Turkish Journal of Science & Technology
17(2), 241-250, 2022

Research Paper
https://doi.org/10.55525/tjst.1063284

Comparison of the Machine Learning Methods to Predict Wildfire Areas

Gözde BAYAT[1*], Kazım YILDIZ[2]
[1] Department of Computer Engineering, Institute of Pure and Applied Sciences, Marmara University, Istanbul, Turkey
[2] Department of Computer Engineering, Technology Faculty, Marmara University, Istanbul, Turkey
[*1]gozdebayat@marun.edu.tr, [2] kazim.yildiz@marmara.edu.tr

(Geliş/Received: 26/01/2022) Kabul/Accepted:21/04/2022)

Abstract: In the last decades, global warming has changed the temperature. It caused an increasing the wildfire in everywhere. Wildfires affect people's social lives, animal lives, and countries' economies. Therefore, new prevention and control mechanisms are required for forest fires. Artificial intelligence and neural networks(NN) have been benefited from in the management of forest fires since the 1990s. Since that time, machine learning (ML) methods have been used in environmental science in various subjects. This study aims to present a performance comparison of ML algorithms applied to predict burned area size. In this paper, different ML algorithms were used to forecast fire size based on various characteristics such as temperature, wind, humidity and precipitation, using records of 512 wildfires that took place in a national park in Northern Portugal. These algorithms are Multilayer perceptron(MLP), Linear regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree and Stacking methods. All algorithms have been implemented on the WEKA environment. The results showed that the SVM method has the best predictive ability among all models according to the Mean Absolute Error (MAE) metric.

Key words: Machine Learning, Random Forest, Support Vector Machine, Decision Tree, WEKA.

Orman Yangınlarını Tahminlemede Makine Öğrenmesi Metotlarının Karşılaştırılması.

Öz: Son on yılda, küresel ısınma sıcaklığı değiştirdi. Orman yangınlarının her yerde artmasına neden oldu. Orman yangınları insanların sosyal yaşamlarını, hayvan yaşamlarını ve ülke ekonomilerini etkiler. Bu nedenle orman yangınları için yeni önleme ve kontrol mekanizmalarına ihtiyaç duyulmaktadır. 1990'lı yıllardan itibaren orman yangınlarının yönetiminde yapay zeka ve sinir ağlarından yararlanılmaktadır. O zamandan beri, çevre biliminde çeşitli konularda makine öğrenmesi (ML) yöntemleri kullanılmıştır. Bu çalışma, yanan alan boyutunu tahmin etmek için uygulanan ML algoritmalarının performans karşılaştırmasını sunmayı amaçlamaktadır. Bu yazıda, Kuzey Portekiz'deki bir milli parkta meydana gelen 512 orman yangınının kayıtları kullanılarak sıcaklık, rüzgar, nem ve yağış gibi çeşitli özelliklere dayalı olarak yangın boyutunu tahmin etmek için farklı ML algoritmaları kullanılmıştır. Bu algoritmalar Lineer regresyon, Destek Vektör Makineleri (SVM), Çok Katmanlı Algılayıcı, K-En Yakın Komşular (KNN), Karar Ağacı ve Yığınlama yöntemleridir. Tüm algoritmalar WEKA ortamında gerçeklenmiştir. Sonuçlar, Ortalama Mutlak Hata (MAE) metriğine göre tüm modeller arasında SVM yönteminin en iyi tahmin yeteneğine sahip olduğunu göstermiştir.

Anahtar kelimeler: Makine Öğrenmesi, Random Forest, Destek Vektör Makineleri, Karar Ağacı, WEKA.

1. Introduction

Forests are one of the most important resources of the world's ecological balance. Also, it provides oxygen for people and natural living areas for animals. The world has been losing its forests rapidly as a consequence of wildfires and tree cutting uncontrollably. Wildfire as a natural disaster has severe effects on all living creatures. Also, it has extremely large economic and social consequences. During the last few decades, gigantic wildfires have occurred in various places of the world. Creating a trustworthy model to predict the size of the burned area in a forest fire is necessary to allocate resources optimally for fire departments. In this paper, ML models have been used to predict how much fire will grow using the dataset that includes wind speed, humidity, location information, temperature, etc. The output of prediction is the burning area and its unit hectares.

A wildfire susceptibility map for the two fire seasons in the Liguria region in Italy was created and validated by using the Random Forest (RF) method [1]. The susceptibility map was investigated considering the dataset of mapped fire environments covering a 21-year period (1997-2017) and different environmental susceptibility factors. Also, the authors aim to compare the performance of ML models. The proposed model is better than the other models to predict areas which were affected by a fire. Hung Van Le and friends [2] suggested a novel deep

* Corresponding author: gozdebayat@marun.edu.tr. ORCID Number of authors: [1] 0000-0003-1116-1881, [2] 0000-0001-6999-1410

neural network model for the prediction of wildfires in a tropical region. They proposed 3 hidden layers to create a wildfire susceptibility map for the Gia Lai province in Vietnam which is called deep neural computing.

A literature review covering 300 publications by the end of 2019 was investigated in [3] to show that ML methods can be used in wildfires. It is shown that the common methods are RF, NN, SVM, Decision trees. Stella and friends use machine learning to address the next day forest fire prediction problem. An ML method utilizing Tree Ensemble and NN, where a large parameter search procedure is performed through cross-validation, has been applied to determine powerful models that are expected to generalize fine on the new data[4] . Meteorological parameters such as temperature, average rains to understand scale of a forest fire can be used. These parameters have been used as a input values for these forecast models, such as long short-term memory (LSTM) backpropagation neural network (BPNN) and recurrent neural network (RNN). The experimental results show, the scale of fire can be predict at the onset onccurence with these informations [5]. ML techniques such as RF, SVM and Logistic Regression (LR) have been exploited to build susceptibility map and compared for the study area of Northern Iran. It was revealed that RF has the highest accuracy and suited for wildfire sensitivty evaluation[6]. Novel gradient boosting models have been applied to predict wildfire activity trained with loss function Extreme-Value theory have been exploited for generate loss function. In the study, the benchmarked against boosting scheme was designed and shown to provide a better proxy for test set performance than pure cross validation. Estimates are compared against reinforcement approaches with different loss functions[7]. BPNN, RNN and LSTM techniques have been applied to data set which include Alberta region meteorological parameters taken from Canadian National Fire Database (CNFDB) [8]. In the study, length of fire time have been exploited along with meteorological parameters to predict burning area.

Authors recommend that to have placed sensors that has massive resolution at the initial phase of fire to predict scale of wildfire. Different synthetic data generation techniques and different ML models have been applied the created synthetic dataset. Results have shown that SVM method has most accuracy to predict large forest fires [9]. Uncertainty is big problem to predict fire, in literature multi-fidelity technqies have became attractive from wildfire researchers, recently. Multi-fidelity techniques have been used to understand fire spread als Monte-Carlo and multi level Monte-Carlo simulation methods have been compared[10]. Not only weather parameters also smoke information has been used to predict wildfire events in early stage. LSTM has been applied with convolutional layers for smoke detection and reached high accuracy 97.8% [11]. Forecast future wildfires is vital point as well in forest fires management. Daily forest fire probability map forecast has been carried out using deep fully convolutinal neural network called AllConvNet. Authors estimated future burn probability map for next seven days using 2006-2017 wildfire period for Australia[12].

Predictive models such as RF, LR, Ridge Regression have been applied for estimate burning area size in [13]. The dataset contains parameters meausured in wildfires between 1911-2015 in the United States. RF algorithm has better performance than LR and Ridge Regression. Beşli and Tenekeci used the data obtained from the satellites for prediction. Forest fires were estimated using Normalized Difference Vegetation Index (NVDI), Land Surface Temperature (LST) and Thermal anomaly (TA) data calculated from satellite data. Decision trees were used to make predictions from the mentioned data. 70% of the data was used to be used as training and remaining as a test. The average performance of the applied method was determined by repeating the training and testing process 10 times with different data. In the experiments carried out, the fires were predicted correctly with an average sensitivity of 98.62%. The actual situation was determined with an average accuracy of 93.11% [14]. RF, linear regression, Stacked Regressor, NN, SVM and KNN algorithms have been used for forecast the burned areas with two different data set. Algorithms have been implemented on the Python environment. Also Data sets have taken from Kaggle and UCI, respectively. Performance of used ML algorithms compared each other. MAE and MAPE error metrics have been used to evaluate the performance of the models[15]. Logistic regression has been applied for predicting areas that can be burned using past meteorological parameters. This technique is easy to implement and also facilitates interpretation of the results obtained and possible duplication of the methodology in other regions or countries. [16]. Trucchia and friends proposed a study which is RF based. Their approach is about to obtain national susceptibility maps in Italy. Each pixel of the study are is classifed by the model. Experimental results show the ability of RF to notice the most senstive areas with defined factors[17].

It is aimed to demonstrate the performance comparison of ML algorithms applied to estimate the burned area size. For this purpose, Linear Regression, SVM, MLP, KNN, Decision Tree and Stacked Regressor methods were applied in WEKA environment. The estimation performance of these methods is compared to the MAE performance metric. In addition, the obtained results are presented by comparing with Moore's study[15].

2. Material and Methods
2.1. Dataset

The dataset was taken from the UCI site. The dataset consists of fires in a national park in northern Portugal between January 2000 and December 2003. It was collected using two different sources. The first of these sources were prepared by the inspector responsible for forest fires in the park. The inspector recorded time, date, location (x and y), Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), Initial Emission Index (ISI), and total burned area data for each fire. The second source was prepared using the meteorological station in the park. The meteorological station recorded various weather information like temperature (Celsius), relative humidity, rain, wind speed. The datas were collected from two sources and converted into a single dataset with a total of 512 entries [18]. In this study, the burning area size attribute was tried to be estimated. The sample dataset content is shown in Table 1. The distribution histogram of the features are given in Figure 1.

Table1. Example of UCI dataset file[18]

X	Y	Month	Day	FFMC	DMC	DC	ISI	Temp	RH	Wind	Rain	Area
7	5	Mar	Fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0	0
7	4	Oct	Tue	90.6	35.4	669.1	6.7	18	33	0.9	0	0
7	4	Oct	Sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0	0
8	6	Mar	Fri	91.7	33.3	77.5	9	8.3	97	4	0.2	0
8	6	Mar	Sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0	0
8	6	Aug	Sun	92.3	85.3	488	14.7	22.2	29	5.4	0	0
8	6	Aug	Mon	92.3	88.9	495.6	8.5	24.1	27	3.1	0	0
8	6	Aug	Mon	91.5	145.4	608.2	10.7	8	86	2.2	0	0
8	6	Sep	Tue	91	129.5	692.6	7	13.1	63	5.4	0	0
7	5	Sep	Sat	92.5	88	698.6	7.1	22.8	40	4	0	0
7	5	Sep	Sat	92.5	88	698.6	7.1	17.8	51	7.2	0	0
7	5	Sep	Sat	92.8	73.2	713	22.6	19.3	38	4	0	0
6	5	Aug	Fri	63.5	70.8	665.3	0.8	17	72	6.7	0	0
6	5	Sep	Mon	90.9	126.5	686.5	7	21.3	42	2.2	0	0
6	5	Sep	Wed	92.9	133.3	699.6	9.2	26.4	21	4.5	0	0
6	5	Sep	Fri	93.3	141.2	713.9	13.9	22.9	44	5.4	0	0
5	5	Mar	Sat	91.7	35.8	80.8	7.8	15.1	27	5.4	0	0
8	5	Oct	Mon	84.9	32.8	664.2	3	16.7	47	4.9	0	0
6	4	Mar	Wed	89.2	27.9	70.8	6.3	15.9	35	4	0	0
6	4	Apr	Sat	86.3	27.4	97.1	5.1	9.3	44	4.5	0	0
6	4	Sep	Tue	91	129.5	692.6	7	18.3	40	2.7	0	0
5	4	Sep	Mon	91.8	78.5	724.3	9.2	19.1	38	2.7	0	0
7	4	Jun	Sun	94.3	96.3	200	56.1	21	44	4.5	0	0
7	4	Aug	Sat	90.2	110.9	537.4	6.2	19.5	43	5.8	0	0
7	4	Aug	Sat	93.5	139.4	594.2	20.3	23.7	32	5.8	0	0
7	4	Aug	Sun	91.4	142.4	601.4	10.6	16.3	60	5.4	0	0

It was considered that some parameters such as temperature, wind, and humidity would be very correlated with the burning area. However, during the preliminary research phase on the data, it was observed that even these highest correlated features of the data set were low correlated with the burning field. As represented in Figure 2(a), the wind has a low relationship to the burning area in the data set. Small-scale fires occur all months of a year whereas large-scale occurred mostly during the summer seasons. This relationship has been shown in Figure 2(b).

The correlation matrix was generated with the help of Python program to measure the correlation between dataset features which can be seen in Figure 3. The correlation value of -0.076 between burned area and relative humidity (RH) indicates an inverse relationship between these data. In this case, it can be seen that the RH data is one of the weak indicators in estimating the fire size. Examining the correlation matrix, the lack of high correlation between burning areas and other data complicates it difficult to predict fire size. Besides, Figure 3 shows the importance of weather parameters such as temperature, wind for estimating burning areas size, which has a high correlation.

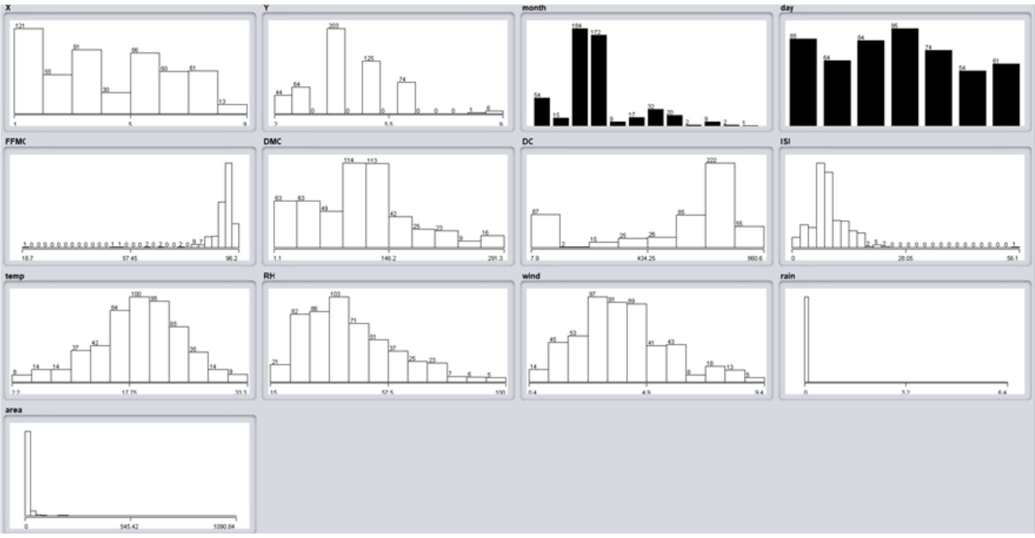

Figure 1. The distrubution histogram of dataset features

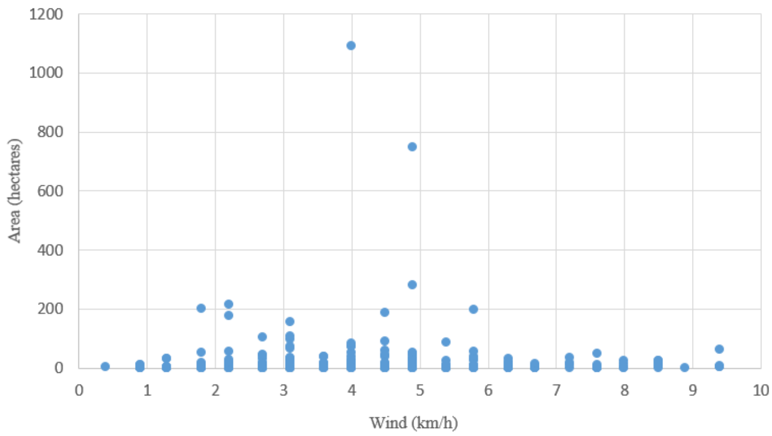

Figure 2(a). Relationship between wind (km/h) and burned area (hectares)

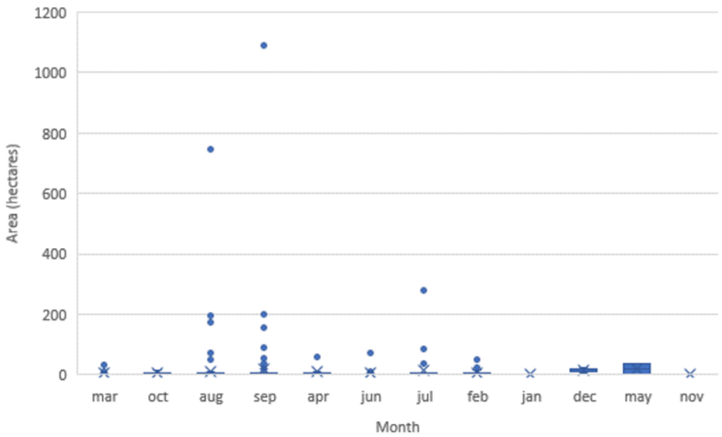

Figure 2(b). Relationship between the months of the fire and the burning area as hectare.

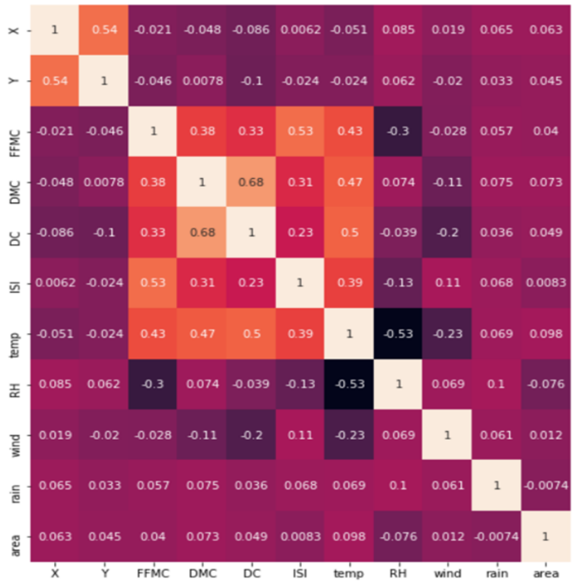

Figure 3. Correlation matrix of features for area estimation

2.2. WEKA

WEKA is a modular data mining tool produced by New Zealand's University of Waikato under a free GNU license [19]. It contains many methods, algorithms, libraries, and ready-made functions. Many features that are newly developed or not included with the standard program can be downloaded free of charge from the WEKA platform and can be integrated into the program as requested by the user. Since WEKA program is produced using Java, during development, its libraries have .jar extensions which provides convenience in the integration process of many programs produced with Java. Data preprocessing, classification, clustering, association and visualization can be done easily on the WEKA platform. In order to perform these operations, the extension of the file must be arff. However, the conversion of data in different extensions can be done easily. So the methods were tested in WEKA environment and 10 fold cross validation was chosen to trained the test pattern.

2.3. Linear Regression

Representing the relationship between two or more variables with a straight line is called Linear Regression [20]. Figure 4 shows an example linear regression model. The blue points describe the data, the red line defines the relationships between these variables. The closest linear result to the connection between the two variables is obtained. A line equation is obtained that will pass to cover as much of the sample data as possible which predicts future data [21].

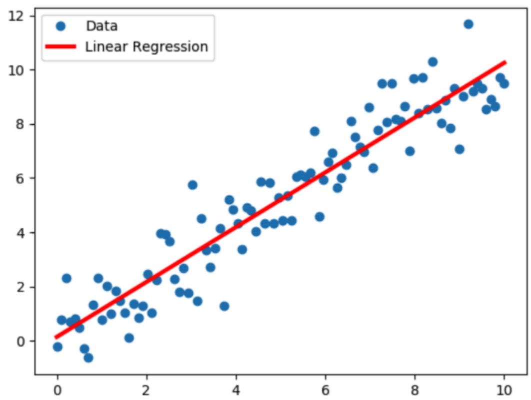

Figure 4. Example a linear regression model [22]

2.4. SVM

SVM [23-24] is frequently used in classfication problems which aims to find the line that has the maximum distance between the points placed on a hyper plane. In order to draw the border, two lines close and parallel to each other are drawn for both groups and these lines are brought closer together to produce the borderline. The most appropriate function is to try to be estimated for separating the data from each other [25]. If a linear boundary cannot be found for classification, the boundary is searched by moving the data to another multidimensional space. It is more common to use for complex small and medium sized datasets [26]. The presented dataset in this study is medium-sized.

2.5. Neural Networks

NN is a ML model built in layers[27-28]. MLP[29] is one of the neural network methods which is proposed in the study. It tries to automatically realize abilities such as deriving new information, creating and discovering through learning. Classification can be made using threshold values. MLP have an input layer, one or more hidden layers, an output layer, and transitions between layers called back-and-forward propagation. The input layer gets the data and sends it to the middle layer. Then this information is send to the next layer. The number of hidden layers is adjusted between at least and need. Each layer's output, becomes the input of the next layer. Thus, the output is reached[30].

2.6. KNN

KNN is one of the ML methods used for regression and classification in supervised learning[31-32]. It is considered the simplest machine learning algorithm. KNN is based on estimating the class of the vector formed by the independent variables of the value to be estimated, based on the information in which class the nearest neighbors are dense [33]. KNN makes predictions on two basic values such as distance and number of neighbors [34]. An example KNN model is shown with three classes in figure 5. In this study, the number of neighbors is selected as n=1, n=5, n=10 and n=50 respectively and the results were compared.

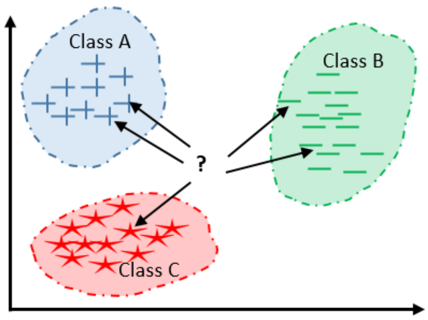

Figure 5. Example a KNN model [35]

2.7. Decision Tree

Decision tree is the tree-based algorithm which is commonly used in regression and classification problems [36]. It can be used in complex datasets [37]. The tree model is created by dividing possible decision groups into small sub groups by applying simple decision-making steps. The first node of the decision tree is the root and the other nodes connected to the roots are leaf nodes [38]. An example decision tree model is given in figure 6.

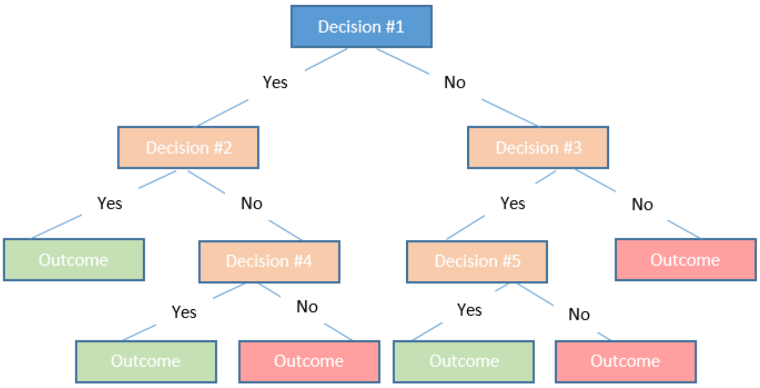

Figure 6. Example a Decision Tree model [39]

2.8. Stacked Regressor

Stacked regressor is used to combine multiple predictors which has been applied since a set of models performs better at burning areas at different places [40]. The principal concept of this technique is to assemble the complementary merits of multiple models to boost the total performance of the ensemble model [41]. Ensemble is one of the machine learning methods that combines the prediction results of more than one base model in order to obtain more powerful and generalizable results compared to a single model. An example stacked regressor model is given in figure 7. The figure illustrates that, the training data is tested on three different models in the first stage and predictions are obtained. At the second level, these predictions are generalized and expressed as a single output.

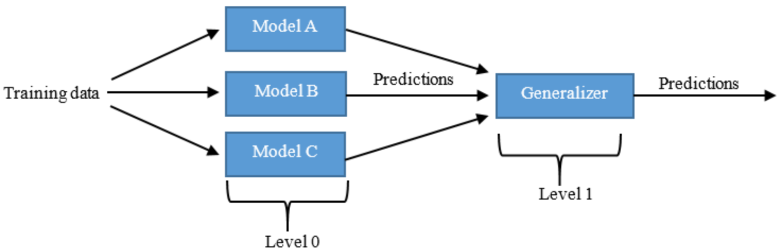

Figure 7. Example a Stacked regressor model [42]

3. Results and Discussion

The identified models were trained on the UCI dataset and the models were evaluated based on the MAE metric on the respective test sets. MAE is a performance metric calculated by dividing the sum of the absolute values of the differences between the actual and the predicted results by the total number of data. Since MAE is easily interpreted, it is frequently used in the fields of machine learning and artificial neural network. Table 2 shows the performance results which are obtained comparatively. The best value of the MAE is obtained with SVM as 12.8879 and the worst one is MLP is 38.7481. SVM has the best overall prediction ability among all models for the MAE metric in the dataset. However, when all the results are examined, it is seen that very high accuracy results are not achieved. This shows that the parameters used cannot predict fire sizes with high accuracy. It means that the correlations between the lit fields and the input parameters are weak. SVM performed better as the values in the UCI dataset were likely to be distributed over a small range. Considering the results in Table 2, the least accurate results are obtained from the MLP according to the MAE metric. This is thought to be due to the small amount of data and insufficient features.

Table 2. Model performances of according to the MAE metric

ML MODELS	MAE
Lineer Regression	20.0857
SVM	12.8879
NN	38.7481
KNN (N=1)	23.8073
KNN (N=5)	20.6596
KNN (N=10)	20.3174
KNN (N=50)	18.9255
Decision Tree	19.1364
Stacking	18.5918

In addition, the obtained results were compared with the results obtained in which was performed in Phyton[15]. The same data set and methods were used. Table 3 shows the comparison results. When Table 3 is examined, although the results obtained from the Linear regression, SVM, KNN, Stacking methods are compatible, the results obtained from the MLP and Decision tree methods are completely inconsistent. Our proposed approach was carried out in the WEKA program. For this reason, it is thought that one of the reasons for the differences in the results obtained may be the difference in the program used. In addition, it is thought that the other reason for the difference may be the difference in the data preprocessing stage. Finally, in the study of [15], more than one dataset was used, but in this study it was studied on a single dataset. For these reasons, it was concluded that there may be differences between the results obtained.

Table 3. Comparison of study results with [15]

	MAE (Our study)	MAE [15]
Linear Regression	20.0857	15.547
SVM	12.8879	6.334
NN	38.7481	8.264
KNN	18.9255	15.53
Decision Tree	19.1364	31.5
Stacking	18.5918	9.45

4. Conclusion

In this study, various machine learning methods were used to estimate fire size based on various characteristics such as temperature, wind, humidity and precipitation using 512 wildfire records that took place in a national park in Northern Portugal. These methods are Linear regression, SVM, MLP, KNN, Decision Tree and Stacking. Models were evaluated based on the MAE metric in the relevant test sets. It has been seen that the SVM method has the best predictive ability among all models for the MAE metric in the data set. According to the MAE metric, it was observed that the least accurate results were obtained from the MLP method.

In this study a dataset from Northern Portugal was used. For future work if a similar dataset belonging Turkey can be found then It will be able to possible to estimate the size of the burning area over the dataset with the alternative ML methods. In addition, the number of data and more parameters such as weather and environmental factors in the dataset used in this study can be increased to improve the prediction success.

References

[1] Tonini, M.; D'Andrea, M.; Biondi, G.; Degli Esposti, S.; Trucchia, A.; Fiorucci, P. A Machine Learning-Based Approach for Wildfire Susceptibility Mapping. The Case Study of the Liguria Region in Italy. Geosciences 2020, 10, 105.

[2] Le, H. V., Hoang, D. A., Tran, C. T., Nguyen, P. Q., Tran, V. H., Hoang, N. D., Amiri, M., Ngo, T. P., Nhu, H. V., Hoang, T. V., & Tien Bui, D. A new approach of deep neural computing for spatial prediction of wildfire danger at Tropical Climate Areas. Ecological Informatics, 2021, 63

[3] Jain, P., Coogan, S.C., Subramanian, S.G., Crowley, M., Taylor, S., & Flannigan, M.D. A review of machine learning applications in wildfire science and management. ArXiv, 2020,abs/2003.00646.

[4] S. Girtsou, A. Apostolakis, G. Giannopoulos and C. Kontoes,A Machine Learning Methodology for Next Day Wildfire Prediction, 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 8487-8490

[5] Liang Hç, Zhang M. and Wang H., "A Neural Network Model for Wildfire Scale Prediction Using Meteorological Factors," in IEEE Access, vol. 7, pp. 176746-176755, 2019

[6] Gholamnia, K.; Gudiyangada Nachappa, T.; Ghorbanzadeh, O.; Blaschke, T. Comparisons of Diverse Machine Learning Approaches for Wildfire Susceptibility Mapping. Symmetry 2020, 12, 604.

[7] Jonathan K., "Gradient boosting with extreme-value theory for wildfire prediction," arXiv, 2021.

[8] V. Zope, T. Dadlani, A. Matai, P. Tembhurnikar and R. Kalani, "IoT Sensor and Deep Neural Network based Wildfire Prediction System," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 205-208

[9] Pérez-Porras, F.-J.; Triviño-Tarradas, P.; Cima-Rodríguez, C.; Meroño-de-Larriva, J.-E.; García-Ferrer, A.; Mesas-Carrascosa, F.-J. Machine Learning Methods and Synthetic Data Generation to Predict Large Wildfires. Sensors 2021, 21, 3694.

[10] Valero, M. M., Jofre, L., & Torres, R. Multifidelity prediction in wildfire spread simulation: Modeling, uncertainty quantification and sensitivity analysis. Environmental Modelling & Software, 141, 2021.

[11] Cao Y., Yang F., Tang Q. and Lu X., An Attention Enhanced Bidirectional LSTM for Early Forest Fire Smoke Recognition. IEEE Access, vol. 7, pp. 154732-154742, 2019

[12] Bergado J. R, Persello C., Reinke K., Stein A. Predicting wildfire burns from big geodata using deep learning. Safety Science, 140, 2021.

[13] Qin L , Shao W. , Du G., Mou J. ve Bi R., Predictive Modeling of Wildfires in the United States. 2021 2nd International Conference on Computing and Data Science (CDS);2021 Stanford, pp. 562-567

[14] Beşli N. And Tenekeci M. Uydu verilerinden karar ağaçları kullanarak orman yangını tahmini. DÜMF Mühendislik Dergisi; 2020.

[15] Moore S. A. Wildfire Burn Area Prediction. 2019. 33rd Conference on Neural Information Processing Systems. Vancouver, Canada,.

[16] Rafaello Bergonse, Sandra Oliveira, Ana Gonçalves, Sílvia Nunes, Carlos DaCamara & José Luis Zêzere (2021) Predicting burnt areas during the summer season in Portugal by combining wildfire susceptibility and spring meteorological conditions, Geomatics, Natural Hazards and Risk, 12:1, 1039-1057, DOI: 10.1080/19475705.2021.1909664.

[17] Trucchia, A.; Meschi, G.; Fiorucci, P.; Gollini, A.; Negro, D. Defining Wildfire Susceptibility Maps in Italy for Understanding Seasonal Wildfire Regimes at the National Level. Fire 2022, 5, 30. https://doi.org/10.3390/fire5010030.

[18] Cortez, Paulo & Morais, A.. (2007). A Data Mining Approach to Predict Forest Fires using Meteorological Data.

[19] Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco (2005).

[20] Vetter TR, Schober P. Regression: The Apple Does Not Fall Far From the Tree. Anesth Analg. 2018 Jul;127(1):277-283.

[21] Seber, G. A.ve Lee, A. J., "Linear regression analysis", Vol. 329, John Wiley & Sons, 2012.

[22] Tran, Hieu. (2019). A survey of machine learning and data mining techniques used in multımedia system.

[23] Willsch D., Willsch M., De Raedt H. , Michielsen K., Support vector machines on the D-Wave quantum annealer. Computer Physics Communications, Volume 248, 2020, 107006, ISSN 0010-4655.

[24] Huang, Y., Zhao, L. Review on landslide susceptibility mapping using support vector machines. 2018. CATENA, 165, 520–529.

[25] Karakoyun, M. ve Hacıbeyoğlu, M.,"Biyomedikal Veri Kümeleri ile Makine Öğrenmesi Sınıflandırma Algoritmalarının İstatistiksel Olarak Karşılaştırılması", Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 2014.

[26] Support Vector Machine - Regression (SVR), http://www.saedsayad.com/support_vector_machine_reg.htm. 27.02.2022.

[27] Zhang, Y., Tuo, M., Yin, Q., Qi, L., Wang, X., & Liu, T. Keywords extraction with deep neural network model. Neurocomputing. 2020 383, 113-121.

[28] Zhang, G., Wang, M., & Liu, K. Forest fire susceptibility modeling using a convolutional neural network for Yunnan province of China. 2019. International Journal of Disaster Risk Science, 10(3), 386-403.

[29] Heidari, A. A., Faris, H., Mirjalili, S., Aljarah, I., & Mafarja, M. Ant lion optimizer: theory, literature review, and application in multi-layer perceptron neural networks. 2020 Nature-Inspired Optimizers, 23-46.
[30] Meha Desai, Manan Shah, "An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN)," Clinical eHealth, Volume 4, 2021, Pages 1-11, ISSN 2588-9141,https://doi.org/10.1016/j.ceh.2020.11.002.
[31] Abu Alfeilat, H. A., Hassanat, A. B., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., Eyal Salman, H. S., & Prasath, V. S. Effects of distance measure choice on k-nearest neighbor classifier performance: a review. 2019,Big data, 7(4), 221-248.
[32] Ali, N., Neagu, D., & Trundle, P. Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets. 2019. SN Applied Sciences, 1(12), 1-15.
[33] M. Toğaçar , "Detection of Phishing Attacks on Websites with Lasso Regression, Minimum Redundancy Maximum Relevance Method, Machine Learning Methods, and Deep Learning Model", Turkish Journal of Science and Technology, c. 16, sayı. 2, ss. 231-243, Eyl. 2021.
[34] Damien Chanal, Nadia Yousfi Steiner, Raffaele Petrone, Didier Chamagne, Marie-Cécile Péra, "Online Diagnosis of PEM Fuel Cell by Fuzzy C-Means Clustering", Reference Module in Earth Systems and Environmental Sciences, Elsevier, 2021, ISBN 9780124095489, https://doi.org/10.1016/B978-0-12-819723-3.00099-8.
[35] Atallah, Dalia & Badawy, Mohammed & El-Sayed, Ayman & Ghoneim, Mohamed. (2019). Predicting kidney transplantation outcome based on hybrid feature selection and KNN classifier. Multimedia Tools and Applications. 78. 20383–20407. 10.1007/s11042-019-7370-5.
[36] Yıldız, Olcay Taner, et al. Bagging Soft Decision Trees. Springer Verlag, 2017. EBSCOhost, https://doi.org/10.1007/978-3-319-50478-0_2.
[37] Jaafari, A., Zenner, E. K., & Pham, B. T. Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers. 2018. Ecological informatics, 43, 200-211.
[38] Altaş, D. & Gülpınar, V. (2012). Karar Ağaçları Ve Yapay Sinir Ağlarının Sınıflandırma Performanslarının Karşılaştırılması . Trakya Üniversitesi Sosyal Bilimler Dergisi , 14 (1) , 1-22 .
[39] Decision tree diagrams: what they are and how to use them, https://blog.mindmanager.com/blog/2021/05/11/decision-tree-diagrams/ 27.02.2022.
[40] Pavlyshenko, B. Using stacking approaches for machine learning models. 2018. 2018 IEEE Second International Conference on Data Stream Mining & Processing .255-258. IEEE.
[41] Kaibing Zhang, Shuang Luo, Minqi Li, Junfeng Jing, Jian Lu, and Zenggang Xiong. 2020. Learning stacking regressors for single image super-resolution. Applied Intelligence 50, 12 (Dec 2020), 4325–4341. DOI:https://doi.org/10.1007/s10489-020-01787-0
[42] Divina, F., Gilson, A., Goméz-Vela, F., García Torres, M., & Torres, J. (2018). Stacking Ensemble Learning for Short-Term Electricity Consumption Forecasting. Energies, 11(4), 949. doi:10.3390/en11040949.