Grenze International Journal of Engineering and Technology, Jan Issue

Prediction and Analysis of Forest Fire using Meteorological Parameters

Shaifali Bhatt[1] and Usha Chouhan[2]
[1]Research Scholar, Department of Mathematics, Bioinformatics and Computer applications, MANIT Bhopal, India
Email id: bhattshalu.93@gmail.com
[2]Associate Professor, Department of Mathematics, Bioinformatics and Computer applications, MANIT Bhopal, India,
Email id: ycchouhan@gmail.com

***Abstract*— Prediction, prevention, and control of forest fires are becoming increasingly crucial for dealing with wildfires on all scales. Developing more effective fire detection systems can aid in the control of environmental threats. For forest fire prediction, the presented study used a kernel ridge regression approach with fuzzy logic (random search) based on meteorological measurements (temperature, wind direction, relative humidity, etc), soil moisture indicators, and geographical data. For the scoring matrix, mean absolute error (MAE) and root mean square error (RMSE) are used, and cross-validation 10 folds is performed to tune the hyper parameters. The result obtained by the presented model has a mean absolute error (MAE) of 9.89 and root mean square error (RMSE) of 23.38, which is minimum as compared to the prior works which means minimum the value of mean absolute error (MAE) and root mean square error (RMSE) more accurate will be the model with high predictive power. This paper analyses kernel ridge regression's significant impact on forecasting forest fire compared with different models with advancements made in technique with the help of substantial parameters. This technique (kernel ridge regression) can aid in the control of environmental threats by effective fire detection systems.**

***Index Terms*— Machine Learning, Ridge Regression, Kernel Ridge regression, Forest fire prediction**.

I. INTRODUCTION

Fire is a significant ecological determinant that has shaped and disrupted numerous plant communities worldwide. Fire most likely manifested as a natural disruption as soon as terrestrial vegetation existed [1]. Prior to the arrival of humans, the primary sources of ignition were activity associated with earthquakes, volcanoes, and lightning [2–4]. Thus, the process of fire was natural and happened on a recurrent basis during the natural vegetation transition cycle, which contributes to the intelligent system's ongoing revitalization and improves the productivity of various plant groupings and ecosystems [5]. The frequency with which ecosystems burn is regulated by the fuel load needed to keep a fire going after a natural ignition.

Wildfires, particularly massive ones, are the product of two key drivers interacting: fuel supply and consistency, and weather patterns. Significant effort has been made in recent years to study historical and current fire regime trends and to assess the relative significance of the aforementioned key factors in determining current and historical fire regimes [6–9].Prediction, prevention, and control of forest fires are becoming increasingly crucial for dealing with wildfires on all scales. Forest fire hazard prediction systems are an important instrument for predicting forest

Grenze ID: 01.GIJET.10.1.339

fire risks, assisting in fire management planning and resource allocation, and backing up the forest fire surveillance and eradication phase [10].

Numerous research articles have been published in the literature that examine the prediction of burnt regions using various Data Mining approaches, artificial intelligence, and machine-learning methods. Several of them were created in[11–13] and are mostly based on neural networks, and their reliability for forecasting burnt regions was evaluated and analyzed. Numerous authors also employed Support vector machines, radial basis functions, genetic algorithms, and random forests to accomplish the task at hand [14–16]. In [17], an Australian case study was used to investigate and assess a strategy for forecasting the risk of wildfires using Bayesian networks and fuzzy logic.

This work employs a predictive regression model to anticipate the spreading of wildfire and the area at risk. As the regression technique, Kernel ridge regression algorithm with Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) as scoring metrics is used. This model was trained over the UCI repository forest fire dataset[18] and finally tested through 10-fold cross-validation.

The rest of this work is divided as follows: Section 2 describes the materials and methods utilized in the field of fire detection and provides an overview of the proposed framework and its components. Section 3 contains the outcomes of the numerous experiments. Section 4 offers more discussions of the collected results, and section 5 provides a conclusion and research suggestions for the future.

II. METHODS

A. Study Area

This study is based on fire information from the Montesinho Natural Park in the Trans-os-Montes area of north-eastern Portugal [19]. This park features a large variety of flora and fauna, as well as a Mediterranean environment with a yearly average temperature of between 8 and 12 degrees Celsius. Our model's data was collected between January 2000 and December 2003. There are 517 data in the database. It comprises 12 variables that were gathered over the course of three years. The forest Fire Weather Index [20], which is Canada's approach for estimating the hazard of wildfires, contains fuel and meteorological data relevant to the fires. It is made up of six parts. (i)FFMC is for Fine Fuel Moisture Code, (ii) DMC stands for Duff Moisture Code, (iii) DC stands for Drought Code (iv) FWI stands for Fire Behaviour Index (v) ISI stands for Initial Spread Index (vii) BUI stands for buildup index [21].Besides we also have temperature, wind, rainfall data, relative humidity, and burned area. We have a total of 12 variables, with the remaining being the x and y components of the location [22].The values of the data features described in Table 1.

B. Machine learning Algorithm

The widely used machine learning algorithm selected in this work is kernel ridge regression with linear kernel function which can work in a high dimensional environment and provide good predictive power.

C. Model Architecture:

This research suggests kernel ridge regression mechanism for learning efficient features extraction to forecast burned area. For training and testing, this work divided data collection in an 80:20 ratio as train and test data. 10-fold cross validation is applied on the dataset. The training is seeded with Kernel ridge regression having linear kernel to fine tune the learning rate.

D. Model Evaluation

TABLE 1 DATASET FOR FOREST FIRE PREDICTION.

SNo.	Variable	Description	Values
1	Y	y-axis coordinate	1 – 9
2	Month	Month of the Year	January - December
3	X	x-axis coordinate	1 – 9
4	RH	Relative humidity (in %)	15.0 -100
5	Wind	Wind speed(in km/h)	0.40 -9.40

6	Area	Total Burned area(in ha)	0.00 - 1090.84
7	Rain	Rain(in mm/m2)	0.0 - 6.4
8	ISI	Initial Spread Index	0.0 - 56.10
9	FFMC	Fine Fuel Moisture Code	18.7 - 96.20
10	DC	Drought Code	7.9 - 860.6
11	DMC	Duff Moisture Code	1.1 -291.3
12	Temp	Temperature (in C)	2.2 -33.30

The performances of the models were explored by applying a 10-fold CV (cross-validation) approach in this work. The model was trained with the help of training data, and the model-generated error was determined with the help of test data. This means the dataset was randomly divided into training and test subsets 10 times, the model was created from the training data, and the testing set error was determined. The mean absolute error (MAE) and root mean square (RMSE) errors are employed to assess the model's performance. Equations respectively define these error measurements.

$$\text{RMSE}=\sqrt{\frac{1}{M}\sum_{i=1}^{M}|a_i - \hat{a}_i|^2}$$

$$\text{MAE}=\frac{1}{M}\sum_{i=1}^{M}|a_i - \hat{a}_i|$$

Where $\hat{a}_i$ represents the predicted output, a_i is the output of the system, and M is the sample size. In this study, RMSE (root mean square error) and MAE (mean absolute error) are conventional accuracy measures for evaluating several techniques on the same dataset. The MAE (mean absolute error) assigned the mean deviations with the same weight, corresponding to a linear score. Since deviations were squared before being averaged, the RMSE (root mean square error) accorded a comparatively high weight to significant errors. This implies RMSE (root mean square error) was most advantageous when major errors were most undesired.

III. RESULTS

This section depicts the outcome of the present study, the performance of kernel ridge regression for predicting the burnt area of forest fires compared to that of other techniques used in recent works. For maximizing fire management efforts, it plays a crucial role to have models with good forecast capability.

A. Ambient conditions and model setup

For training and testing, this work divided data collection in an 80:20 ratio as train and test data. RMSE is sensitive to outliers in a dataset hence an outlier elimination approach was used on the dataset using Cook's Distance with a threshold of 4/n. Figure 2 Shows the outliers and threshold.

To reduce overfitting and bias, irrelevant features such as days were eliminated as the burnt area is similarly distributed for all days as shown in Figure 3. Different days of the week show no significant impact on the fire which day of the week has more area burnt by forest fire so these weeks were converted into months. Further distribution among months was transformed into seasons shown in Figure 4 as some specific months had less observation compared to others. Seasons show that in the fall and summer seasons the area burnt by the fire is more compared with the spring and winter seasons.

As confirmed by this paper [23], features like X, Y, DC, ISI, wind, month, and DMC were also dropped, and only 5 features were used: temperature, wind, relative humidity (RH), FFMC, and rain. But to improve the predictive power of the model in terms of RMSE, the month, which was dropped by the previous researcher, was used in the form of a season, so six features were used: temperature, wind, season, relative humidity (RH), FFMC, and rain. 10-fold cross-validation was done, as previously stated. The RMSE obtained by the Kernel ridge regression in this fold is 23.38 and MAE = 9.89, respectively.

In Figure 7, each point on the plot represents forest fire. Our model was able to predict large and small fires significantly well. Blue color dots depict the real burnt areas while orange color dots depict the burnt area observed by our model.

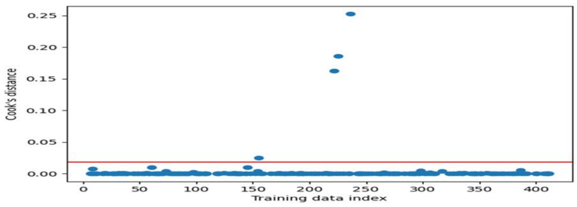

Figure 2: Outlier Detection and threshold value using relevant 6 features.

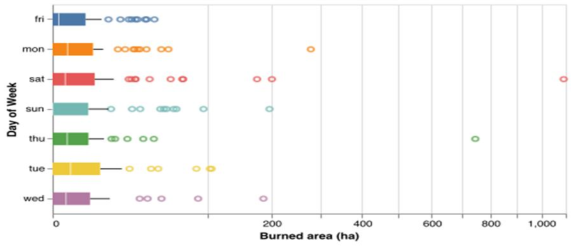

Figure 3: Burnt area distributed throughout different days of the week

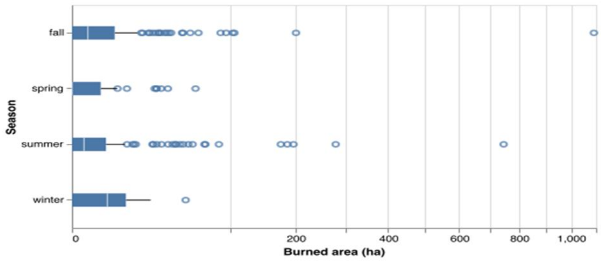

Figure 4: Burnt area distributed throughout different season.

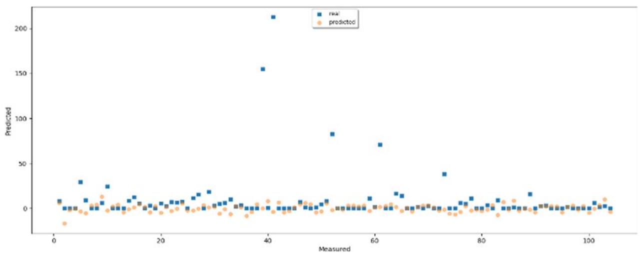

Figure 7: Predicted versus observed forest fire area for Kernel ridge regression model trained for 10 folds.

IV. DISCUSSION

To assess the predictive accuracy of these sorts of models, it is necessary to compare the outcomes with other machine learning and artificial intelligence (AI) techniques. This part displayed the results of other published studies pertaining to the same issue for this purpose. For this comparison, the root mean square error (RMSE) and mean absolute error (MAE) are used.In the literature, Table 2 demonstrates the errors of several models applied to identical data sets [22, 24, 25]. Here, we describe prior research that may be directly compared to the kernel ridge regression model used in this work, given that the data treatment, model construction, and model validation criteria were the same.Table 2 shows that Kernel Ridge Regression, with RMSE = 23.38 and MAE = 9.89, has the lowest error for RMSE and MAE compared with other published techniques. In comparison to the other models in the table, this means that the Kernel Ridge Regression models have greater prediction accuracy.The difference between the RMSE value of the Kernel Ridge Regression model and other models is found to be significant. This means that the Kernel Ridge Regression model was better able to handle outliers than the MAE model.Within its internal model identification structure, Kernel Ridge Regression executed a feature selection procedure, which we believe contributed to the improvement is achieved. Kernel Ridge Regression was used to determine that just 6 of the 12 variables (temperature, wind, season, relative humidity (RH), FFMC, and rain) were employed to derive the model with the maximum predictive power for this problem.

TABLE II MAE (MEAN ABSOLUTE ERROR) AND RMSE (ROOT MEAN SQUARE ERROR) ERRORS OBTAINED BY DIFFERENT MODELS.

Method	**Parameters**	**MAE**	**RMSE**
Decision Tree(DT) [22]	Reduction of the sum of squares	13.18± 0.05	64.5± 0.0
Multiple Regression(MR) [22]	Least square Algorithm	13.01± 0.0	64.5± 0.0
Neural Network(NN) [22]	100 epochs; 3 times; MLP 1HL; linear/logistic; BFGS alg.; Hidden Lay. = 4 neurons	13.71± 0.69	66.9± 3.4
Support Vector Machine(SVM) [22]	Sequential Minimal Optimization alg.; C = 3; Gamma = 2−5; RBF kernel	12.71± 0.01	64.7± 0.0
Random Forest(RF) [22]	T=500	12.93± 0.01	64.4± 0.0
Cascade Correlation Network(CCN) [25]	Hidden Lay. = 2 neurons; Num. Layers=3; Gaussian /Sigmoid	551.4	62.6
Multi-layer Perceptron Neural Network [25]	Num. Layers = 3; logistic/linear; Hidden Lay. = 1 neuron	617.4	63.1
Polynomial Neural Network [25])	Gaussian kernel; p1 = 31.30; p2 = −27.30; p4 = 6.30; p3 = 1.13	516.5	63.2
Radial base function Neural Network [25]	Min. Rad. = 0.01; Num. Neur. = 100; Max. Rad. = 519.669; Min. Lambda = 0.01328; Maximum Lambda = 9.95337	911.2	54.2
Support Vector Machine(SVM) [25]	RBF kernel; C = 84.18; Epsilon = 0.001; Gamma = 3800.2	282.4	54.0

Adaptive Neuro-Fuzzy Inference System(ANFIS) [23]	Hybrid alg.; constant func.; 50 epochs	13.02± 0.04	64.6± 0.0
Fuzzy Inductive Reasoning(FIR) [23]	EWP-EFP; 2–3 FS per var	11.93± 0.01	48.9± 0.0
Kernel Ridge Regression(proposed)		9.89± 3.56	23.38± 10.8

V. CONCLUSION

Forest fires are a common occurrence in both wildlife fauna and flora. Every year, millions of hectares of forest are lost all over the world. Forest fires are a severe environmental concern that endangers forest preservation and damages the economy and ecology. This paper implements the Kernel ridge regression, a method for predicting forest fires using real meteorological parameters like temperature, wind, season, relative humidity (RH), FFMC, and rain. Furthermore, to improve the predictive power of the model, the parameter month which was dropped by the previous researcher [23] was added and it has been verified that using the season as a new parameter the performance of the model has been increased and it will have a significant impact for predicting forest fire. The Proposed Kernel ridge regression model had shown significantly good predictive power with minimal RMSE (root mean square error) and MAE (mean absolute error) values. Minimum the value of RMSE (root mean square error) and MAE (mean absolute error) more the model is accurate and has high predictive power. This technique (kernel ridge regression) can aid in the control of environmental threats by effective fire detection systems.
Kernel ridge regression is used as a feature selection process to determine which parameters are relevant for high predictive power, out of 12 parameters only temperature, wind, season, relative humidity (RH), FFMC, and rain were relevant. The result obtained was compared with other machine learning techniques applied to the same dataset as shown in Table 2. The model thus obtained yielded a better result than most of the recently proposed solutions to this problem. Future studies in this work can employ a combination of filters and feature categories, or extract deep features as well.

DECLARATIONS

Competing interests: No financial interests
*Author's contributions***:** Shaifali Bhatt: Conceptualization, Methodology, Done the main research work .
Funding: No funding
Availability of data: The data set used has been obtained from the UCI machine learning repository, https://archive.ics.uci.edu/ml/datasets/forest+fires

BIBLIOGRAPHY

[1] Pausas JG, Keeley JE (2009) A burning story: The role of fire in the history of life. Bioscience 59:593–601
[2] Edwards D, Booysen Pd, Tainton NM (1984) Ecological Effects of Fire in South African Ecosystems. Springer-Verlag 19–37
[3] Kruger LM, Midgley JJ, Cowling RM. (1997) Resprouters vs reseeders in South African forest trees; a model based on forest canopy height. Funct Ecol 11:101–105
[4] Whelan RJ. (1995) The Ecology of Fire. Cambridge University Press
[5] Moreno JM, Oechel WC (1994) The Role of Fire in Mediterranean Type Ecosystems. Springer-Verlag 107:1–15
[6] Dimitrakopoulos AP, Vlahou M, Anagnostopoulou CG (2011) Impact of drought on wildland fires in Greece: Implications of climate change? Clim Change 109:331–347
[7] Pausas JG, Fernandez-Munoz S (2012) Fire regime changes in the Western Mediterranean Basin: From fuel-limited to draught-driven fire regime. Climatic Change 110:215–226
[8] Koutsias N., Xanthopoulos G., Founda D., et al (2013) On the relationships between forest fires and weather conditions in Greece from long-term national observations (1894-2010). Int J Wildland Fire 22:493–507

[9] Turco, Marco, Bedi, et al (2016) Decreasing fires in mediterranean Europe. PLoS One 11:. https://doi.org/10.1371/JOURNAL.PONE.0150663
[10] Alonso-Betanzos A.;, Fontenla-Romero O.;, Guijarro-Berdi˜nas B.;, et al (2003) An intelligent system for forest fire risk prediction and fire fighting management in Galicia. Expert Syst Appl 25:545–554
[11] Liang H, Zhang M, Wang H (2019) A neural network model for wildfire scale prediction using meteorological factors. IEEE Access 7:176746–176755
[12] Al_Janabi S, Al_Shourbaji I, Salman MA (2018) Assessing the suitability of soft computing approaches for forest fires prediction. Applied Computing and informatics 14:214–224
[13] Gholamnia K, Gudiyangada Nachappa T, Ghorbanzadeh O, Blaschke T (2020) Comparisons of Diverse Machine Learning Approaches for Wildfire Susceptibility Mapping. Symmetry 604:
[14] M. Naderpour, H. M. Rizeei, F. Ramezani (2020) Wildfire Prediction: Handling Uncertainties Using Integrated Bayesian Networks and Fuzzy Logic. In: Environmental Science 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
[15] Jaafari A, Pourghasemi HR (2019) Factors Influencing Regional-Scale Wildfire Probability in Iran: An Application of Random Forest and Support Vector Machine. Spat. Model. GIS R Earth Environ Sci 607–619
[16] García-Llamas P, Suárez-Seoane S, Taboada A, et al (2019) Environmental drivers of fire severity in extreme fire events that affect Mediterranean pine forest ecosystems. Ecol Manag 433:24–32
[17] Naderpour M, Rizeei HM, Ramezani F (2020) Wildfire Prediction: Handling Uncertainties Using Integrated Bayesian Networks and Fuzzy Logic. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, UK. pp 1–7
[18] UCI Machine Learning Repository: Forest Fires Data Set. https://archive.ics.uci.edu/ml/datasets/forest+fires. Accessed 9 Sep 2021
[19] UCI Machine Learning Repository: Forest Fires Data Set. https://archive.ics.uci.edu/ml/datasets/forest+fires. Accessed 18 Sep 2021
[20] Júnior JSS, Paulo JR, Mendes J, et al (2022) Automatic forest fire danger rating calibration: Exploring clustering techniques for regionally customizable fire danger classification. Expert Syst Appl 193:. https://doi.org/10.1016/J.ESWA.2021.116380
[21] Taylor SW, Alexander ME (2006) Science, technology, and human factors in fire danger rating: The Canadian experience. Int J Wildland Fire 15:121–135. https://doi.org/10.1071/WF05021
[22] Cortez P, Morais A (2007) A Data Mining Approach to Predict Forest Fires using Meteorological Data. In: —Portuguese Conference on Artificial Intelligence, Guimarães, Portugal. pp 512–523
[23] Nebot À, Mugica F, Pellizzaro G, et al (2021) Forest Fire Forecasting Using Fuzzy Logic Models. Forests 2021, Vol 12, Page 1005 12:1005. https://doi.org/10.3390/F12081005
[24] Nebot À, Mugica F (2021) Forest fire forecasting using fuzzy logic models. Forests 12:. https://doi.org/10.3390/f12081005
[25] Al_Janabi S, Al_Shourbaji I, Salman MA (2017) Assessing the suitability of soft computing approaches for forest fires prediction. https://doi.org/10.1016/j.aci.2017.09.006