Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

PREDICTION OF OCCURRENCE OF FIRES IN FOREST

. K L Ganapathi Reddy [1*], P Yamini Sri Chandrika [2], Shaik Masthan Vali [3], D Nishanth [4]

*[*1] Asst. Professor Dept. of CSE, Raghu Engineering College, Vishakhapatnam, Andhra Pradesh, India*

*[*2,3,4] Students Dept. of CSE, Raghu Engineering College, Vishakhapatnam, Andhra Pradesh, India*

Ganapathi.karri@raghuenggcollege.in, 19981a05c6@raghuenggcollege.in, 19981a05f0@raghuenggcollege.in, 19981a05i5@raghuenggcollege.in

ABSTRACT: Predicting Forest Fires heads a major role in managing, preventing fires. These fires are a significant environmental issue that causes ecological devastation by endangering natural landscapes, disrupting the balance of ecosystems, increasing the likelihood of other natural disasters, and depleting resources such as water. This not only contributes to global warming but also pollutes water sources, thereby exacerbating the problem further. Accurately predicting forest fires is crucial in managing them, enabling effective allocation of resources and facilitating mitigation and recovery efforts. Our project aims to develop and analyse artificial intelligence-based strategies for predicting forest fires, specifically utilizing decision trees and random forest algorithms. These algorithms rely on past weather conditions to determine the likelihood of a forest fire occurring.

Keywords : Forest Fire, Decision Tree, Classification, Regression, Random Forest.

INTRODUCTION: Forest fires pose a significant threat to the environment, property, and human lives, making it imperative to detect and respond to them promptly. The rise in global temperatures is a primary contributor to forest fires, alongside factors such as lightning, thunderstorms, and human carelessness. To predict and manage forest fires, simulation tools have become popular, but they face challenges related to accuracy and speed.[1] Machine learning, a branch of computer science, offers a promising approach to forest fire prediction. This involves using supervised, unsupervised, and reinforcement learning algorithms such as SVM (Support Vector Machine), Regression, ANN (Artificial Neural Network), Decision trees. These methods utilize various data sources, including laboratory experiments and simulations, to specify and forecast the spread of fires in different locations.

The major goal of forest fire forecasting is to properly allocate resources and to assist the fire management team's firefighters as much as possible.[2]

Weather factors have the greatest impact on fire danger in a given location. The signals of high fire risk on the land are carefully analysed to anticipate the chance of a fire developing in a location. These indications are derived from local sensors and combined at the nearest meteorological stations.

Every year, fires damage enormous tracts of land, emitting more carbon monoxide than all the vehicles on the road combined. These wildfires caused extensive damage and have devastated many regions.[3]

Early detection of fires and monitoring of potentially dangerous locations can significantly minimise response times, as well as the possibility of damage and fire-fighting expenses.

LITERATURE SURVEY:

The Forest Fire System prediction involves steps such as the collection of datasets, meteoro-logical conditions, spatial data, Population Census, Forest fire cells and picture classification using Machine learning and extraction of data of occurrences of Fire.

In 2010, George E and his team proposed a methodology for studying forest fire prediction approaches based on artificial intelligence. They used support vector machines as the foundation for their forest fire risk forecast method and applied it to data from Lebanon. The results demonstrated that the system could effectively estimate the risk of fire. [1]

Nizar Hamadeh Laris EA (2015) did research on forecasting forest fires in Lebanon by taking temperature, relative humidity, and wind speed into account. Artificial Neural Networks (ANNs) were used to adapt to these criteria and make predictions about forest fires.[2]

Divya T. L. (2015) proposed an image mining technique to predict the spread of forest fires by analyzing a series of pixel values in satellite images. The model utilizes satellite photos to forecast forest fires.[3]

Multivariate linear regression can be a powerful tool for predicting forest fires, as it allows for the examination of multiple variables and their relationship to the outcome variable. The use of different methods for computing the regression coefficients, such as least-squares, Gauss-Seidel and Gauss-Jordan can help to optimize the model's performance and accuracy. Overall, the study by Mukhammad Wildan Alauddin highlights the potential of linear regression for forest fire prediction and emphasizes the importance of selecting appropriate methods for computing regression coefficients.[4]

Wiering and Dorgio aimed to find optimal locations for cutting fire-lines to mitigate the impact of forest fires. Their approach involved developing a simulator to predict the spread of fire and identify effective decision-making strategies. To construct the simulator, they relied on two critical parameters: fuel-type and wind speed, which were used to build the spread index.[5]

Clarke employed cellular automata to predict the spread and extinction of wildfires by utilizing fractal geometry.[6]

Muzy and colleagues created a fire spread prediction system using a cellular automation model coupled with Rothermel's physical model of fire spread. [7]

Dunn and Milne developed a new approach to address the challenge of modeling terrains with heterogeneous structures. Previous techniques required these terrains to be divided into sub-terrains and modeled individually.[8]

Cheng and Wang utilized spatiotemporal data mining methodologies to develop a model that is specifically designed for predicting the burnt area in the event of a forest fire.[9]

Another, more contemporary approach used data mining and meteorological data to anticipate forest fires and estimate the burnt area caused by the fire. As a result, they determined that the most important factors were oxygen level, temperature, humidity, and wind speed. Among the data mining tools used were SVM (Support Vector Machine), MR (Multiple Regression), Random Forest, NN (Neural Networks), and Decision tree.

METHODOLOGY

Forest fires have become a global concern, with numerous countries finding it difficult to prevent and manage them. The severity and frequency of forest fires have surged in recent times, resulting in extensive damage to forests, wildlife, and even human settlements.

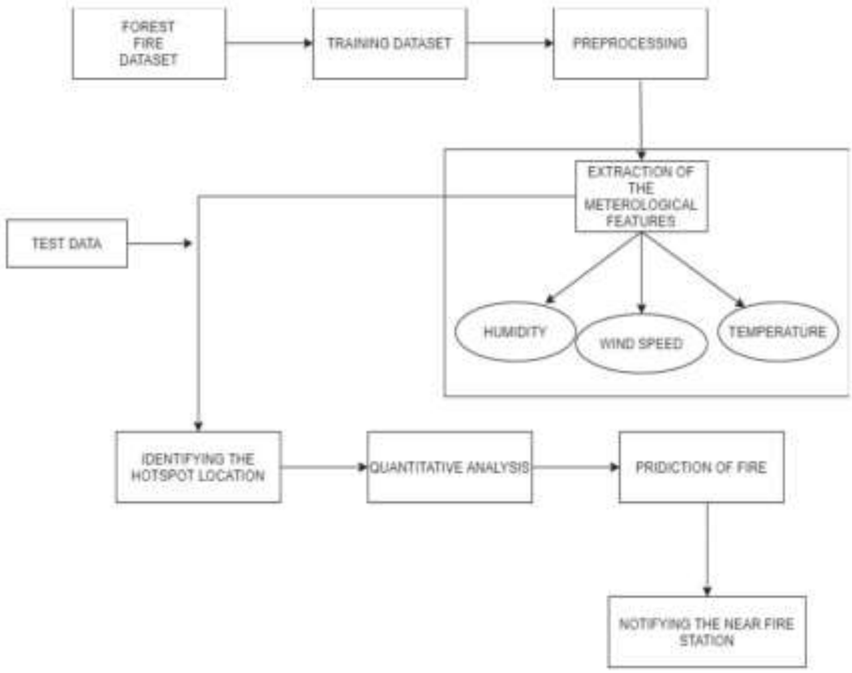
The random forest is an ensemble learning strategy that improves prediction accuracy while decreasing the likelihood of overfitting. This method employs multiple decision trees and combines their output

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

to make a final prediction. To accomplish this, it creates a collection of decision trees using random samples from both the training data and features. Forecasts from these trees are then blended using either a majority vote or the average.

For predicting forest fires, a combination of decision trees and random forest algorithms can be used. The model is trained on historical data of forest fires and environmental factors named as oxygen, temperature, humidity, wind-speed, and precipitation. By analyzing these factors, the model can provide predictions of the likelihood and severity of forest fires based on current also future environmental conditions. The decision tree identifies the most significant features that contribute to fire risk, while the random forest helps to enhance the prediction accuracy and generalization ability of the model. This combined approach can result in a highly accurate and robust model for predicting forest fires.

3.1. Research Design



3.2 Data Set

For this study, we used a publicly available dataset from Kaggle in the format of csv files containing information about previous occurred forest fire information. The dataset consists of 517 instances with 10 features, including meteorological features like oxygen, temperature, humidity, wind-speed, sea level, weather conditions, precipitation, solar radiation, ecosystem imbalance where the fire can occur in those conditions. The target variable for the study is mainly changing in levels of the features, which indicates whether the fire can occur or not. In this context we design a web page that contains Home, About, user registration and login details. After login successfully we have to upload datasets that are downloaded from Kaggle.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

Area	Oxygen	Temperature	Humidity	Fire Occurrence
Balaghat	30	33	35	0
Barwani	54	43	22	1
Betul	69	43	22	1
Burhanpur	28	26	78	0
Chhatarpur	54	43	22	1
Chhindwara	27	28	27	0
Damoh	35	30	28	0
Dewas	64	43	26	1
Guna	49	30	33	0
Jabalpur	65	44	32	1
Khandwa	71	46	26	1
Katni	38	29	34	0
Mandla	63	46	30	1
Morena	57	39	19	1
Panna	29	27	31	0
Rison	22	25	30	0
Sagar	67	38	28	1
Satna	71	41	29	1
Seoni	53	39	17	1
Cachar	66	40	21	1
Dimadlasao	42	33	47	0
Karbianglong	28	27	33	0
Tinsukia	56	38	21	1
Sonitpur	48	39	22	1
Sibsagar	48	38	19	1
Naogaon	64	35	12	1
Kokrajhar	26	28	30	0
Karimganj	71	39	24	1
Kamrup	63	36	25	1
Jorhat	26	24	33	0

To analyse the dataset, we used the PyCharm community environment with several Python libraries, including pandas, NumPy, Matplotlib, and scikit-learn.

3.3 Data Pre-Processing

Analysis: When we refer to data, we often imagine large databases with numerous rows and columns. However, data can come in various formats, including structured tables, images, audio files, videos, and more. Therefore, it is not uncommon for data to exist in different forms beyond the traditional database format.

Data pre-processing: Data processing comprises a series of steps for verifying, arranging, converting, amalgamating, and extracting data to generate an appropriate output. It is imperative to thoroughly document the processing methods to ensure data's veracity and usefulness. Since our dataset involves text data in a few columns, the foremost task is to convert this textual data into numerical format. To divide the dataset into training and testing subsets, the 'train_test_split' function from the 'sklearn' library can be employed. The correlation matrix was created to demonstrate the link between meteorological variables such as relative humidity, wind speed, temperature, and rainfall in the pre-processed data. This correlation analysis helps in understanding the dependence between two variables and how they move together. The results indicate that Oxygen level, temperature and humidity are less correlated, as the correlation value between them is negative. A positive value indicates a positive correlation, whereas a negative number indicates a negative correlation. Therefore, the negative value of correlation suggests that oxygen level, temperature and humidity are less correlated. One method for doing this is one-hot encoding, which assigns numeric values to each category. In this project, categorical data such as months and days were converted into numerical data using one-hot encoding. After encoding, the data was split into training and testing datasets in a 75:25 ratio. A Random Forest Regression model was then fitted to the dataset, and the results were predicted.

Algorithms Used

1.Random Forest Classifier: Random Forest is a machine learning method used to address classification and regression problems. It employs ensemble learning, which combines multiple classifiers to solve complex problems. The algorithm derives its output based on the predictions of decision trees. The results from different trees are averaged or combined to provide predictions. The accuracy of the predictions improves as more trees are added. Random Forest eliminates the limitations of a decision tree algorithm and reduces overfitting of datasets, which in turn improves precision.

2. Decision Tree Classifier: A decision tree is a type of supervised learning algorithm utilized in machine learning and data mining. It provides a graphical illustration of all potential solutions to a decision based on a set of conditions. A decision tree comprises nodes, edges, and leaves. Each node signifies a decision, whereas each edge denotes a condition or decision rule. The leaves of the tree depict the ultimate outcome or classification.

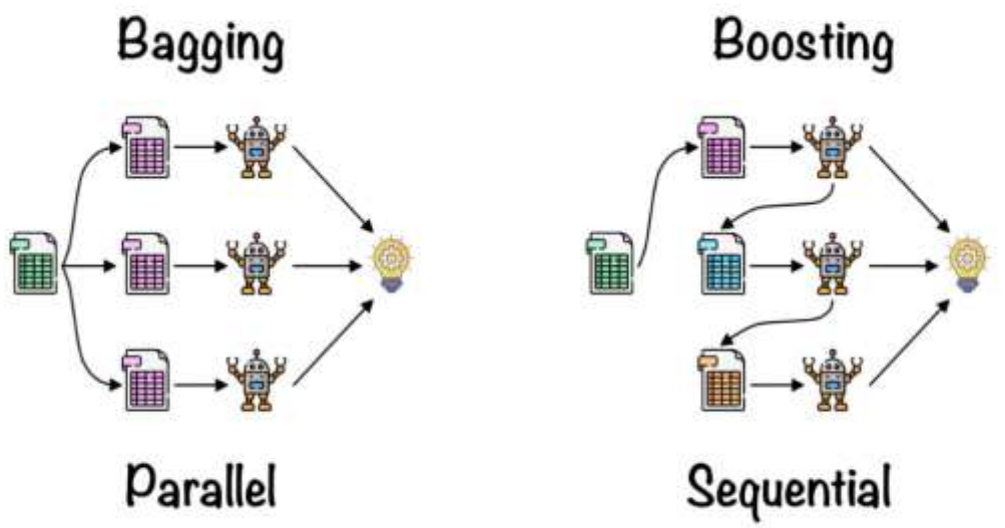
Decision trees can be applied to both classification and regression tasks. In classification, the decision tree is utilized to classify an input into one of several predetermined categories. On the other hand, in regression, the decision tree is utilized to predict a numerical value.

Decision trees are easy to understand and interpret, making them a popular choice for both beginners and experts in machine learning. They are also highly scalable and can be used with large datasets. However, decision trees can be prone to overfitting, which can be mitigated by pruning the tree or using ensemble methods like random forests or boosting.

Working of Random Forest Algorithm

To comprehend the workings of a random forest, it is important to the first grasp the concept of ensemble technique. Essentially, ensemble technique refers to the merging of several models. This means that instead of relying on a single model to make predictions, a set of models is utilized.

1. **Bagging-** Random Forest utilizes the Bagging principle, which involves creating a unique training subset by randomly sampling with replacement from the original training data. The final output is then determined based on the majority vote of the individual models. To delve deeper into the Bagging principle, it is an ensemble technique employed by Random Forest, also referred to as Bootstrap Aggregation. This technique selects a random sample from the dataset.



Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

Mathematical Concepts

To assess the effectiveness of a machine learning algorithm for earthquake prediction, it is common practice to divide the dataset into training and testing sets. The algorithm is trained on the training set and assessed on the testing set using different evaluation metrics, including accuracy, precision, recall, and F1-score. Additionally, Mean Absolute Error and Root Mean Squared Error can be computed to determine the algorithm's predictive performance.

Accuracy is the proportion of correctly classified samples, and it's computed as:

Accuracy = (True Positives + True Negatives) / (True Positives + False Positives + True Negatives + False Negatives)

Mean Absolute Error (MAE) measures the average magnitude of the errors in the predicted values compared to the actual values. It's computed as:

$$MAE = (1/n) * \sum|yi - \hat{y}i|$$

Where yi is the actual value, ŷi is the predicted value, and n is the number of samples. Root Mean Squared Error (RMSE) is similar to MAE but gives more weight to larger errors. It's computed as:

$$RMSE = sqrt((1/n) * \sum(yi - \hat{y}i)\char`^2)$$

Where yi is the actual value, ŷi is the predicted value, and n is the number of samples.

It should be emphasized that factors such as class imbalance, dataset size, and the selection of hyperparameters can influence the accuracy of these evaluation metrics. Hence, it is crucial to meticulously assess the performance of machine learning algorithms in predicting forest fires and to compare the outcomes produced by different algorithms and evaluation metrics.

Result Analysis and Discussions:

The data collected has been used for training and prediction of forest fires using various attributes such as oxygen, temperature, humidity, rain, and wind speed. The regression techniques employed include Random Forest (RF) and Decision Tree (DT), which were implemented using the Python platform. Comparison research was done to examine the model training and testing outcomes, with accuracy and MAE (Mean Absolute Error) determined for both the RF and DT models. MSE (Mean Square Error) is a metric used to determine how near a regression curve is to a collection of points by squaring the distances between the points and the curve, giving greater weight to bigger variances. Variance, on the other hand, quantifies how much the observed values differ from the mean of the expected values, with a low number being preferred. The variance graphs of the models were investigated prior to standardisation.

Mean Square Error was experimented with using a random forest classifier, and the results are as follows:

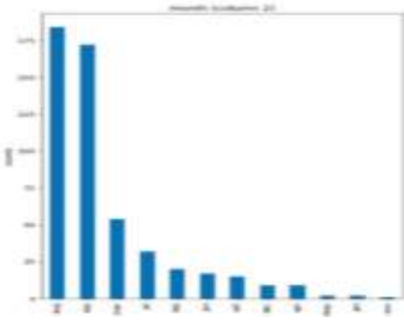
- Mean Absolute Error (MAE) was 0.04
- Mean Squared Error (MSE) was 0.01
- Root Mean Squared Error (RMSE) was 0.11.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

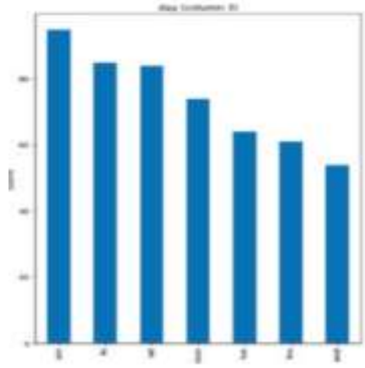
From the experiment Random Forest Classifier Algorithm can get 96.7% accuracy

Exploratory Analysis Results:

We have plotted a column graph to display the distribution of values in the "month" column against the "count" variable. The graph includes only those values which have between 1 and 50 unique occurrences.



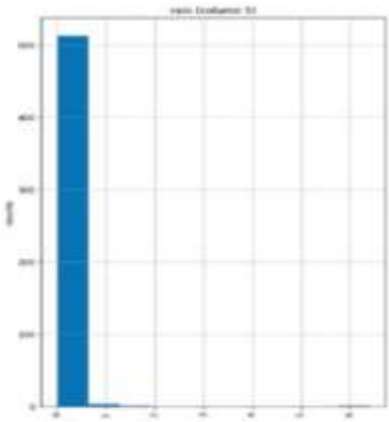
We have plotted a column graph to display the distribution of values in the "day" column against the"count" variable. The graph includes only those values which have between 1 and 50 unique occurrences.



Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

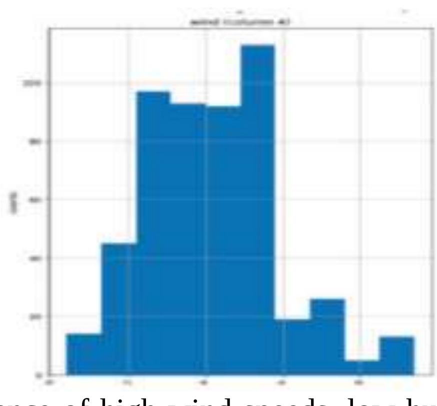
We have plotted a column graph to display the distribution of values in the "wind" column against the "count" variable. The graph includes only those values which have between 1 and 50 unique occurrences.

We have plotted a column graph to display the distribution of values in the "rain" column against the "count" variable. The graph includes only those values which have between 1 and 50 unique occurrences.



Conclusion And Future Scope

The study analyzed the factors that affect the frequency of forest fires and utilized data mining techniques for fire prediction purposes. The experiment considered three meteorological variables, namely temperature, relative humidity, and wind speed, which were found to significantly increase the



risk of burning in the presence of high wind speeds, low humidity, and extreme temperatures. The study also revealed that forests have a higher risk of fire compared to other surface areas. The study's findings may be utilised to determine the number of training examples and evaluation instances needed for fire prediction. The study can be further expanded to improve the models and their effects. Additionally, a real-time performance can be integrated into the application's UI, where the user can enter their local and zip code to retrieve meteorological data for a specific day, including the maximum and minimum temperatures, humidity levels, wind speeds, and other relevant data, which can be used as inputs for the fire prediction model.

References

[1] "An intelligent system for forest fire risk prediction and firefighting management in Galicia" by M. Inmaculada Paz Andrade, O. Fontenla-Romero, E. Jim´enez, B. Guijarro-Berdi nas, A. Alonso-Betanzos, T. Carballas, J. LuisLegido Soto and E. Hern´andez-Pereira in Expert systems with applications, vol. 25, no. 4, pp. 545–554, 2003, march 18 2023.

[2] “Introduction to Hilbert Space Theory. Research [sic] Foundation” by N. Aronszajn, Stillwater, Oklahoma, 1950, march 20 2023.

[3] "Applications of spatiotemporal data mining and knowledge for forest fire" by T. Cheng and J. Wang, in ISPRS Technical Commission VII Mid-Term Symposium, Enschede, 2006, pp. 148-153, march 21 2023.

[4] "Integrated Spatiotemporal Data Mining for Forest Fire Prediction" Transactions in Geographic Information Science, vol. 12, no. 5, pp. 591-611, 2008, march 21 2023

[5] “A cellular automaton model of wildfire propagation and extinction” by, J. Brass, P. Riggan, K. Clarke, Photogrammetric Engineering and Remote Sensing, vol. 60, no. 11, pp. 1355–1367, 1994, march 22 2023.

[6] “A review of AVHRR-based active fire detection algorithms: Principles, limitations, and recommendations” by A. Trishchenko, L. Giglio, C. Ithoku, Z. Li, Y. Kaufman, X. Yu, R. Fraser, Global and Regional Vegetation Fire Monitoring from Space, march 23 2023.

[7] “Statistics Based Predictive Geo-Spatial Data Mining: Forest Fire Hazardous Area Mapping Application” by K. Ryu, Y. Yeon, J. Han and K. Chi Lecture notes in computer science, pp. 370–381, 2003, march 23 2023.

[8] “Forest fire risk zone mapping from satellite imagery and gis” by D. Kumaran, R. Jiaswal, S. Rajesh and M. Saumitra in International Journal of Applied Earth Observation and Geo-information, vol. 4, pp. 1–10,2002, march 25 2023.

[9] “A semi-automated object-oriented model for burned area mapping in the Mediterranean region using Landsat-TM imagery” International Journal of Wildland Fire, vol. 13, no. 3, pp. 367–376, 2004, march 25 2023.

[10] “An object-oriented environment applied to a semi-physical model of fire spread across a fuel bed” by P. Santoni, A.Muzy, J. Balbi, J. Santucci, T. Marcelli and A. Aiello in Actesde la conf´erence ESS 2001 conference, 2001, pp. 641–643, march 26 2023.

[11] “Fire detection based on video processing method” by Ti& Nguyen-Phuc, Tuan, Hong, Nguyen (2013) 106-110. 10.1109/ATC.2013.6698087.

[12] “Computer vision techniques for forest fire perception” by J. Ramiro &Arrue, Begoña&Ollero, Luis & GómezRodríguez, Anibal& Merino, Martinez-de Dios, Francisco. (2008). 26. 550- 562.

[13] “Forest Fire Detection Using a Rule-Based Image Processing Algorithm and Temporal Variation” by M. and Ren, Mahmoud, 2018 Mathematical Problems in Engineering, 2018, pp.18.

[14] “Forest Fires Detection in Digital Images Based on Color Features” by Eva Tuba, Viktor Tuba, Romana Capor-Hrosik. (2017) International Journal of Environmental Science, 2, 66-70.