Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı 49, S. 12-15, Mart 2023

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue 49, pp. 12-15, March 2023

Research Article

Time Series and Data Science Preprocessing Approaches for Earthquake Analysis

Mustafa Kanber[1*], Yunus Santur[2]

[1*] Firat University, Institute of science, Departmant of Software Engineering, Elazığ, Turkey, (ORCID: 0009-0001-4417-8105), kanbermustafa10@gmail.com
[2] Firat University, Institute of science, Departmant of Software Engineering, Elazığ, Turkey, (ORCID: 0000-0002-8942-4605), ysantur@firat.edu.tr

(2nd International Conference on Scientific and Academic Research ICSAR 2023, March 14-16, 2023)

(**DOI:** 10.31590/ejosat.1265261)

ATIF/REFERENCE: Kanber, M. & Santur, Y. (2023). Electronic Properties of FLG/InP Schottky Contacts. *European Journal of Science and Technology*, (49), 12-15.

Abstract

Time series are frequently used today to analyze data that changes over time and to predict future trends. Usage areas of time series data include many applications such as financial market forecasts, weather forecasts, sales forecasts, medical diagnostics and stock management. Among the methods, there are techniques such as autoregressive integration, moving average, long-short-term memory neural network, time series condensation, wavelet transform and Frequency Domain. These techniques are chosen depending on the characteristics of the time series data and their intended use. For example, the ARIMA model is used for variable variance and non-stationary time series, while the LSTM model may be more suitable for capturing long-term dependencies. In this article, it has been tried to prove that time series based artificial intelligence systems can be built on fault movements, which are very difficult to predict on earthquake time series data, and it is quite possible to get useful results. In particular, deep learning methods are among the prominent methods in the article. Deep learning methods are used to detect complex structures and analyze large datasets to produce accurate results. These methods include multilayer perceptrons, long-short-term memory neural network, and radial-based function network. It is also emphasized that factors such as the selection of features used in earthquake prediction, data preprocessing, feature engineering and correct model selection are also important. As a result, the use of artificial intelligence techniques on earthquake time series data has great potential in estimating earthquake risk. Deep learning methods perform better, especially for large datasets, and more accurate results can be obtained with the right model selection. However, factors such as data preprocessing and feature selection also need to be considered.

Keywords: API, Data Science, Earthquake.

Deprem Analizi için Zaman Serileri ve Veri Bilimi Ön İşleme Yaklaşımları

Öz

Zaman serileri günümüzde zaman içinde değişen verileri analiz etmek ve gelecekteki eğilimleri tahmin etmek için sık biçimde kullanılır. Zaman serisi verilerinin kullanım alanları arasında finansal piyasa tahminleri, hava durumu tahminleri, satış tahminleri, tıbbi teşhisler ve stok yönetimi gibi birçok uygulama yer almaktadır. Yöntemler arasında ise otoregresif entegrasyon hareketli ortalama, uzun-kısa vadeli bellekli sinir ağı, zaman serisi yoğunlaştırması, dalgacık dönüşümü ve Frekans Domaini gibi teknikler yer almaktadır. Bu teknikler, zaman serisi verilerinin özelliklerine ve kullanım amaçlarına bağlı olarak seçilir. Örneğin, ARIMA modeli, değişken varyanslı ve durağan olmayan zaman serileri için kullanılırken, LSTM modeli, uzun vadeli bağımlılıkları yakalamak için daha uygun olabilir. Bu makale, deprem zaman serisi verileri üzerinde tahmin etmesi oldukça güç olan fay hareketleri üzerine zaman serisi tabanlı yapay zeka sistemleri kurgulanabileceğini ve faydalı sonuçlar almanın oldukça mümkün olduğu kanıtlanmaya çalışılmıştır. Özellikle, makalede öne çıkan yöntemler arasında derin öğrenme yöntemleri yer almaktadır. Derin öğrenme yöntemleri, karmaşık yapıları algılamak ve doğru sonuçlar üretmek için büyük veri kümelerini analiz etmek için kullanılır. Bu yöntemler arasında, çok katmanlı perceptronlar, uzun-kısa vadeli bellekli sinir ağı ve radyal bazlı fonksiyon ağı yer almaktadır. Ayrıca, deprem tahmininde kullanılan özelliklerin seçimi, veri ön işleme, özellik mühendisliği ve doğru model seçimi gibi faktörlerin de önemli olduğu vurgulanmaktadır. Sonuç olarak, deprem zaman serisi verileri üzerinde yapay zeka tekniklerinin kullanımı, deprem riskinin tahmin edilmesinde büyük bir potansiyele sahiptir. Derin öğrenme yöntemleri, özellikle büyük veri kümeleri için daha iyi performans gösterir ve doğru model seçimi ile daha doğru sonuçlar elde edilebilir. Ancak, veri ön işleme ve özellik seçimi gibi faktörlerin de dikkate alınması gerekmektedir.

Anahtar Kelimeler: API, Veri Bilimi, Deprem.

[*] Corresponding Author: kanbermustafa10@gmail.com

1. Introduction

X Time series are data sets in which data measured over a period of time is ordered in an orderly fashion. These data are often used in the economic, meteorological and social sciences. Time series analysis is used to predict future values by identifying trends and patterns of historical data.

Time series analysis uses many different mathematical methods and statistical techniques. These techniques are used to characterize time series, identify trends and fluctuations, and predict future trends [1-5].

An earthquake is an event that occurs as a result of sudden breaking of rocks in the earth's crust, releasing great energy [6]. Earthquakes can be divided into 3 classes as foreshocks, main and aftershocks. Major earthquakes are earthquakes with large destructive effects, which can cause a large energy release and subsequent relatively smaller earthquakes. Major earthquakes occur at certain periodic intervals. In this way, it allows to make predictions about when the main earthquakes may occur.

Precursor earthquakes are small earthquakes that occur before a major earthquake will occur. Thanks to the analysis of the forerunners, it can be predicted that the main earthquakes will occur and the opportunity to take early measures can be created. Aftershocks are small earthquakes that occur in nearby areas after the main earthquake.

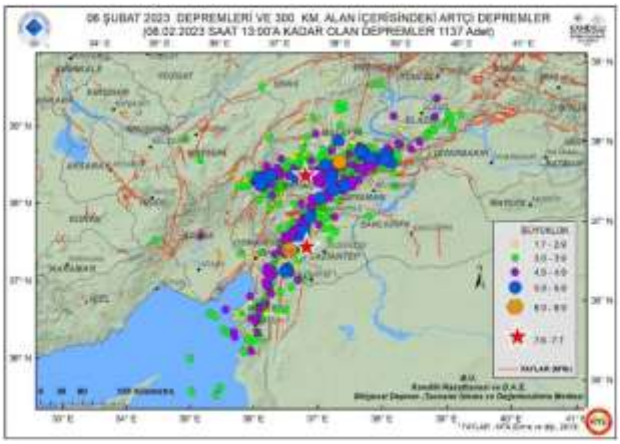

Fig. 1 Main earthquake and its aftershocks [7]

Earthquake data are examples of time series data [8]. The content of these data can be summarized as the earthquake's time, depth, intensity, region name and coordinate information.

In this study, it is aimed to prove that various analyzes of forerunner, aftershock and main earthquakes, which are difficult to obtain effectively by using various pre-processing methods, can be adapted to machine learning models, which are very advanced today, and thus can be beneficial in vital issues such as taking early precautions [9].

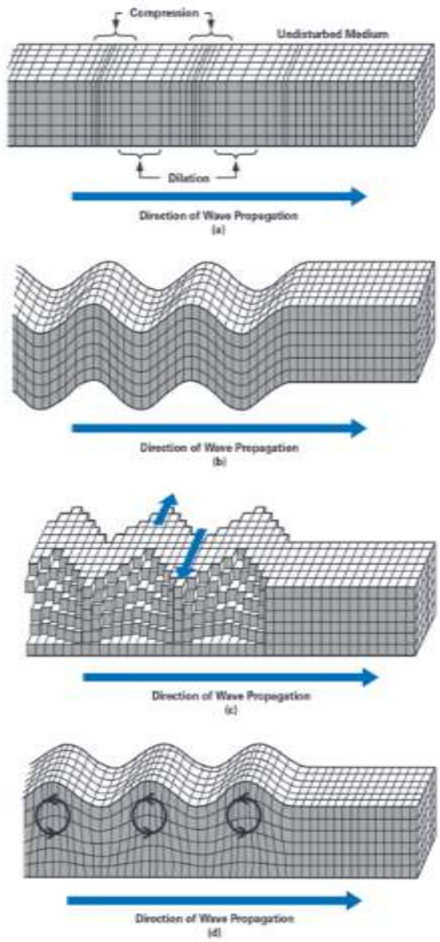

Fig. 2 Earthquake wave types [10]

2. Material and Method

In this section, the acquisition of the data set used in the study and the proposed approach are summarized under sub-titles.

2.1. Dataset

In this study, earthquake data from Kandilli observatory was used as a data source [10]. The data was obtained through the API, transferred to the SQL database and presented as CSV format as a result.

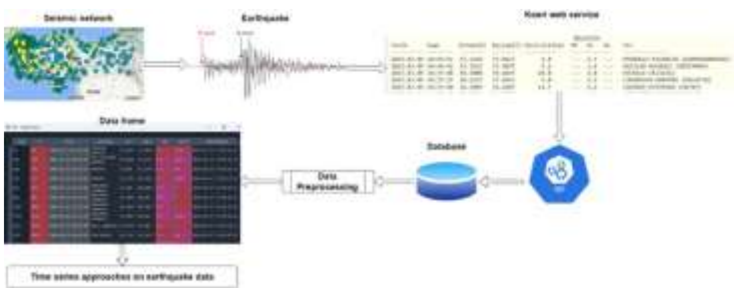

Fig. 3 Material and method pipeline of this study

2.2. Data preprocessing

• The data source is loaded into the Pandas DataFrame.

• The data source was read using the 'pd.read_csv()' function.

Fig. 4 Loading earthquake dataset

2.3. Data exploration

- After the data set was added to the system as a Pandas Dataframe, it was visualized and analyzed using the 'df.head()' function.
- Requirements have been identified so that the data can be ready for use.
- The amount of data examined, missing data etc. presence of conditions has been checked

Fig. 5 Earthquake dataset as a pandas dataframe

2.4. Data preprocessing

• Earthquake intensity (MAG) data has been converted to float data format with PYTHON programming language.

• Earthquake depth (DEPTH) data has been converted to float data format with PYTHON programming language.

• Earthquake latitude (LAT) and longitude (LNG) coordinate data have been converted to PYTHON float data format.

• Earthquake date (DATE_) and record date (RECORDDATE) data have been converted in accordance with Python date data format.

2.5. Data analysis

• The data are grouped and the number of earthquakes in intensity ranges from 2005 to date has been determined.

• Python Numpy library is used to group the data.

• Python Matplotlib library is used to visualize the data [11].

2.6. Data visualization and analysis

As a result of the analysis of the data, the earthquake numbers were grouped according to their intensity and the opportunity to observe was presented. No missing data was found in the data.

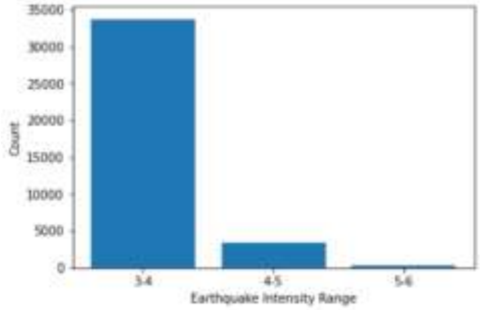

Fig. 7 Earthquakes that between 3 – 6 Mag.

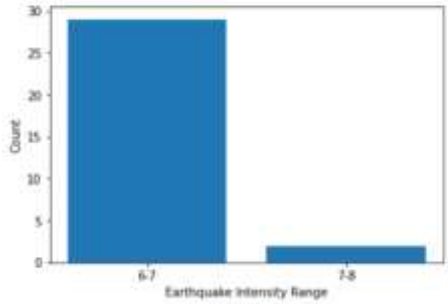

Fig. 8 Earthquakes that between 6-8 Mag.

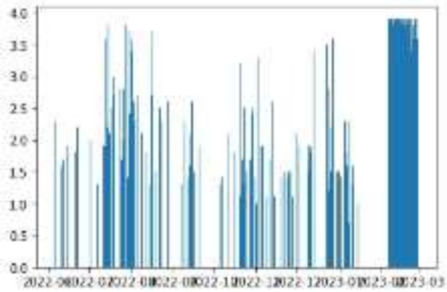

Fig. 9 Earthquake (7.7 Mag) in Kahramanmaraş/Turkey in 2023 and its frequency distributions

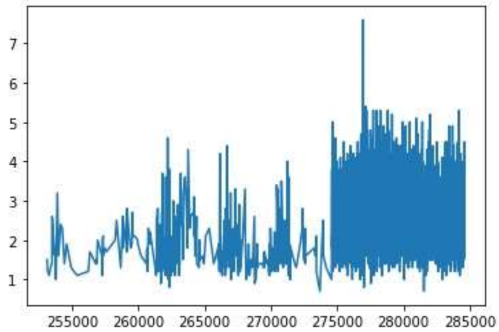

Fig. 10 7 Earthquake (7.7 Mag) in Kahramanmaraş/Turkey in 2023 and its historical line graphics

Python 3.x version was used in this study. Pandas library is used for data preprocessing steps. Matplotlib library is used to visualize data. Numpy library is used to filter data. Google Colab online notebook was used as the development environment [12].

3. Discussion and Conclusion

Earthquakes are unstoppable movements of the earth that have occurred since the formation of the earth. These movements have many parameters such as intensity, depth, location, wave type (p, s, etc.) [13]. By looking at these parameters, predictions can be made about the geological consequences of earthquakes and their effects on future earth movements.

Obtained earthquake data are examples of time series data. Time series data is frequently used in artificial intelligence model training, especially for the prediction of periodic events such as earthquakes [14].

In this study, no estimation or estimation method has been applied, but it is aimed to give information and ideas about the data that will be used or used in artificial intelligence and its sub-branches models that are planned to be developed in the future. For this purpose, the data of earthquakes that occurred in Turkey between the years 2005-2023 were obtained from the Kandilli observatory with APIs, and each data column was converted into a form that can be processed on the Python programming language. As a result, 284.615 processed data obtained were presented to the reader through figures.

5. Acknowledge

This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK, Grant No. 121E733).

References

[1] A. Raiyani, A, A. Lathigara, M. Mehta. Usage of time series forecasting model in Supply chain sales prediction. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1042, No. 1, p. 012022) Publishing. 2020.

[2] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, Z. Yan, L. Sun. Transformers in time series: A survey. *arXiv preprint arXiv:2202.07125*. 2022.

[3] U. Özkaya, Ş. Öztürk. Gaussian Regression Models for Day-Level Forecasting of COVID-19 in European Countries. *Understanding COVID-19: The Role of Computational Intelligence*, 339-356.2022.

[4] Z. Zheng, Z. Wu, Y. Chen, C. Guo, F. Marinello. Instability of remote sensing based ecological index (RSEI) and its improvement for time series analysis. *Science of the Total Environment*, *814*, 152595. 2022.

[5] Y. Xie, M. Ebad Sichani, J. E Padgett, R. DesRoches. The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. Earthquake Spectra, 36(4), 1769-1801. 2020.

[6] Online, available: koeri.boun.edu.tr/sismo/2/06-subat-2023-ml7-5-ekinozu-kahramanmaras-depremi/

[7] H. Shah, R. Ghazali, N. M. Nawi. Using artificial bee colony algorithm for MLP training on earthquake time series data prediction. *arXiv preprint arXiv:1112.4628*. 2021

[8] S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, G. C. Beroza, G. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, *11*(1), 3952. 2020.

[9] Online, available: www.analog.com/en/analog-dialogue/articles/understanding-the-fundamentals-of-earthquake-signal-sensing-networks.html

[10] W. McKinney. *Python for Data Analysis*. "O'Reilly Media, Inc.".2022.

[11] N. Zhang, X. Xu, L. Tao, H. Yu, H. Ye, X. Xie ,H. Chen, H. Deepke: A deep learning based knowledge extraction toolkit for knowledge base population. *arXiv preprint arXiv:2201.03335*. 2022.

[12] J. Santos, A. N. Catapang, E. D. Reyta. Understanding the fundamentals of earthquake signal sensing networks. *Analog Dialogue*, *53*(4). 2019.

[13] M. Akpınar, N. Yumuşak. Günlük temelli orta vadeli şehir doğal gaz talebinin tek değişkenli istatistik teknikleri ile tahmini. Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 35(2), 725-742. 2020.