

FOREST FIRE PREDICTION USING MACHINE LEARNING

Koruprolu Bhargavi¹, P Lakshman Kumar², P Satya Charan³

Dr. Sanivarapu Prasanth Vaidya⁴

^{1,2,3}UG Student, Department of CSE, Aditya Engineering College, Surampalem, A.P., India

⁴Associate Professor, Department of CSE, Aditya Engineering College, Surampalem, A.P., India

ABSTRACT

Over the last few decades, deforestation and climate change have caused an increasing number of forest fires. This is a serious environmental issue that threatens the landscape's natural resources, disrupts the ecosystem's stability, raises the risk of other natural disasters, and depletes resources like water, which contributes to global warming and water pollution. forest blaze Controlling Forest fires relies heavily on prediction. Forest fire forecasting is predicted to lessen the impact of forest fires in the future. In order to model fire threat using data, this study uses data-mining techniques regarding, topography, climate, and some other fire-danger influencing factors. The classification techniques used in this work include Support Vector Machine, Decision Tree, and Random Forest (RF).By managing land use, planning, and making wise decisions, forest managers can utilize this process to identify the places that pose the greatest fire risk and prevent future fires..

Keywords: Machine Learning; Support Vector machine, Random Forest and Decision Tree.

1. INTRODUCTION

The technique of detecting patterns, correlations, and anomalies within big data sets is known as machine learning. The Data mining technique can handle the numerous parameters like temperature, slope, latitude, longitude, rainfall that is responsible for the fire in the forest. One of the key reasons for the decline in human life and loss of ecosystem services is forest fire. The heating brought on by the world's rising average temperature is one of the main causes of forest fires. The ability to predict forest fires is essential components of forest fire control. Predicting forest fires is primarily done in order to allocate resources effectively and to assist the fire management team's firefighters as much as possible. Data mining algorithms have the ability to handle non-linear, highly dimensional data and provide solutions to complex issues. This study involves some classical classification methods, such as Random Forest, Support Vector Machine algorithms. With the help of strategic land use decisions, management, and planning, forest managers can utilize this methodology to identify areas that are at risk of fire and take action to prevent future fires. Based on fire hazard maps in the forests of western Iran or in forests of other locations with similar conditions, the results allow forest managers to choose the optimum management, control, and monitoring strategies for fire occurrence. In recent years, Lands have been Changing and fire has been increasing. Although other factors like drought, wind, topography, plants, etc., have significant indirect effects on fire appearance and its expansion, a large majority of forest fires are caused by humans. These fires also violate the functions of natural systems and cause significant material damage to the environment. For a firefighting operation to be effective, quick detection is essential. The development of automatic solutions has received a lot of attention because conventional human surveillance is expensive and influenced by subjective factors.

2. RELATED WORK

1. Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models:Forest fires are viewed as a serious threat to the environment, infrastructure, and human lives. In order to create the best possible combination of characteristics connected to forest fires, they use genetic algorithms. They also use data mining techniques to create a map of the vulnerability of different types of forest fires. We take into account factors like height, slope angle, aspect, etc. that are associated with forest fires. The creation of the maps showing the susceptibility of forests to fire was done using two cutting-edge data mining techniques, RF and Support Vector Machine (SVM). He reduced the emphasis on variables like mean annual temperature and mean annual wind speed in this.

2.Spatial pattern analysis and prediction of forest fire using new data mining approach:This study's primary goal was to suggest novel data mining methods for analyzing and forecasting regional patterns of forest fire threat. Ten contributing elements (slope, aspect, elevation, land use, distance to road, etc.) were included in a Geographical Information System database created for the study area. To build forest fire models and generate spatial trends in fires in forests, an innovative mixed machine learning model called MARS-DFP, which was Multivariate Adaptive Regression Splines (MARS) optimized by Differential Flower Pollination (DFP), was presented. They got the least accuracy by simply taking into account tropical forests.

3. Predicting the minimum height of forest fire smoke within the atmosphere using data mining and data from the CALIPSO satellite: In this work, they used information regarding nearby fire activity, geographic location, and climatic conditions to forecast the lowest height of the smoke layer as detected by CALIPSO. These data points are frequently accessible in close to real-time, assuring the operational applicability of the resulting model. The predictions can be used to directly apply to smoke detections from the current remote sensing products, integrate into statistical frameworks with inputs from remote sensing products, and be used to influence estimates of injecting height and downwind dispersion in deterministic models. The evaluation of ground-level population exposure to forest fire smoke for epidemiologic research and public health surveillance is anticipated to be improved by these possible uses. Using data on fire activity, they used Random Forest to forecast the smallest height of the smoke layer. For a firefighting operation to be effective, quick detection is essential. The development of automatic solutions has received a lot of attention because conventional human surveillance is expensive and influenced by subjective factors. Satellites have high acquisition costs, slow localization times, and insufficient resolution for all procedures. It requires a large database set and that should be effective. This is an Expensive process.

4. A data mining approach to predict forest fires using meteorological data: Forest fires are a serious environmental problem that endangers human lives and damages the environment and economy. A crucial component of controlling such a problem is quick detection. Utilizing automated tools based on nearby sensors, such as those given by meteorological stations, is one way to accomplish this. A number of forest fires, like the forest Fire Weather Index (FWI), employ information on the meteorological conditions (such as temperature and wind), which are known to affect fire indices. putting human lives in danger while causing economic and ecological damage. A crucial component of mitigating such situations is quick detection. Finding sufficient data that isn't obviously proprietary can also be challenging.

5. Predicting Forest fire risk based on mining rules with ant-miner algorithm in cloud-rich areas: Every year, flames burn off millions of hectares of forest area around the world. More research on the risk prediction of forest fires must be done in order to reduce the losses brought on by fire. It is anticipated that the collaborative use of operational forecasting systems and remote sensing-based models will consistently perform well at predicting the risk of forest fires. In this work, a brand-new model based on an ant-miner algorithm was put out. This method does an excellent job of tackling multivariable and nonlinear issues in the synergetic modeling of data from several sources. The suggested approach performs better at estimating danger levels. This study has the flaw that creating algorithms takes time and that it is challenging to create algorithms for large tasks. Algorithm branching and looping are challenging to demonstrate. For this algorithm, understanding through this convoluted logic might be very challenging.

6. Predictive modeling of wildfires: A new dataset and data mining approaches: The simulation's dataset will only contain a small number of cases because the experiment is intended to serve as a case study for demonstrating what can be done at greater sizes. We used two of the most well-known data mining algorithms in our investigation. In Databricks, these methods were put into practice. The approach taken in this study combines Big Data, Remote Sensing, and Data Mining methods (Artificial Neural Network and SVM) to analyze data gathered from satellite photos across vast areas and extract knowledge from them to anticipate the occurrence of wildfires and prevent such calamities. One of the most damaging natural disasters in the world is a wildfire. We outline a strategy for estimating the likelihood of wildfires. Based on data from remote sensing, we construct a dataset. One of the pricy ways used in this work to analyze repetitive photos is a limitation. Consequently, it will be challenging to carry out this project.

3. METHODOLOGY USED

The proposed system tries to implement more factors than the previously used inputs like Relative humidity, wind speed, Moisture, Drought, Despite the fact that wildfire additionally has a significant negative impact on the environment, society, and economy. Natural (climate, vegetation, topography) and anthropogenic (human activities) processes both have an impact on fire potential. The research area's strongest fire location predictors were the Drought Moisture Code and Drought Code. With the help of strategic land use decision-making, management, and organizing, forest managers may utilize this methodology to identify the places that provide the greatest fire risk and take action to reduce that risk. The proposed system has to be put into practice using various algorithms, including Support Vector Machine, Random Forest, and Decision Tree. We will finalize the forecast of fire through contrasting both algorithms and utilizing a voting procedure to determine which prediction is made the most frequently.

Advantages:

- Given that the suggested method will take more elements into account than the current system does, it will be simple to forecast the fire.
- Accurate and efficient detection of fire through mathematical computations.

- By involving data mining, each aspect of data is inspected clearly to give the best output through implementing data processing.
- It will be helpful to balance the environmental resources and ecosystem.

4. OBJECTIVES OF THE PROJECT

The project's primary goal is to offer a system that can analyze the occurrences of a Fire Accident using Machine Learning Algorithms. The System can prevent many environmental problems. So that it will be helpful to the forest managers. This project provides a system that can analyze and predict the fire accurately by using machine learning algorithms based on the voting process.

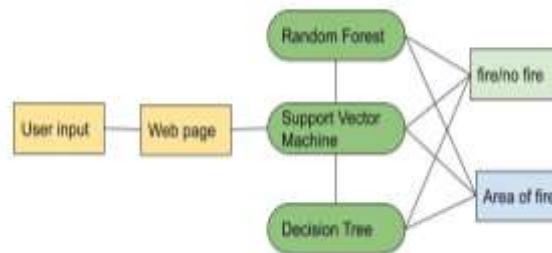


Fig. 1. Prediction of forest fire implementation flow

The forecast for forest fires is based on taking into the parameters that impact the fire occurrence. The project consists of a webpage with an interacting interface. The user inputs are based on the real-time environment, and the website's data collection is combined with algorithms like random forest, support vector machine, and decision tree. These algorithms are responsible to predict the forest fire and display the prediction and also calculate the area of forest affected by the fire.

Tools and Technologies:

Streamlit: Streamlit is used to make it easy to create interactive web applications which can be used to explore the data, visualize results, and share your work with others. Some of the features that Streamlit include are, Streamlit provides a simple API that allows you to create custom web applications with just a few lines of code. Streamlit automatically updates the web application in real-time as you modify the code or data. It offers a variety of interactive widgets that may be used to study the data and change the parameters, like sliders and dropdown menus. Streamlit includes built-in visualizations that make it easy to create charts and graphs from your data. Streamlit allows you to create custom components that can be used to add new functionality to your web application. Overall, the Streamlit is a very powerful tool for creating the custom web applications for machine learning.

Google colab: You can write and run Python code using a browser that works with Google Colab, which is a free cloud-based Jupyter Notebook development environment. It is an effective instrument for machine learning, data analysis, and research, providing access to a range of Google Cloud services, and the Google Cloud Storage. Colab comes with many popular Python libraries pre-installed, such as NumPy, Pandas, and Matplotlib.

Python: Python facilitates the creation of applications using an Object-Oriented programming methodology. Providing and learning a wide variety of high-level data structures is simple and easy. Python is a programming language that is appealing for application development since it is simple to learn yet also flexible and strong. This is a very useful capability that allows you to type in a python program and to have it executed immediately in an interactive mode. To implement the project, we have used the following libraries in python.

Sklearn: Machine learning is used by the Python library Scikit-learn. Classification, regression, clustering, and reduction of dimensionality are just a few of the features it offers for supervised and unsupervised learning. Python's Scikit Learn (Sklearn) package is a reliable and practical tool for machine learning. Through a consistent Python interface, it offers effective tools for statistical modeling and machine learning, such as dimensionality reduction, classification, and regression. This library was created using the programming languages NumPy, SciPy, and Matplotlib.

NumPy: NumPy is a python library which is used for computing numerical values. Which provides the powerful tools for arrays and matrices. It provides a number of functions for performing the mathematical operations on arrays and for random number generation. The major key features of NumPy are to perform the element wise operations on arrays and efficient calculations. It provides functions for manipulating the arrays. The NumPy is widely used in many scientific and engineering fields.

Pandas: Pandas is a Python package that offers tools and data structures for working with time series and other types of structured data. It adds additional capability for data processing, cleaning, and analysis on top of NumPy. The Series and Data Frame objects are the two main data structures in Pandas. A Series is a one-dimensional array-like object that can hold any data type, while a Data Frame is a two-dimensional table-like object that can hold multiple data types. In this example, we import Pandas using the import statement and give it the alias pd. We then create a Data Frame df from a dictionary data, where the keys of the dictionary correspond to the column names and the values correspond to the data for each column. Overall, Pandas is a powerful library for data analysis and manipulation in Python.

Data Collection

A dataset is a collection of data that is organized in a certain format for the use and analysis. Many disciplines, including machine learning, employ datasets. To train the machine learning models, it is employed. Datasets can be in many different formats, such as text, graphic, and tabular data. Here, we've used tabular data, which includes data from SQL databases, CSV files, and Excel files. With the use of machine learning methods like Decision Tree, Random Forest, and Support Vector Machine, we have taken this dataset from Kaggle and will use data preparation techniques to extract the necessary elements from the forest in order to predict the forest fire. It is crucial to correctly preprocess and clean the data before using datasets too.

A	B	C	D	E	F	G	H	I	J	K	L	M
x	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
7	3	mar	fri	86.2	26.2	94.3	5.1	8.2	33	6.7	0	0
7	4	oct	tue	90.6	33.6	669.1	6.7	18	11	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	37	4	0.2	0
8	8	mar	sun	89.3	51.3	102.2	9.8	11.4	39	1.8	0	0
8	8	aug	sun	92.3	85.3	488	14.7	22.2	29	5.4	0	0
8	8	aug	mon	92.3	88.9	495.6	8.5	24.1	27	1.1	0	0
8	8	aug	mon	91.5	145.4	608.2	10.7	8	88	2.2	0	0
8	8	sep	tue	91	129.5	692.6	7	13.1	63	5.4	0	0
7	3	sep	sat	92.5	88	696.6	7.1	22.8	40	4	0	0
7	3	sep	sat	92.5	88	696.6	7.1	17.8	51	7.2	0	0
7	3	sep	sat	92.8	71.2	711	22.8	19.3	18	4	0	0
6	3	aug	fri	63.5	70.8	605.3	0.8	17	72	6.7	0	0
6	3	sep	mon	90.9	126.5	686.5	7	21.3	42	2.2	0	0
6	3	sep	wed	92.9	111.3	699.6	9.2	26.4	21	4.5	0	0
6	3	sep	fri	93.3	141.2	713.9	19.9	22.9	44	5.4	0	0
5	3	mar	sat	91.7	35.8	80.8	7.8	15.1	27	5.4	0	0
8	3	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
8	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	280	16.1	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	10.3	23.7	32	5.8	0	0
7	4	aug	sun	93.4	142.4	601.4	10.6	18.3	40	5.4	0	0

Fig. 2.Forest fire Dataset

Data Pre-Processing: Data Preprocessing refers to the process of preparing raw data for analysis by cleaning, transforming, and organizing it into a format that is suitable for machine learning or statistical modeling. Data pre-processing primary objective is to increase the data's quality and accuracy by removing noise, inconsistencies, and errors that could skew or produce inaccurate conclusions.

Dataset Splitting: In statistical modeling and machine learning, the division of data into training and testing sets is a standard strategy. This method's goal is to assess how well the model performs when applied to fresh, untested data. There are two sets in the dataset, a testing set and a training set are used to train the model and assess its efficacy, respectively. Divide the dataset into a training set and a testing set at random. To fit the model and improve its parameters, utilize the training set. Utilize the testing set to gauge how well the model performs with fresh, untested data.

Algorithm: Algorithm implementation on training dataset refers to the process of using a machine learning algorithm to train a model on the training dataset. This involves feeding the algorithm with the training data to learn the underlying patterns and relationships between the input variables and the target variable. The proposed system uses machine learning algorithms Decision Tree, Random Forest, Support Vector Machine are used. The suggested system yields effective outcomes. Algorithms for machine learning are processes that are written in code and applied to data. The quality and size of the training dataset, the algorithm's complexity, and the data analyst or data scientist's level of experience all affect how well an algorithm performs when applied to a training dataset. To guarantee that the model is reliable and successful in making predictions or spotting patterns in the data, it is crucial to properly choose and preprocess the data as well as an appropriate algorithm.

Performance evaluation: Algorithms are evaluated based upon their performance and their accuracy.

Making Predictions: Making predictions is the process of utilizing a machine learning model that has been trained to anticipate the outcome of fresh, unforeseen data using the patterns and connections discovered during training.

5. RESULTS AND DISCUSSION

Utilizing a machine learning model that has been trained to predict the result of novel, unanticipated data using the patterns and connections found during training is the process of making predictions.

TABLE I. VARIOUS TECHNIQUES FOREST FIRE

S. No	Techni-que	Advantages	Disadvantages
1	Optimal combination of forest fire related variables is set by the genetic algorithm application....data mining models are used for forest fire susceptibility	They use data mining techniques to design a forest fire and Genetic Algorithms to find the best combination of characteristics connected to forest fires.	He limited the attention on the factors like mean yearly temperatures and mean yearly wind speed.
2	Analyzing spatial patterns and forecasting forest fires with new machine learning techniques.	In this, they used a case study of a tropical forest fire to analyze and forecast geographical patterns of forest fire danger.	They considered only Tropical Forest and gained least accuracy.
3	employing data mining from the Calypso satellite to predict the minimum height of forest fire smoke in the atmosphere.	Using data on the fire activity, they used Random Forest to forecast the minimum height of the smoke layer.	It requires a large database set and that should be effective. This is an expensive process.
4	Data mining method to forecast forest fires using meteorological information.	putting human lives in risk while causing economic and ecological damage. A crucial component of mitigating such situations is quick detection.	It is a complicated process. It can also be difficult to find adequate data that isn't already proprietary in nature.
5	using an ant-miner algorithm with mining rules to predict the likelihood of forest fires in places with a lot of clouds.	In this work, a brand-new model based on an ant-miner algorithm was put out. This method does an excellent job of tackling multivariable and nonlinear issues in the synergetic modeling of data from several sources.	The drawback of this study is Algorithms are time consuming and understanding through algorithms can be very difficult.
6	modeling that anticipates wildfires, a fresh dataset and method for data mining.	The approach employed in this research integrates methods from Big Data, Remote Sensing, and Data Mining to analyze information gathered from satellite photos taken across vast distances.	When measuring or analyzing smaller areas, remote sensing is a pretty expensive way of analysis.

6. CONCLUSION

Modeling forest fires is a difficult task because its spread is frequently influenced by a number of variables, including the climate, topography, people, and flora. As a result, in terms of success rate, the Decision Tree Algorithm ensemble obtained the highest Accuracy value, followed by Support Vector Machine and Random Forest. Algorithms are used to estimate the extent of the forest damaged by the forest fire, which improves analysis of the predicted fire and damaged area of the eco forest. These findings suggest that the maps produced here could be a very useful management tool for planning and analyzing forest fire control. The methodology can also be applied to other regions that have a related problem.

7. FUTURE SCOPE

The changes in the algorithms are simple to implement since we use a modular approach, and work can be extended later on to adapt the model's implementation. We insert the model's pickle file in the necessary places so that it may be quickly copied onto products. Therefore, this may easily avoid compiling the big code in its entirety each time. We may also put it into practice by forecasting utilizing image processing. As we did with Python, we can also use the forthcoming, simple, comprehensible programming language Julia for implementing our project.

8. REFERENCES

- [1] Hong, Haoyuan, et al. "Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China." *Science of the total environment* 630(2018):1044-1056.<https://doi.org/10.1016/j.scitotenv.2018.02.278>
- [2] ien Bui, D., Hoang, N.-D., Samui, P., 2019. Spatial pattern analysis and prediction of forest fire using new data mining approach of Multivariate Adaptive Regression Splines and Differential Flower Pollination optimization: a case study at Lao Cai province (Viet Nam). *J. Environ. Manage.* 237, 476–487. <https://doi.org/10.1016/j.jenvman.2019.01.108>.
- [3] Yao, Jiayun, et al. "Predicting the minimum height of forest fire smoke within the atmosphere using data mining and data from the CALIPSO satellite." *Remote sensing of environment* 206 (2018): 98-106..
- [4] Cortez, P., & Morais, A. D. J. R. (2007). A data mining approach to predict forest fires using meteorological data.
- [5] Zheng, Zhong, et al. "Predicting Forest fire risk based on mining rules with an ant-miner algorithm in cloud-rich areas." *Ecological Indicators* 118 (2020): 106772.
- [6] Sayad, Younes Oulad, Hajar Mousannif, and Hassan El Moatassime. "Predictive modeling of wildfires: A new dataset and data mining approach." *Fire safety journal* 104 (2019): 130-146.
- [7] Tang, Xianzhe, et al. "A novel optimized repeatedly random under sampling for selecting negative samples: A case study in an SVM-based forest fire susceptibility assessment." *Journal of Environmental Management* 271 (2020): 111014.7
- [8] Pourtaghi, Zohre Sadat, et al. "Investigation of general indicators influencing forest fire and its susceptibility modeling using different data mining techniques." *Ecological indicators* 64 (2016): 72-84.
- [9] Eskandari, Saeedeh, Hamid Reza Pourghasemi, and John P. Tiefenbacher. "Relations of land cover, topography, and climate to fire occurrence in natural regions of Iran: Applying new data mining techniques for modeling and mapping fire danger." *Forest Ecology and Management* 473 (2020): 118338.<https://doi.org/10.1016/j.foreco.2020.118338>
- [10] Ajbilou, R., Marañón, T., Arroyo, J., 2006. Ecological and biogeographical analyses of Mediterranean forests of northern Morocco. *Acta Oecologica* 29, 104–113.