Research Article

Submission Date
29 / 07 / 2022
Admission Date
11 / 08 / 2022

Forecasting Precipitation by Machine Learning Algorithms to Adapt Climate Change

İklim Değişikliği Uyumu için Makina Öğrenmesi Algoritmaları ile Yağış Tahmini

Erman Ulker [1]

How to Cite: Ulker E., (2022). Forecasting of Precipitation by Machine Learning Algorithms to Adapt Climate Change. *Journal of Environmental and Natural Studies*, 4 (2), 109-118. DOI: https://doi.org/10.53472/jenas.1150975

ABSTRACT:
Throughout history, water has been the main factor to choose settlement for living beings and civilizations. Indeed, there are many advantages of being closer to water basins such as needing fewer transportation, having wealthy crops, and saving energy for water delivery. However, there are some disadvantages as well such as flooding, and erosions. Therefore, it has been understood that predicting precipitation accurately to be able to take necessary measures before any hazardous events must be an utmost aim. In this study, precipitation prediction is investigated by implementing several machine learning algorithms in Python. The data used in this study is taken from two distinct cities in Turkey. The results show that the random forest regression algorithm performs more accurately than other regression models used in this study. Additionally, the prediction for the next 4 years illustrates that it should be expected more rainfall and should be stored in either ground by directing the rainfall to the green areas or harvesting the rainfall for dry seasons. While the climate change occurs dramatically and changes dry and wet seasons duration, the prediction of precipitation amount will help us to adapt to the change more gently.

KEYWORDS: *Precipitation, Climate Change, Forecasting, Python, Regression*

ÖZ:

Tarih boyunca su, canlılar ve uygarlıklar için yerleşim yeri seçiminde en önemli etken olmuştur. Gerçekten de su havzalarına yakın olmanın daha az ulaşım ihtiyacı, zengin mahsul, su dağıtımında enerji tasarrufu gibi birçok avantajı olduğu açıktır. Ancak sel, erozyon gibi dezavantajları da göz önünde bulundurulması gerektiği yaşanan doğal felaketlerle tarih boyunca hissedilmiştir. Bu nedenle, insanoğlu için herhangi bir doğal afetten önce gerekli önlemlerin alınması için yağışların doğru tahmin edilmesi önemli bir amaç olmuştur. Bu çalışmada, Python'da makine öğrenmesi algoritmaları uygulanarak yağış tahmini incelenmiştir. Bu çalışmada kullanılan veriler Türkiye'nin iki ilinden elde edilmiştir. Sonuçlar, Rassal Orman regresyon algoritmasının bu çalışmada kullanılan diğer regresyon modellerinden daha iyi performans gösterdiğini göstermektedir. Ayrıca gelecek 4 yılın öngörüsü, daha fazla yağış beklenmesi ve yağışların yeşil alanlara yönlendirilmesi ile toprakta depolanması veya kurak mevsimler için hasat edilmesi gerektiğini gösterilmektedir. İklim değişikliği dramatik bir şekilde meydana gelip kuru ve yağışlı mevsimlerin süresini değiştirirken, yağış miktarının tahmini, değişime daha yumuşak bir şekilde uyum sağlamamıza yardımcı olacaktır.

Anahtar Kelimeler: *Yağış, İklim Değişikliği, Tahmin, Python, Regresyo*

[1] **Corresponding Author :** Izmir Katip Celebi University, erman.ulker@ikc.edu.tr ORCID: 0000-0001-9279-7288

https://www.jenas.org
ISSN: 2146-9229

INTRODUCTION:

During the last half of the twentieth century, climate change problems have become the focus of researchers, engineers, and scientists. The population growth, rising living standards, and industrial developments have been the main reasons for climate change problems. The change of duration of wet and dry seasons is one of the most remarkable effects that occur nowadays. While the duration of the seasons' changes, society starts to struggle to face either drought or flood due to the current planning of the cities. Therefore, accurate prediction of precipitation for consecutive years is significant for us to adapt ourselves to climate change.

According to the 1st National Declaration of Climate Change (2007), Turkey is marked as a region with drought risk, which will be caused by the increase in temperature values together with the irregular precipitation and intensity values that are thought to continue decreasing. Many studies have been carried out on the distribution of precipitation throughout the year and the trends of change between years. It has been proven by the studies that the differences between the rainfall distribution regions have emerged and it has been revealed that it has undergone great changes in time. Especially precipitation, one of the climatic elements, is one of the most important indicators of change or difference in the climate of a region. For this reason, studies on precipitation have gained importance in every period. In literature, analyzes were carried out according to daily, monthly, and annual rainfall data. (Turkes, 1996, 2007; Cicek, 2001a, 2001b).

Climate change is now a trending topic among various scientific branches. In the literature, it can be seen many studies from different fields in climate change effects our understanding its diverse impact of it. Recently, Bayrak and Kup (2021) studies green infrastructures and Bioretention soil for precipitation impurities. Gursoy – Haksevenler et al. (2021) studied hazardous substances in water. Yucel (2022) stated that global climate change in recent years has caused sudden rainfalls and changed the precipitation regime. O'Gorman (2015) wrote a report on precipitation extremes under climate change and a detailed explanation is given. Using machine learning for the prediction of complex tasks has become common in literature. When the task has many parameters and it is hard to find the trend or the solution to the problem, statistical approaches play a role to give a hint to overcome the problem, at least temporarily. Tosunoglu et al. (2021) used a machine learning algorithm to evaluate trend topics in academic research. Furthermore, Saygin and Baykara (2021) used a machine learning algorithm to detect medical illness by using a machine learning algorithm. Yalcinkaya and Uzer (2022) used GIS-based machine learning to plan waste management for municipalities. Indeed, precipitation is also a complex task to forecast. That is why there are many studies for predicting precipitation by using machine learning algorithms as well. Recently, Sachindra et al. (2018) used statistical downscaling to construct a model by using machine learning techniques for precipitation. Ahmed et al. (2020) also used machine learning algorithms to predict precipitation as well as temperature.

Nowadays, the authorities and governments through incentives focus on sustainable cities (Campisano et al., 2017). Thus, the regulations of water infrastructures become a trend topic among the parliaments and the United Nations. Rainwater harvesting is thought to be an alternative water resource that can be stored and used for variable utilities such as irrigation, flushing toilets, and drinking (Helmreich and Horn, 2009). Spreading the rainwater harvesting systems among the communities leads to increase in self-sufficient locations, which postpones needing for new water infrastructure construction and energy consumption (Steffen et al.,2013; Morales-Pinzón et al., 2014; Devkota et al., 2015). The effective usage of rainwater harvesting, and water infrastructure significantly depends on optimized tank capacity according to precipitation amount (Gardner and Vieritz, 2010; Domènech and Saurí, 2011). To obtain this optimized tank capacity, one should know the precipitation amount versus the demand of the community. In another Word, it is a must to know precipitation to construct viable rainwater harvesting systems.

In this study, two different climatic regions' precipitation data is utilized to perform in the five most popular regression model algorithms in Python. 30 years of precipitation data of two cities are performed and the best model is checked to obtain the best prediction of previous years' rainfall. With the proposed model, the precipitation in the coming years can be foreseen, measures can be taken and adopt the cities to the coming climate change impacts.

https://www.jenas.org
ISSN: 2146-9229

METHODOLOGY:

1. **Machine Learning in Python**

Python is seen as a more advantageous and correct choice among other programming languages, including libraries such as TensorFlow, Keras, Pytorch, and Scikit-learn in engineering and many scientific branches. Python is a dynamic programming language that developers declare variable types, meanwhile it verifies types and errors at compile time when source code is converted to machine code. This provides less code, faster development, and more flexibility.

Machine Learning is programmed through libraries in Python language to improve performance or reduce error according to a criterion. Many approaches to machine learning have been studied using sample data or experience, and most of these approaches can predict, classify, and reduce error. With this acquired ability, we can predict the solution of that problem within a meaningful error by creating sample data or experience sets for complex tasks, which is shown in Figure 1.

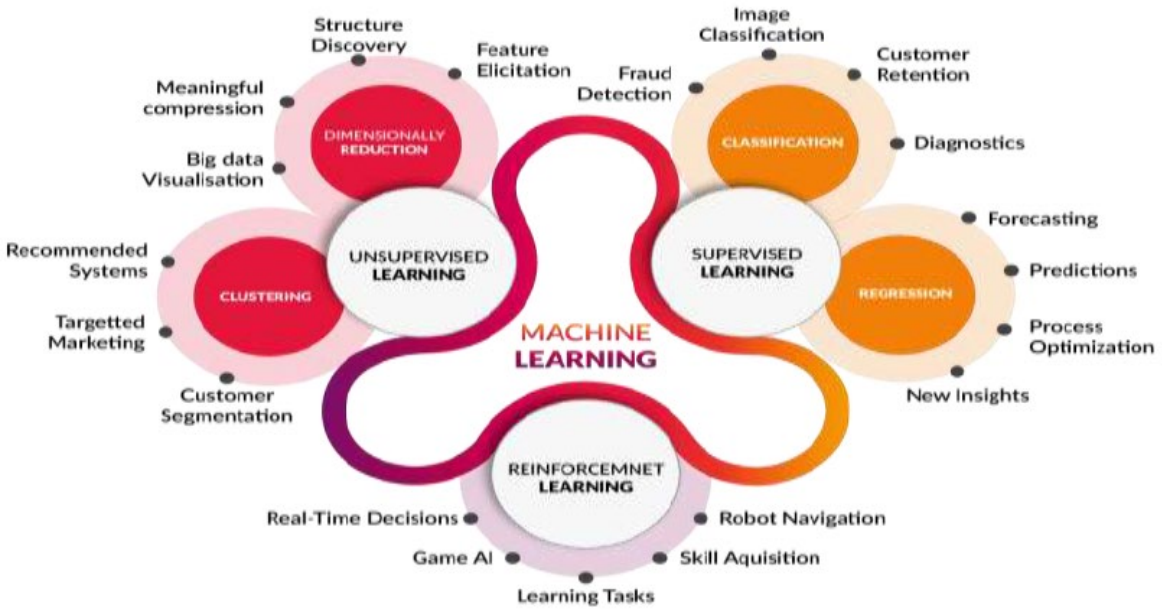

Figure 1 Schematic Representation of Machine Learning Library in Python (Image Source: http://www.cognub.com/index.php/cognitive-platform/)

In supervised learning, machine learning techniques are utilized to obtain regression analysis for predicting continuous solutions based on single or multivariable inputs. For regression analysis, dependent and independent variables are divided into linear and non-linear variables. Generally, linear models are used in studies and all parameters in this model are simple, that is, first-order models. In this study, as regression methods are:

- Linear Regression
- Polynomial Regression

https://www.jenas.org
ISSN: 2146-9229

- Decision Tree Regression
- Random Forest Regression
- Support Vector Regression

2. Dataset Preparation

The dataset used in the present study were gathered from Turkish Meteorology General Directorate (TMGD). The dataset for two distinct cities is obtained. Namely, the precipitation dataset from 1970 to 2011 is obtained from TMGD for Aydın Province. The other dataset is from 1990 to 2020 for Diyarbakır Province. Both datasets are divided into subsets such as input(s) and output. It is preprocessed as to check how many years it needs to be known to predict the best approximation with chosen machine learning algorithm. Therefore, the datasets are divided from 1 to 5 previous years precipitation to train the machine. The table 1 illustrates the schematic representation of dataset for 3 previous annual precipitation as an input to obtain current year's precipitation amount.

Table 1 Sample Illustration of Three Consecutive Years as an Input to Train Machine Learning Algorithm

Input 1	Input 2	Input 3	Output
Precipitation amount in 1990	Precipitation amount in 1991	Precipitation amount in 1992	Precipitation amount in 1993
Precipitation amount in 1991	Precipitation amount in 1992	Precipitation amount in 1993	Precipitation amount in 1994
Precipitation amount in 1992	Precipitation amount in 1993	Precipitation amount in 1994	Precipitation amount in 1995

3. Precipitation Distribution in Study Areas

The variations of annual precipitation in Aydın, where the precipitation is the second highest in the Aegean Region, is significantly high. Annual average rainfall for Aydın is 500 mm. However, the amount of precipitation reaches 1000 mm due to factors such as movement of pressure systems (atmospheric circulation), changes in the air masses and facade systems affecting the region during the year, namely planetary factors, urbanization, change of topographic structure. On the other hand, it drops below 300 mm as well. The reason may be that the hot fronts formed by the maritime tropical air masses, which are effective over the Middle Mediterranean during the winter period, are effective in the form of frontal-orographic precipitation depending on the orography.

A harsh continental climate prevails in Diyarbakır Province, particularly summers are very hot. On the other hand, the winter cold is not as severe as in Eastern Anatolia. The main reason for this is that the Southeastern Taurus arc cuts the cold winds from the North. According to the observations of the meteorology station in the city center, the average of the hottest month is 31 degrees, and the average of the coldest month is 1.8 degrees. The highest temperature ever measured was 46.2 degrees on July 21, 1937, and the lowest temperature was -24.2 degrees on January 11, 1933. Only about 2 percent of the annual average precipitation amount of 496 millimeters falls during the summer months. As it is gone towards the foothills of the mountains in the North, the precipitation increases.

https://www.jenas.org
ISSN: 2146-9229

As observed from Figure 2, the variation of precipitation all over Turkiye shows great differences. Diyarbakır Province faces severe variations of precipitation between 2021 and 2020. It is obvious to expect to have a deeper variation while the climate change prevails its impact on Earth.

Figure 2 Aeral Precipitation Comparison between 2020 and 2021 (Image Source: https://public.wmo.int/en/media/news-from-members/june-2020-precipitation-assessment-turkey)

RESULT AND DISCUSSION:

As it is stated before, two distinct climatic regions are studied in this study. Also, these two provinces have different annual average precipitation amount. Aydın Province is located in the West Region of Turkiye and average of 647 mm annual rainfall is observed. On the other hand, Diyarbakır Province is in the Southeast of Turkiye and average of 496 mm annual rainfall is observed. Leading the climatic change, Diyarbakır Province has change from semi-arid to arid region occasionally. Therefore, the importance and necessity of utilizing rainfall is obvious. The effectiveness of utilizing the rainfall can lead to accurate prediction of rainfall.

In this study, Aydın and Diyarbakır Provinces' annual rainfall amount is gathered and trained in machine learning algorithm. The reason why two different climatic regions data was used in this study is to cross check the outcomes. In the light of these data, regression models were implemented into the machine learning codes created in Python, and the data from 1 to 5 years were analyzed one by one, respectively. Table 2 shows training data for Diyarbakır Province for past 5 years' precipitation amount as input and current year as an output. There may be variations between models. For example, changing the degree of polynomial, changing the number of samples in the Random Forest model. To obtain better results, all the variables are checked for limited numbers.

https://www.jenas.org
ISSN: 2146-9229

Table 2 Training Data for Diyarbakır Province for 5 Years Input

	A	B	C	D	E	F	G
1	YEAR	ANNUAL PRECIPITATION (mm)	YEAR -1	YEAR -2	YEAR -3	YEAR -4	YEAR -5
2	1990	369.9	569.6	428.5	516.8	577.1	411.6
3	1991	569.6	428.5	516.8	577.1	411.6	663.1
4	1992	428.5	516.8	577.1	411.6	663.1	426.9
5	1993	516.8	577.1	411.6	663.1	426.9	442.7
6	1994	577.1	411.6	663.1	426.9	442.7	260.2
7	1995	411.6	663.1	426.9	442.7	260.2	387.7
8	1996	663.1	426.9	442.7	260.2	387.7	605.3
9	1997	426.9	442.7	260.2	387.7	605.3	383.4
10	1998	442.7	260.2	387.7	605.3	383.4	598.7
11	1999	260.2	387.7	605.3	383.4	598.7	476.5
12	2000	387.7	605.3	383.4	598.7	476.5	408.2
13	2001	605.3	383.4	598.7	476.5	408.2	592.5
14	2002	383.4	598.7	476.5	408.2	592.5	397.1
15	2003	598.7	476.5	408.2	592.5	397.1	371.3
16	2004	476.5	408.2	592.5	397.1	371.3	456.7
17	2005	408.2	592.5	397.1	371.3	456.7	398.9
18	2006	592.5	397.1	371.3	456.7	398.9	574
19	2007	397.1	371.3	456.7	398.9	574	625.7
20	2008	371.3	456.7	398.9	574	625.7	429
21	2009	456.7	398.9	574	625.7	429	484.5
22	2010	398.9	574	625.7	429	484.5	472.3
23	2011	574	625.7	429	484.5	472.3	497.6
24	2012	625.7	429	484.5	472.3	497.6	302.6
25	2013	429	484.5	472.3	497.6	302.6	768.6
26	2014	484.5	472.3	497.6	302.6	768.6	726.4
27	2015	472.3	497.6	302.6	768.6	726.4	417.6
28	2016	497.6	302.6	768.6	726.4	417.6	179.1
29	2017	302.6	768.6	726.4	417.6	179.1	
30	2018	768.6	726.4	417.6	179.1		
31	2019	726.4	417.6	179.1			
32	2020	417.6	179.1				
33	2021	179.1					

For the performance criteria, both dataset is constructed and checked in terms of their coefficient of determination (R^2). Decision Tree algorithm is given the best R^2 value as it is seen in Figure 3. However, the maximum value of the

https://www.jenas.org
ISSN: 2146-9229

coefficient of determination also indicates that developed model is not predicting but rather giving the result of the memorized data. For this reason, random forest algorithm is seen as a better option to predict. The detailed quantitative results for Aydın Province can be found in Ulker (2022). It was observed that the best solution was given by random forest and with 5 years and the model is predicted next 5 years with 16% error margin. After cross checking the algorithm, it is now possible to utilize the same algorithm for Diyarbakır Province to predict the precipitation for the next several years. The Figure 4 shows the trend of the previous years and coming next year's precipitation prediction for Diyarbakır Province.

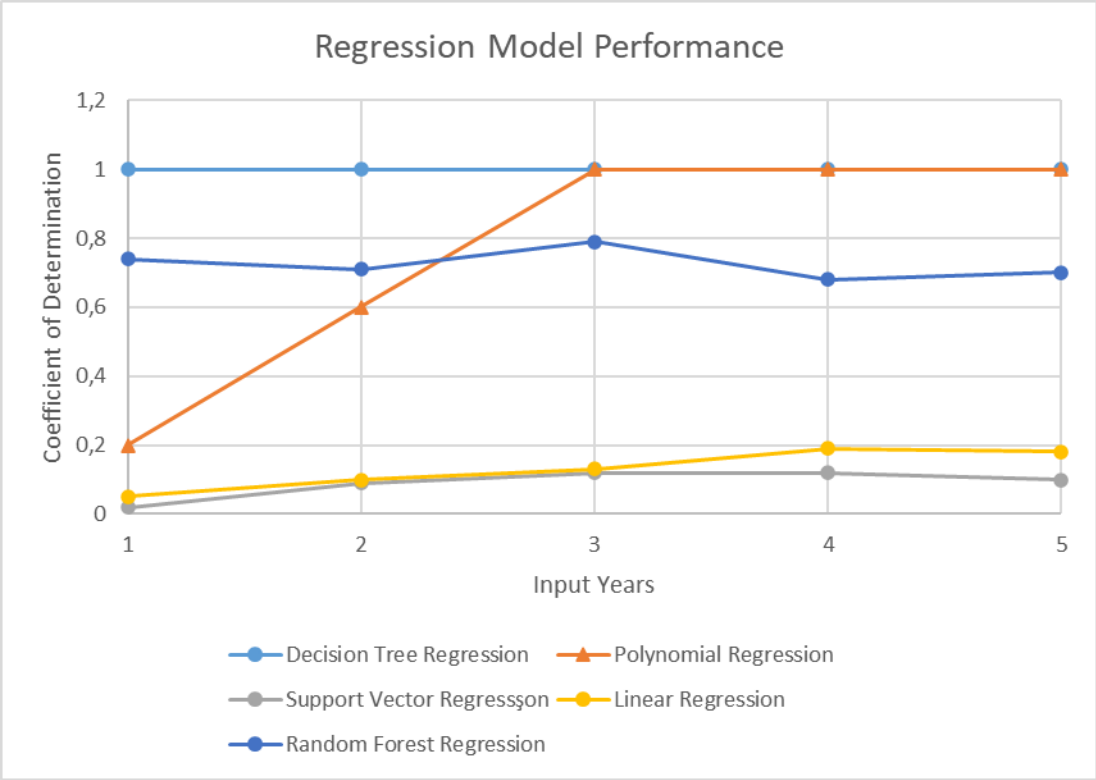

Figure 3 Regression Model Performace Comparison for Diyarbakır Province

https://www.jenas.org
ISSN: 2146-9229

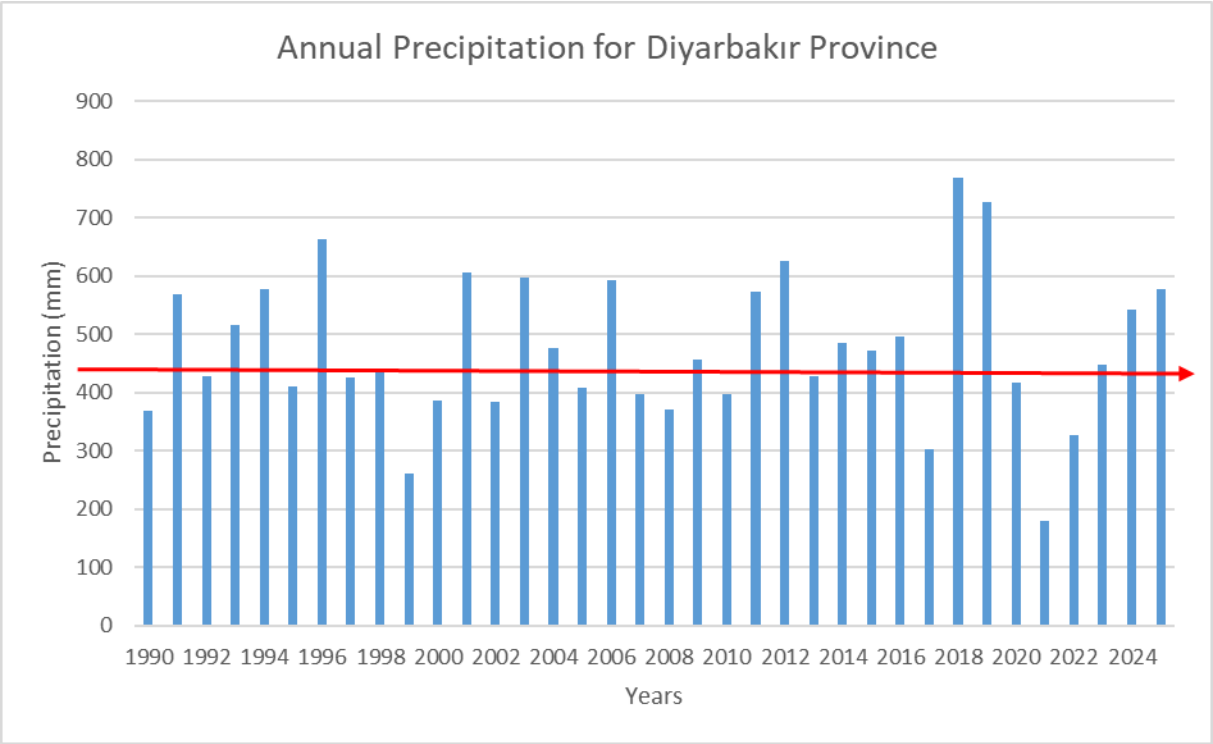

Figure 4 Annual Precipitation for Diyarbakır Province

As we face the climate change and its impact shows destructive consequences in arid regions, Diyarbakır Province should be taken into serious considerations. When it is observed in Figure 2, last year (2021) is the driest season. The machine learning algorithm is compared to its performance by checking Mean Absolute Average Error. It is calculated that the model is predicting the precipitation of that region within 16% error margin. The proposed model gives the prediction of next four years from 2022 to 2025. It shows that the following year annual precipitation is again below the average. However, rising above the average when it reaches to 2025. It is obvious that it is going to be seen great deflection year by year. However, it can be predicted, and necessary measures can be taken while the technological knowledge is becoming more and more advanced. With this information, the rainfall is directed and harvested with respect to utilization and saved for driest years. Hence, the impact of climate change is less destructive for civilizations.

CONCLUSION:

Planning and designing water resources which play a very important role for civilizations faces climate change. Therefore, modeling studies are significant for the future generations to carry sufficient water resources. In this study, annual precipitation amounts were forecasted by python machine learning.

The results show that the best machine learning algorithm used in this study are random forest regression. It shows that the algorithm forecasts the annual precipitation of a location within a 16% error margin. With this knowledge, Diyarbakır province which faces converting the arid region is analyzed. The result shows that the annual precipitation of Diyarbakır is expected below the average again. But there is a rising trend next 4 years. It is important to divert and harvest the rainfall during these years for coming next driest period since last years' precipitation amount was very scarce. With enhancing technology and software libraries, it is relatively easy to predict precipitation and use this knowledge to take necessary measures, particularly for authorities and governments.

KAYNAKÇA:

Ahmed, K., Sachindra, D. A., Shahid, S., Iqbal, Z., Nawaz, N., & Khan, N. (2020). Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms. *Atmospheric Research*, *236*, 104806.

Bayrak, G., & Cansu, K. Ü. P. Yeşil Altyapı Uygulamaları Kapsamında Biyotutma Sistemlerinin Yağmur Suyu Kirletici Giderim Verimlerinin Değerlendirilmesi. *Kent Akademisi*, *14*(3), 853-866.

Campisano, A., Butler, D., Ward, S., Burns, M. J., Friedler, E., DeBusk, K., ... & Han, M. (2017). Urban rainwater harvesting systems: Research, implementation and future perspectives. Water research, 115, 195-209.

Cicek, İ. (2001a). Türkiye'de günlük yağış şiddetleri ve frekansları. *Ankara Üniversitesi Türkiye Coğrafyası Araştırma ve Uygulama Merkezi Dergisi*, *8*, 27-48.

Cicek, İ. (2001b). Türkiye'de mevsimlere göre yağış şiddetleri ve sıklıkları. *Ankara Üniversitesi, Türkiye Coğrafyası Araştırma ve Uygulama Merkezi Dergisi, S*, *8*, 1-26.

Devkota, J., Schlachter, H., & Apul, D. (2015). Life cycle-based evaluation of harvested rainwater use in toilets and for irrigation. Journal of cleaner Production, 95, 311-321.

Domènech, L., & Saurí, D. (2011). A comparative appraisal of the use of rainwater harvesting in single and multi-family buildings of the Metropolitan Area of Barcelona (Spain): social experience, drinking water savings and economic costs. Journal of Cleaner production, 19(6-7), 598-608.

Gardner, T., & Vieritz, A. (2010). The role of rainwater tanks in Australia in the twenty first century. Architectural Science Review, 53(1), 107-125.

Gursoy-Haksevenler, B. H., Atasoy-Aytis, E., Dilaver, M., Yalcinkaya, S., Findik-Cinar, N., Kucuk, E., ... & Yetis, U. (2021). A strategy for the implementation of water-quality-based discharge limits for the regulation of hazardous substances. *Environmental Science and Pollution Research*, *28*(19), 24706-24720.

Helmreich, B., & Horn, H. (2009). Opportunities in rainwater harvesting. Desalination, 248(1-3), 118-124.

Morales-Pinzón, T., Lurueña, R., Gabarrell, X., Gasol, C. M., & Rieradevall, J. (2014). Financial and environmental modelling of water hardness—Implications for utilising harvested rainwater in washing machines. Science of the Total Environment, 470, 1257-1271.

O'Gorman, P. A. (2015). Precipitation extremes under climate change. *Current climate change reports*, *1*(2), 49-59.

Sachindra, D. A., Ahmed, K., Rashid, M. M., Shahid, S., & Perera, B. J. C. (2018). Statistical downscaling of precipitation using machine learning techniques. *Atmospheric research*, *212*, 240-258.

Saygin E., & Baykara, M. (2021). Karaciğer Yetmezliği Teşhisinde Özellik Seçimi Kullanarak Makine Öğrenmesi Yöntemlerinin Başarılarının Ölçülmesi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, *33*(2), 367-377.

Steffen, J., Jensen, M., Pomeroy, C. A., & Burian, S. J. (2013). Water supply and stormwater management benefits of residential rainwater harvesting in US cities. JAWRA Journal of the American Water Resources Association, 49(4), 810-824.

Tosunoglu, E., Yilmaz, R., Ozeren, E., & Saglam, Z. (2021). Eğitimde makine öğrenmesi: Araştırmalardaki güncel eğilimler üzerine inceleme. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi*, *3*(2), 178-199.

Turkes, M. (1996). Spatial and temporal analysis of annual rainfall variations in Turkey. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *16*(9), 1057-1076.

Turkes, M., Telat, K. O. C., & Saris, F. (2007). Türkiye'nin Yağiş Toplami ve Yoğunluğu Dizilerindeki Değişikliklerin ve Eğilimlerin Zamansal ve Alansal Çözümlemesi. *Coğrafi Bilimler Dergisi*, *5*(1), 57-73.

Ulker, E. (2022). Investigation of Precipitatıon Potential of Aydin Province by Using Python with Regression Analysis Methods. *2nd International Antalya Scientific Research and Innovation Studies Congress, March 17-21, 526-531.*

Yucel, G. (2022). Risk Altındaki Kültürel Miras: Karadeniz Bölgesi'ndeki Geleneksel Ahşap Yığma (Çantı) Camiler için Değerlendirme. *Kent Akademisi*, *15*(2), 481-505.

Yalcinkaya, S., & Uzer, S. (2022). A GIS-based multi-criteria decision support model for planning municipal solid waste collection points: A case study of Çağdaş Neighbourhood, Çiğli District, Izmir, Turkey. *Waste Management & Research*, *40*(8), 1297-1310.