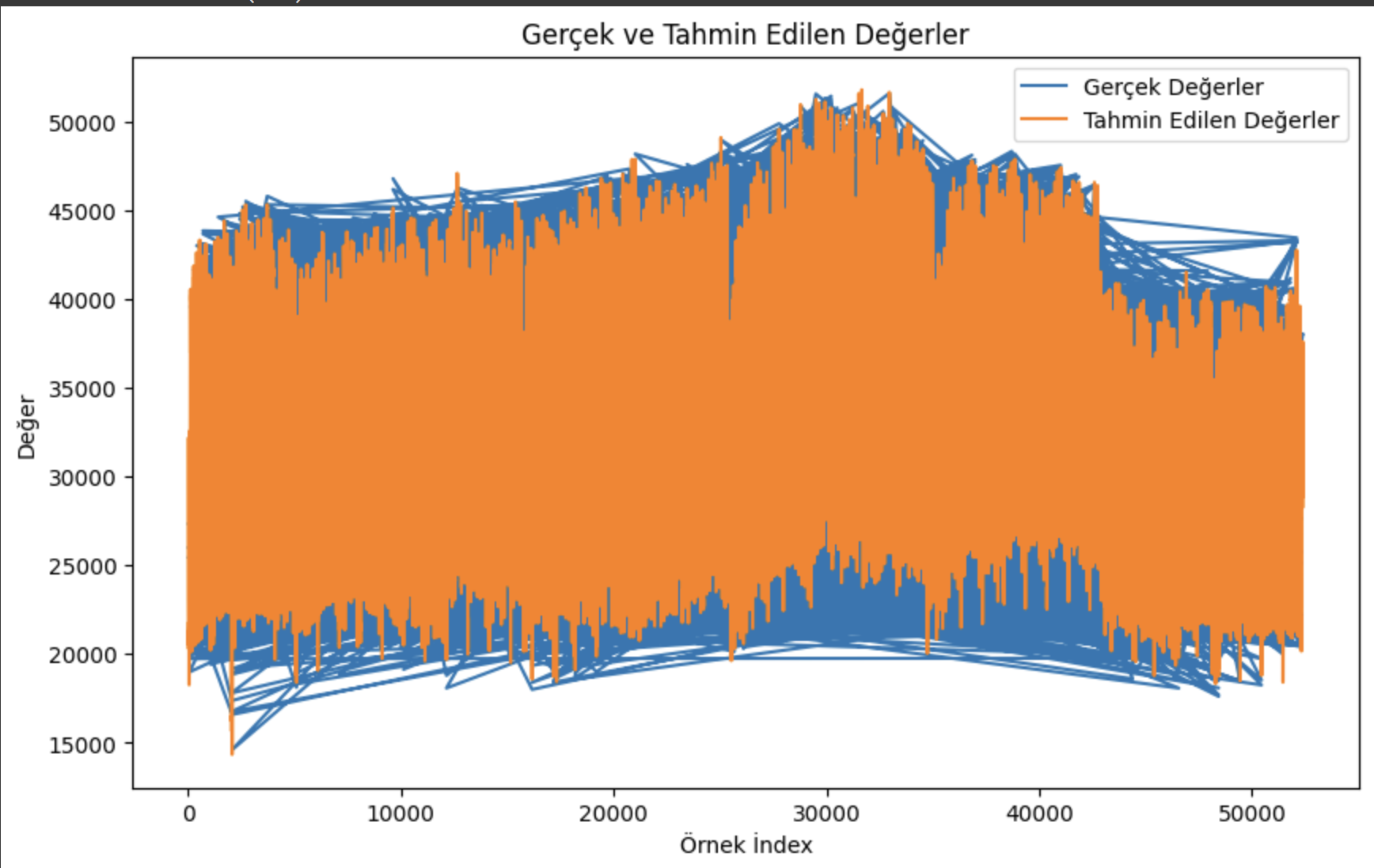
Yeni bir veri noktası geldiğinde, Rastgele Orman modeli bu veri noktasını her bir ağaca besler ve her bir ağaçtan gelen tahminleri birleştirerek güç tüketimi tahmini yapar.

Rastgele Orman modeli, veri setindeki değişken önem sıralamasını da sağlar. Bu sayede, hangi bağımsız değişkenlerin güç tüketimini daha fazla etkilediğini belirleyebiliriz. Örneğin, hava sıcaklığı değişkeni güç tüketiminde en önemli faktör olabilirken, saat dilimi veya rüzgar gibi diğer değişkenler daha az önemli olabilir.

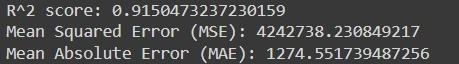
Sonuç olarak, Rastgele Orman modeli, Tetouan Şehri'nin güç tüketimini tahmin etmek için kullanılan etkili bir yöntemdir. Bu model, veri setindeki kompleks ilişkileri yakalayabilir, değişken önem sıralamasını sağlar ve overfitting'e karşı dirençlidir. Yeni veriler geldiğinde, modeli kullanarak etkili bir güç tüketimi tahmini yapabiliriz.

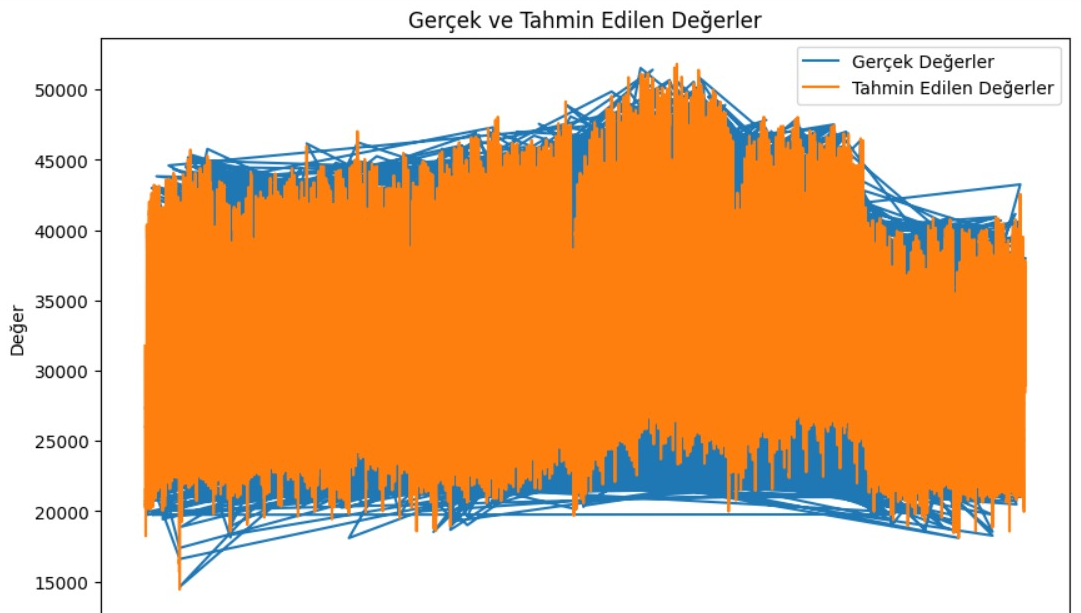
* Kullanılan Modellerin Mean Squared Error ve Mean Absolute Error değerleri görsellerde belirtildi:



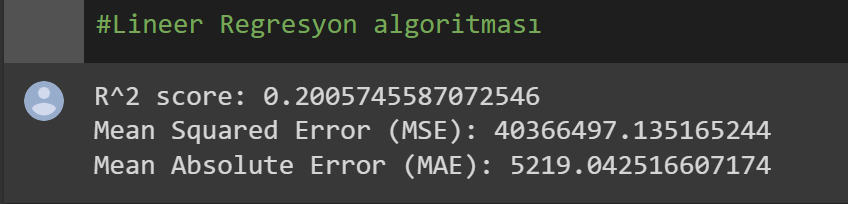


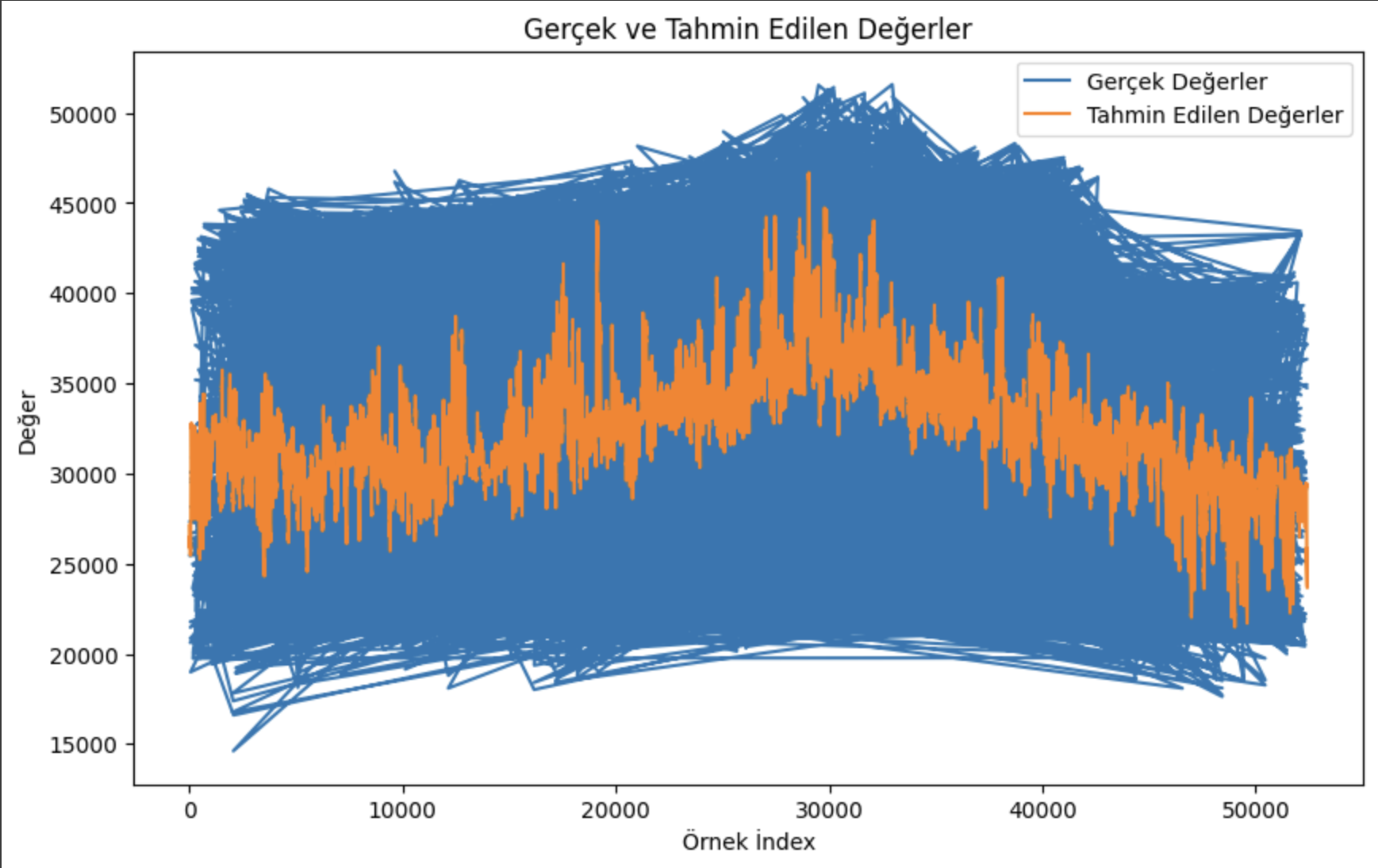
**Random Forest Algoritmasında yapılan çalışmalar ile ulaşılan %4’lük geliştirmenin sonucu aşağıda belirtilmiştir.**

****

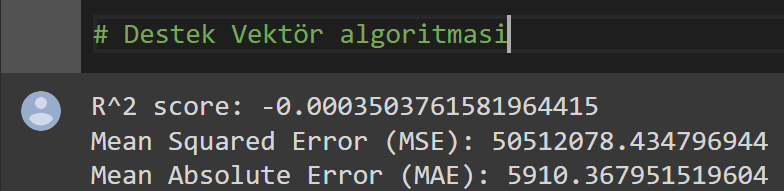
****

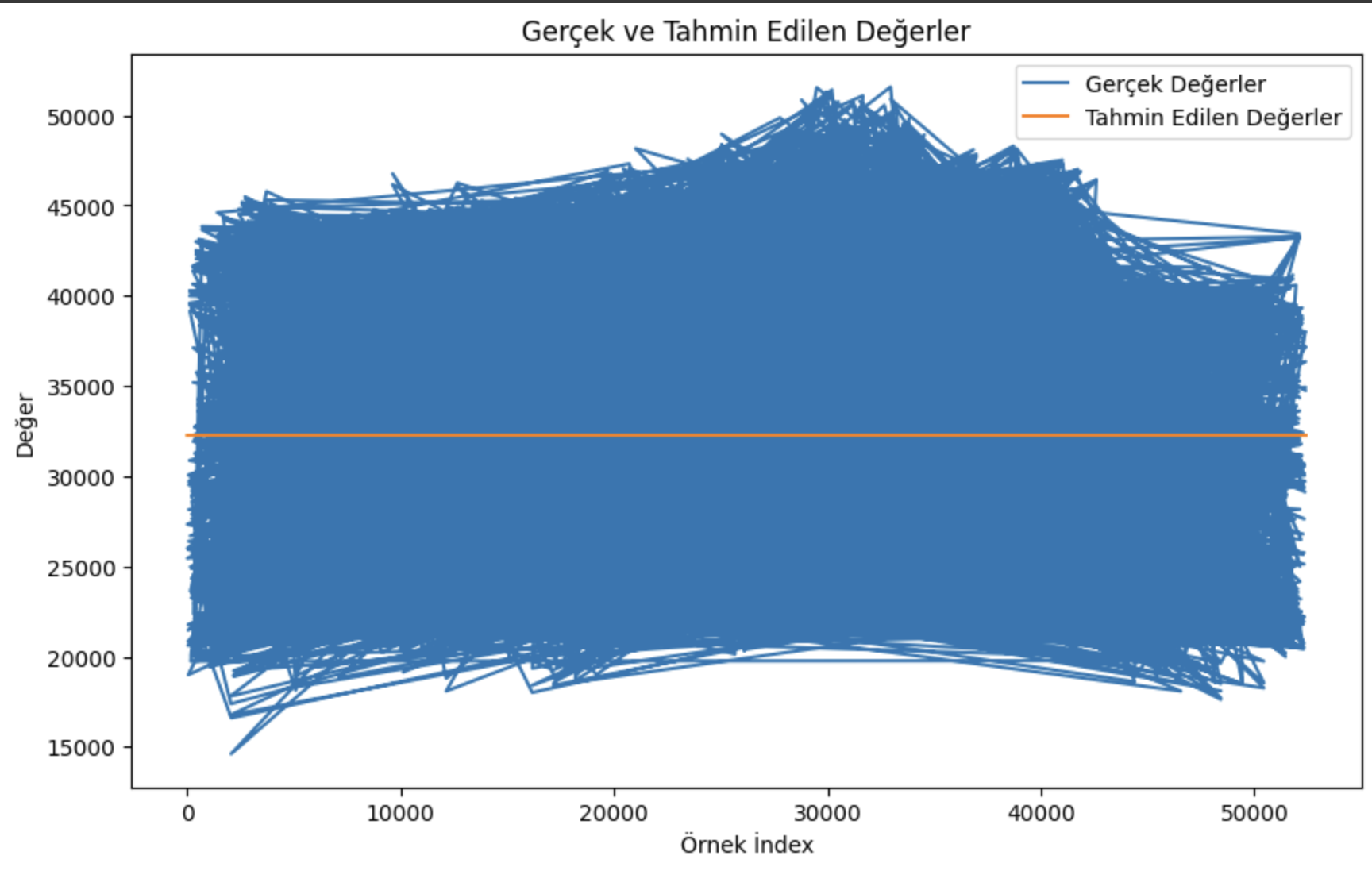
**Lineer Regresyon:** Lineer regresyon, bir bağımlı değişkenin (enerji tüketimi) bir veya daha fazla bağımsız değişkene (örneğin, hava sıcaklığı, rüzgar, saat dilimi) olan doğrusal bir ilişkisini modellemek için kullanılır. Lineer regresyon, en iyi uyan doğruyu (veya hiperdüzlemi) bulmak için verileri analiz eder. Bu model, tahminlemek istediğimiz enerji tüketimini tahmin etmek için kullanılabilir.



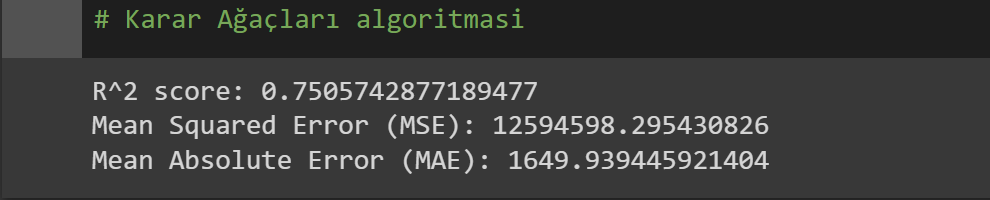


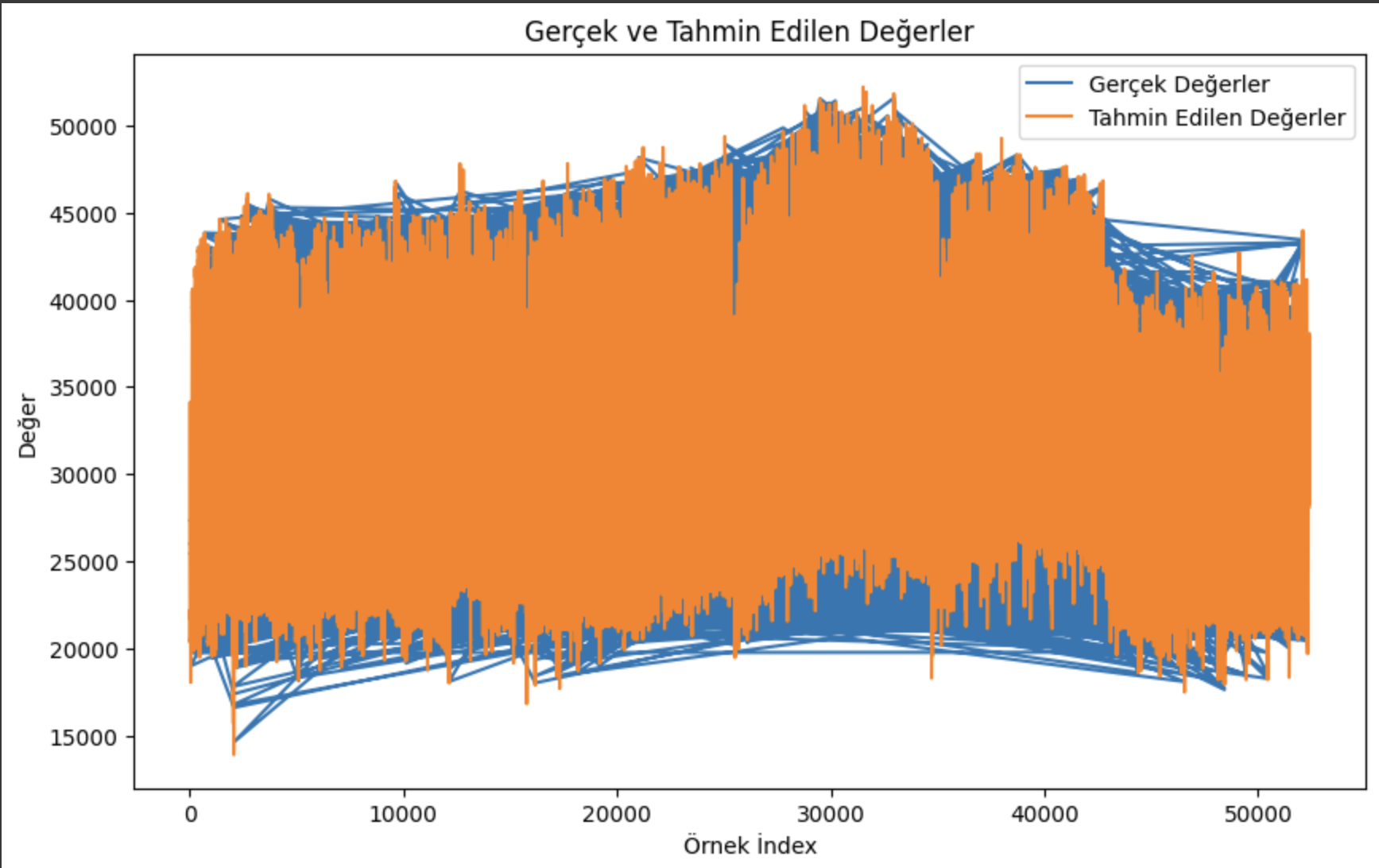
**Destek Vektör Makinesi (SVM):** Destek vektör makinesi, sınıflandırma ve regresyon problemleri için kullanılan bir öğrenme algoritmasıdır. SVM, bir veri kümesini, etiketlenmiş örneklerin iki sınıf arasında bir hiperdüzlemle ayrılmasını sağlayacak şekilde, en iyi şekilde ayırmayı amaçlar. Enerji tüketimi tahmini yapmak için SVM, enerji tüketimi ile diğer bağımsız değişkenler arasındaki ilişkiyi modelleyerek verileri analiz eder.



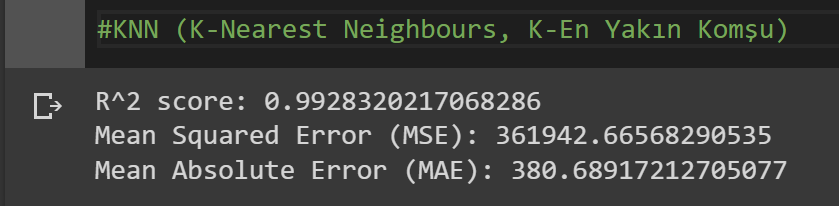


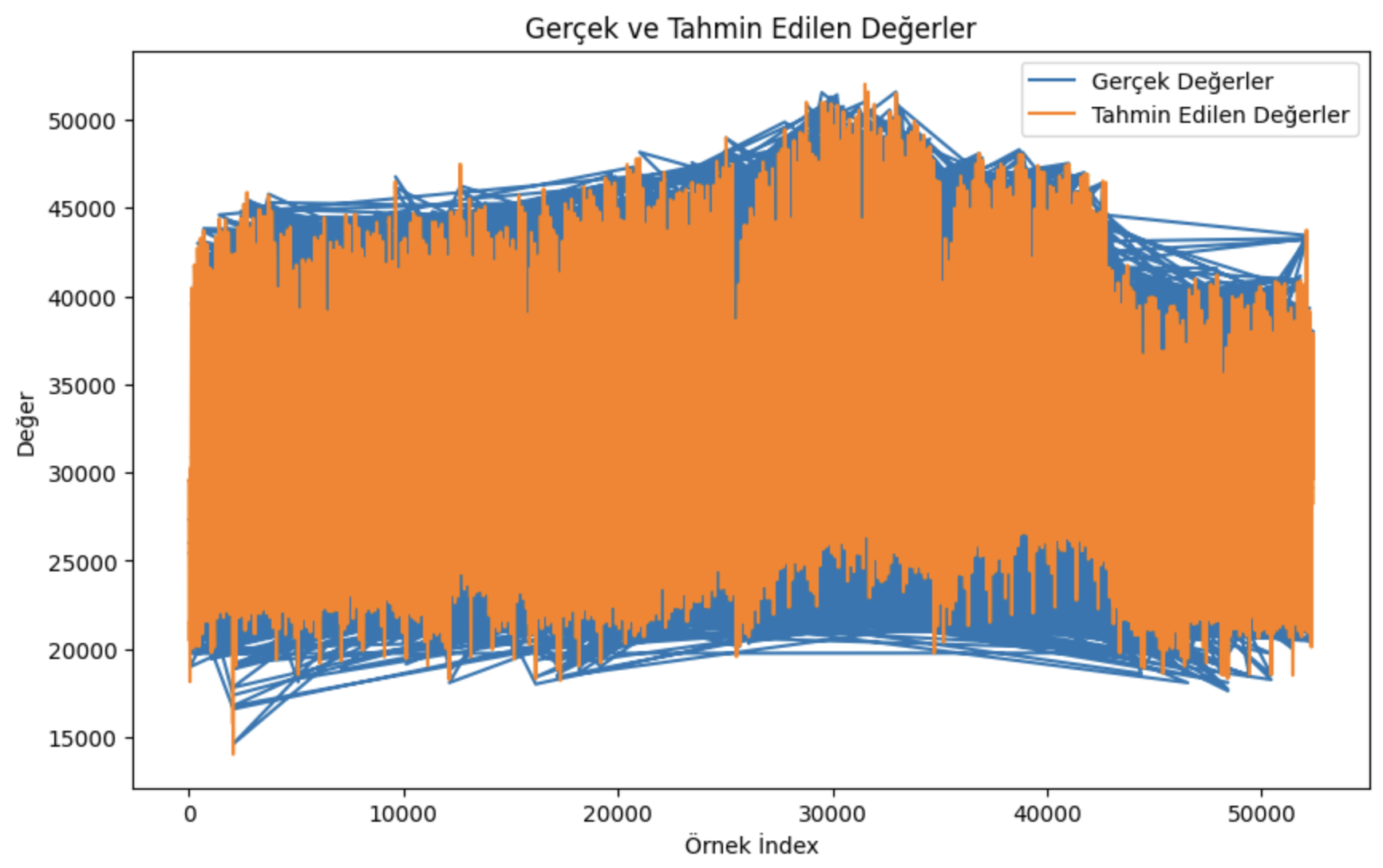
**Karar Ağaçları:** Karar ağaçları, bir dizi karar kuralı kullanarak verileri sınıflandırmak veya regresyon analizi yapmak için kullanılan bir yöntemdir. Her bir karar düğümünde, bir bağımsız değişkenin değerine göre bir dal seçilir ve sonuç olarak enerji tüketiminin tahmini elde edilir. Karar ağaçları, verilerdeki önemli desenleri tanımlayabilir ve bu desenlere dayanarak enerji tüketimi tahminleri yapabilir.





**K-En Yakın Komşu:** K-En Yakın Komşu (k-NN), yeni bir veri noktasının etrafındaki komşu noktaların çoğunluğuna dayanarak sınıflandırma veya regresyon yapmak için kullanılan bir algoritmadır. Enerji tüketimi tahmini için, yeni bir veri noktasının komşuları arasında yer alan enerji tüketimi değerlerine bakılır ve tahmini bu komşuların değerlerine dayanarak yapılır. K-en yakın komşu algoritması, verilerdeki benzerliklere dayanarak enerji tüketimi tahminleri yapabilir.





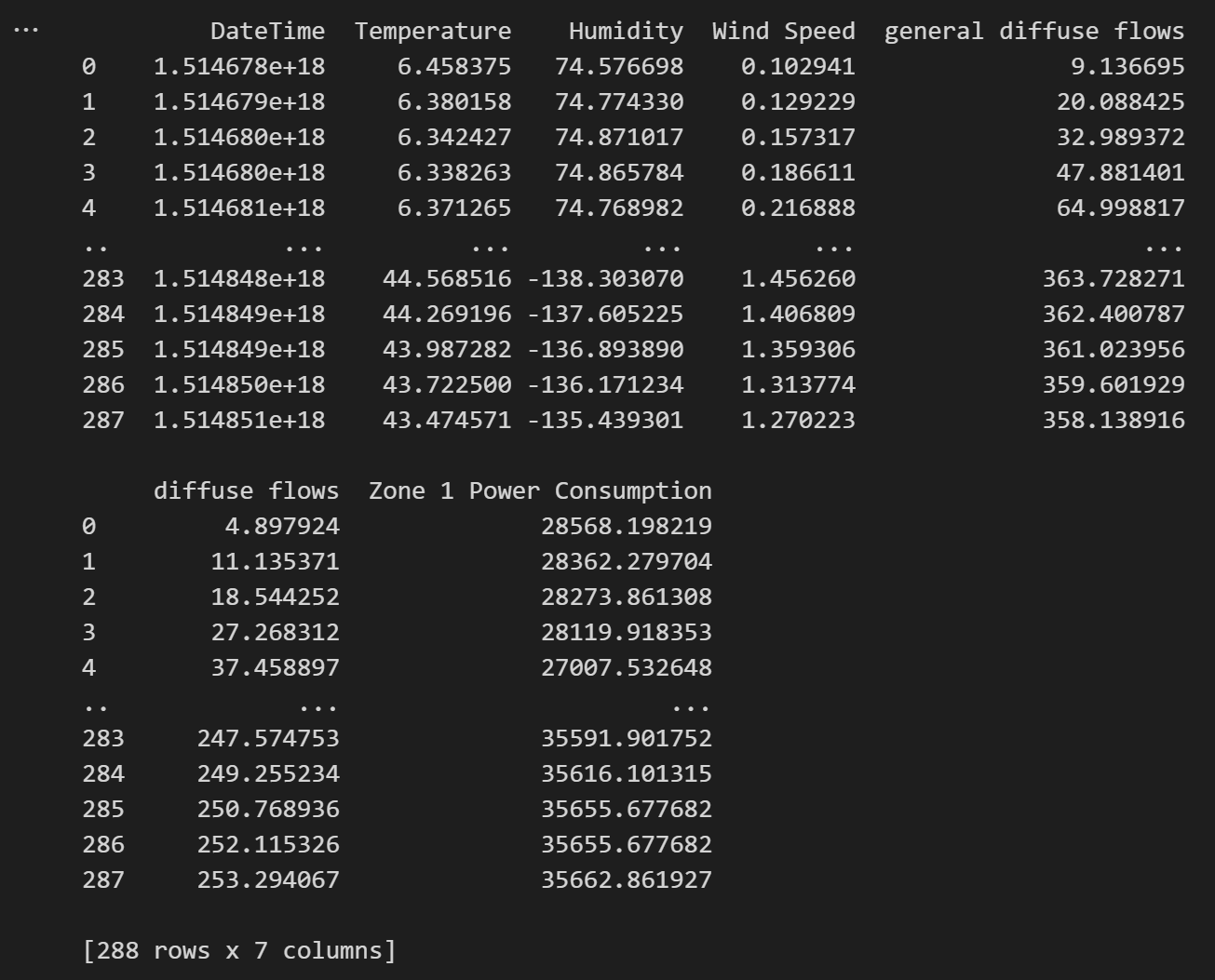
# 3. Bulgular (results)

**Çalışmada kullanılan beş ayrı modelin R^2 Skorları, Ortalama Kare Hata ve Ortalama Mutlak Hata değerleri elde edildi. Bu değerler kıyaslandığında, K-En Yakın Komşu ve Rastgele Orman modellerinin daha iyi sonuç verdiği gözlemlendi. En iyi skor ise Rastgele Orman modelinde alındı. Rastgele Orman** **modeli bu çalışmada en iyi tahmin modeli olarak belirlendi.**

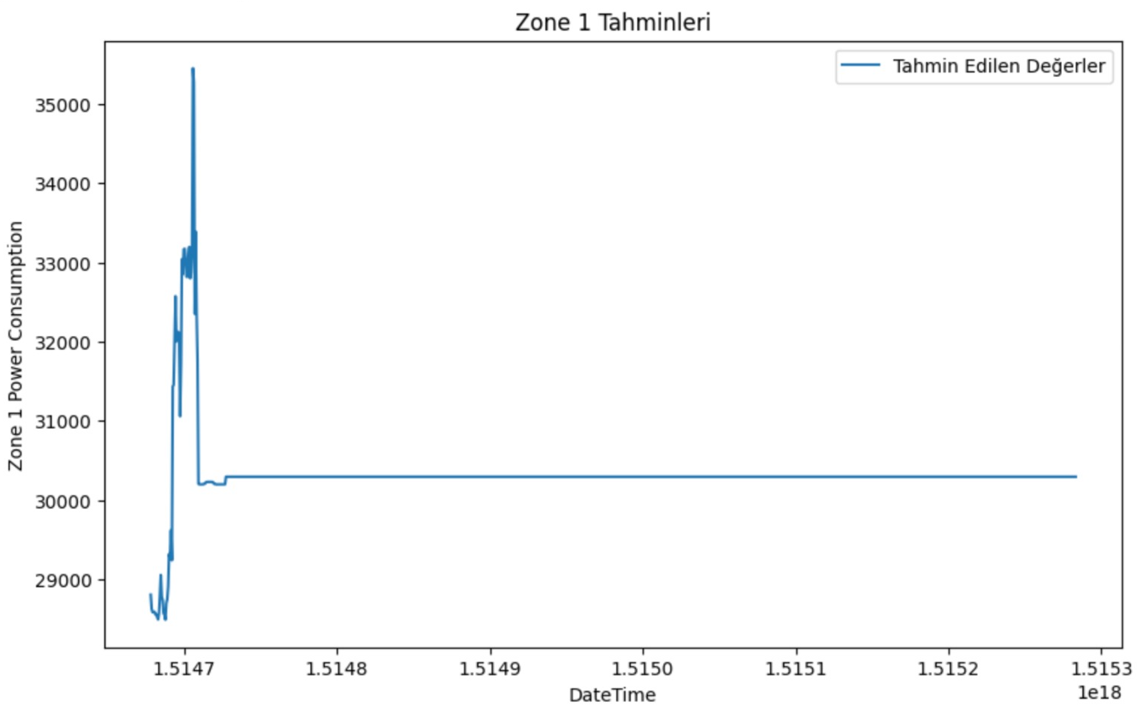
**Sonuçlar, Rastgele Orman modelinin en yüksek R^2 Skoru'na ve en düşük Ortalama Kare Hatası ve Ortalama Mutlak Hata değerlerine sahip olduğunu göstermektedir. Bu model, Tetouan Şehri'nin güç tüketimini tahmin etmek için en etkili yöntem olarak belirlenmiştir.**

.

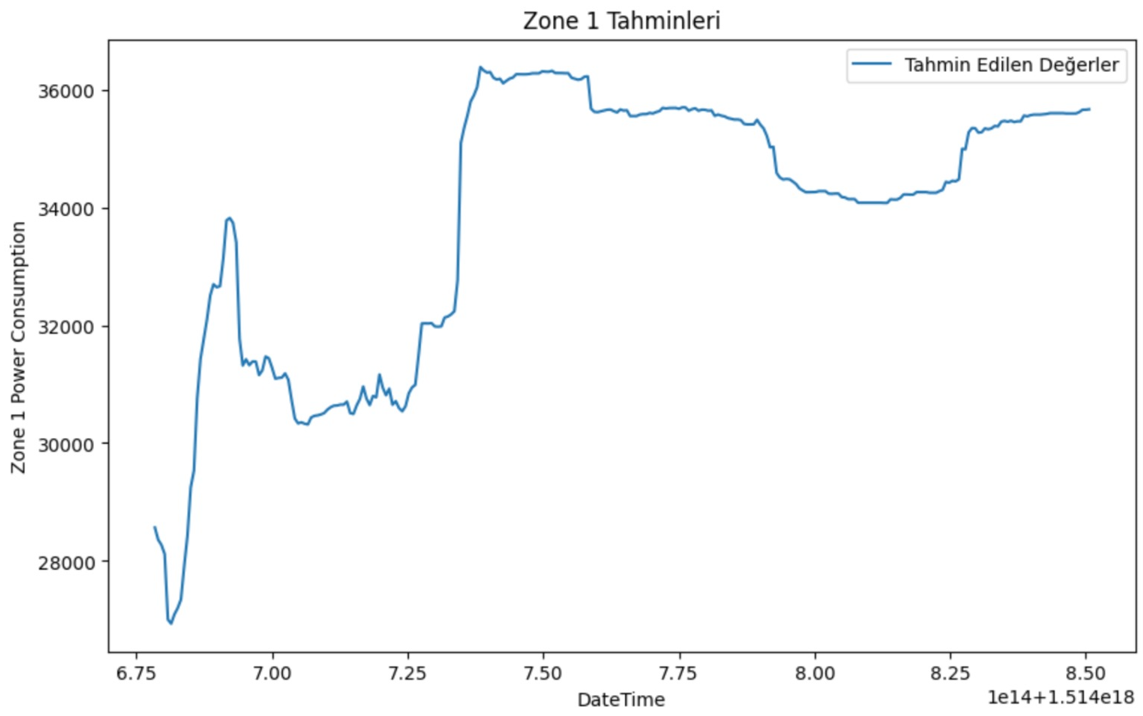
**Şekil 1.** Veri setinin bazı değerleri ve Zone-1 için tahmin edilen değerler.



**Şekil 2.** Zone-1 için bir haftalık enerji tüketimi tahmin değerleri şekil 2.’de gösterilmektedir.



**Şekil 3.** Zone-1 için 2 günlük enerji tüketimi tahmin değerleri.



Aşağıdaki tabloda kullanılan beş model için R^2 Skoru, Ortalama Kare Hatası ve Ortalama Mutlak Hata değerleri gösterilmektedir:

**Tablo 3**. Kullanılan beş model için R^2 Skoru, Ortalama Kare Hatası, Ortalama Mutlak Hata değerleri.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lineer Regresyon | Karar Ağaçları | K-En Yakın Komşu | Destek Vektör | Rastgele Orman |
| 0.22780164454656449 | 0.7514442513611623 | 0.993430868345864 | -0.0003503761581964415 | 0.9150239635644875 |
| 38991682.142094575 | 12550670.01512195 | 331704.27209370903 | 50512078.434796944 | 4243904.892595624 |
| 5130.741562235543 | 1644.475274965662 | 370.4063455034337 | 5910.367951519604 | 1274.1475632488546 |

# 4. Tartışma ve Sonuç (dıscussıon and CONCLUSION)

Güç tüketimi tahmini, enerji yönetimi ve enerji kullanımının iyileştirilmesi açısından önemli bir konudur. Bu çalışmalarda, farklı makine öğrenme modelleri kullanılarak güç tüketimi tahmini yapılmış ve elde edilen sonuçlar değerlendirilmiştir.

Tétouan şehrindeki üç farklı enerji dağıtım ağı için yapılan bir çalışmada, 2017 yılındaki tarihsel güç tüketimi verileri kullanılmıştır. Destek vektör makinesi (SVM), yapay sinir ağı (ANN) ve rastgele orman (RF) modelleri gibi farklı makine öğrenme modelleri karşılaştırılmıştır. Enerji krizi döneminde güç tüketiminin önemi vurgulanarak, RF modelinin diğer modellere göre daha güvenilir ve doğru tahminler yaptığı sonucuna varılmıştır. Ayrıca, RF modelinin daha kısa sürede eğitim ve test aşamalarını tamamladığı belirtilmiştir. [2]

Başka bir çalışmada ise, doğrusal regresyon, karar ağacı, besleme ileri yönlü sinir ağı, destek vektör makinesi ve rastgele orman modelleri karşılaştırılmıştır. Sonuçlar, rastgele orman modelinin diğer modellere göre daha düşük tahmin hataları elde ettiğini göstermiştir. Bu nedenle, rastgele orman modeli tercih edilmiştir. [4]

Her iki çalışma da, rastgele orman modelinin güç tüketimi tahmini için etkili bir seçenek olduğunu göstermektedir. Rastgele orman modeli, diğer modellere kıyasla daha düşük tahmin hataları ve daha yüksek doğruluk oranları sağlamaktadır. Bu nedenle, güç tüketimi tahmininde rastgele orman modelinin tercih edilmesi önerilebilir.

Literatürdeki çalışmaların da dikkate alındığı bu çalışmada, rastgele orman modelinin gelecekteki enerji tüketimini tahmin etmede en düşük hata oranına sahip model olduğu tespit edilmiştir. Literatürdeki sonuçlarla yapılan karşılaştırmalar da göstermektedir ki Random Forest modeli, bu çalışmada, benzer çalışmalardan yaklaşık %4 daha iyi tahmin sonuçları elde edebilmektedir.

Gelecekteki çalışmalarda, rastgele orman modelinin daha da geliştirilmesi ve daha geniş veri setleriyle test edilmesi planlanmaktadır.

**Etik Hususlar (Ethical Considerations)**

**Makine öğrenimi teknolojileri hızla gelişmekte olup, bu teknolojilerin etik açıdan doğru bir şekilde kullanılması son derece önemlidir. Random Forest modeli gibi bir makine öğrenimi algoritması kullanırken, yüksek hassasiyet ile yaklaştığımız etik hususlar:**

**1.Veri Gizliliği ve Mahremiyet:**

* Random Forest modeli, eğitim verilerini analiz ederek tahminler yapar. Bu nedenle, kullanılan veri setinde kişisel bilgiler veya hassas veriler bulunmamalıdır.
* Eğer veri setiniz kişisel bilgiler içeriyorsa, gizlilik ve mahremiyet konularında uygun önlemler almanız önemlidir. Veri anonimleştirme ve şifreleme gibi yöntemler kullanabilirsiniz.

**2.Önyargı ve Adalet:**

* Random Forest modeli, birden fazla karar ağacının bir araya gelmesiyle oluşur. Eğitim verilerindeki önyargılar, modelin sonuçlarında da yansıyabilir.
* Eğitim verilerinin çeşitliliği ve temsil gücü, modelin önyargılı sonuçlar üretme riskini azaltabilir. Veri kümesinin dengeli olmasına ve çeşitli gruplara ait örnekleri içermesine dikkat etmek önemlidir.

**Şeffaflık ve Açıklanabilirlik:**

* Random Forest modeli karmaşık bir yapıya sahiptir ve içerdiği birçok karar ağacının birleşimi sonucunda tahminler yapar. Bu nedenle, modelin sonuçlarını açıklamak ve anlaşılır kılmak zor olabilir.
* Modelin karar sürecini ve önemli özellikleri belirlemek için açıklanabilirlik tekniklerini kullanabilirsiniz. Özellik ağırlıklarını veya karar ağaçlarının çıktılarını görselleştirmek bu konuda yardımcı olabilir.

**Haksız Rekabet:**

* Random Forest modeli, işletmeler ve rekabetçi ortamlarda kullanıldığında, haksız rekabet ve ticari sırların korunması önem kazanır.
* Modelin eğitimi ve sonuçlarının kullanımıyla ilgili gizlilik politikalarına ve yasal düzenlemelere uymanız gerekmektedir.

**Sorumluluk ve İnsan Kontrolü:**

* Makine öğrenimi modelleri, insan kararlarına dayanarak sonuçlar üretse de, son kararı tamamen insanlara bırakmak önemlidir.

**Etik kurallara uyum (Compliance with ethical guidelines)**

Bu çalışma, insan gönüllüleri veya deneysel hayvanları içeren herhangi bir prosedür içermemektedir. Araştırmamız, kurumsal ve/veya ulusal araştırma komitesinin etik standartlarına ve 1964 Helsinki deklarasyonu ve sonraki değişikliklerine uygun olarak gerçekleştirilmiştir. Etik izin alınması gerektiren bir durum olmadığından, "Etik İzin" belge numaraları sunulmamıştır.

**Finansman (Funding)**

Yazarlar, araştırmalarının kamu, ticari veya kar amacı gütmeyen sektörlerdeki fon kuruluşlarından özel bir hibe alıp almadıklarını açıklamalıdır.

**Çıkar çatışması (Conflict of interest)**

Yazar herhangi bir çıkar çatışması olup olmadığını bildirmelidirler.

**Teşekkür (Acknowledgment)**

Bu kısımda; çalışmada yardımları olan kişi veya kişilere teşekkür edilmelidir. Aynı zamanda maddi destek alınan kurum ve kuruluşlara yardım ve desteklerinden dolayı bu bölümde teşekkür edilmelidir. Makale, bir projeden üretilmişse, proje desteği belirtilmelidir. Bölüm başlığı “Times New Roman” yazı tipinde, 10 punto ve kalın olarak yazılmalıdır. Eğer çalışmanızı İngilizce sunmak istiyorsanız başlığı “**Acknowledgment**“ olarak değiştiriniz.

# Kaynakça (REFERENCES)

**[1]** International Trade Administration | Trade.gov. (2019). Morocco — Energy. Retrieved September 25, 2022, from

https://www.trade.gov/country-commercial-guides/morocco-energy#:~:text=Per%20the%20state%2Downed%20power,%2C%20others%20(2%20percent).

**[2]**  Zogaan, W. A. (2022). Power Consumption prediction using Random Forest model. International Journal of Mechanical Engineering, 7(Special Issue 5), 329-341. ISSN: 0974-5823

**[3]** Beheshti, N. (2022, March 2). Random Forest Regression. Towards Data Science.

**[4]** Salam, A., & El Hibaoui, A. (2019). Comparison of Machine Learning Algorithms for the Power Consumption Prediction - Case Study of Tetouan city. In International Renewable and Sustainable Energy Conference (IRSEC) Proceedings, Tetouan, Morocco.

**[5]** Ouahabi, M. H., Elkhachine, H., Benabdelouahab, F., & Khamlichi, A. (2020). Comparative study of five different methods of adjustment by the Weibull model to determine the most accurate method of analyzing annual variations of wind energy in Tetouan - Morocco. Procedia Manufacturing, 46, 698-707.