PROJECT REPORT

on

Wildfire Prediction and Detection

(CSE VI Semester Mini project PCS-604)

2021-2022

Submitted to:

Dr.Satvik Vats

(CC-CSE-B-VI-Sem)

Guided by:

Dr.Satvik Vats

(Resource Person)

Submitted by:

Mr. Shreeyash Pandey

Roll. No.: 1918709

CSE-B-VI-Sem

Session: 2021-2022

DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

GRAPHIC ERA HILL UNVERSITY, DEHRADUN

CERTIFICATE

Certified that Mr. Shreeyash Pandey (Roll No.-1918709) has developed mini project on “Wildfire Prediction and Detection” for the CS VI Semester Mini Project Lab (PCS-604) in Graphic Era Hill University, Dehradun. The project carried out by Students is their own work as best of my knowledge.

Date:

(Dr.Satvik Vats)	Dr.Satvik Vats
Class Co-ordinator	**Project Guide**
CSE-B-VI-Sem	**Resource Person**
(CSE Department)	(CSE Department)
Gehu Dehardun	Gehu Dehradun

ACKNOWLEDGMENT

We would like to express our gratitude to The Almighty Shiva Baba, the most Beneficent and the most Merciful, for completion of project.

We wish to thank our parents for their continuing support and encouragement. We also wish to thank them for providing us with the opportunity to reach this far in our studies.

We would like to thank particularly our project Co-ordinator Dr. Satvik Vats and our Project Guide Dr. Satvik Vats for his patience, support and encouragement throughout the completion of this project and having faith in us.

We also acknowledge to teachers like Dr.Satvik Vats who help us in developing the project.

At last but not the least We greatly indebted to all other persons who directly or indirectly helped us during this work.

Mr. Shreeyash Pandey

Roll No.- 1918709

CSE-B-VI-Sem

Session: 2021-2022

GEHU, Dehradun

TABLE OF CONTENTS

ABSTRACT

Wildfires are the most deadly and dangerous accidents across the World, especially in Forest Regions. Many lives are lost, and billions of dollars worth of property damages occur in wildfire every year. Wildfires are fueled and accelerated by several different factors, such as weather, climate, vegetation types, land cover, and human activities. This projects aims to develop a machine learning fire risk prediction model considering the different geographical factors outlined above. We propose a systematic way to make fire risk prediction and detection models that analyzes satellite data, weather data, and historical fire data to predict fire. We tried to analyze the Uttra-khand Region situated in India with an Latitude : 30° 15' N and Longitude : 79° 15' E.

Literature Review

Many researchers have recently started using machine learning models for wildfire risk prediction based on parameters such as a month, day, temperature, relative humidity, wind, and rain . Guruh F. S. and Khabib M. in proposed a hybrid Model using clustering and classification approaches to improve the wildfire risk prediction accuracy by considering eight parameters: temperature, relative humidity, wind, and rain. They trained a Back-Propagation Neural Network (BPNN) to get the output in three categories: no burn, light burn, and heavy burn. BPNN is one of the methods of neural network in which weights of a neural network are finetuned based on the error rate got in the last iteration. Marcos et al. introduce a wildfire risk prediction system based on three different machine learning models to study the effect of humancaused factors in wildfires in Spain . They implemented Random Forest (RF), Boosting Regression Trees (BRT), and Support Vector Machines (SVM) to detect fire risk prediction and found improvement in the accuracy in terms of the area under the curve (AUC). AUC determines the area covered by the Receiver operating characteristic (ROC) curve. For the perfect classifier, the AUC score is 1.0, while for the random classifier AUC value is 0.5. Researchers Caroline Famiglietti et al. predict fire risk in Northern California using different machine learning models, Logistic Regression (LR), Decision Trees (DT), and Multilayer Perceptron (MLP), based on remote sensing data. LR, also known as logit regression, uses a logistic function to model the probability of a specific event, DT is the non-parametric approach used for regression and classification. MLP is a type of neural network that connects multiple layers of the perceptron. Wonjae et al. developed a wildfire detection system based on deep convolutional neural networks (DNN) using crewless aerial vehicles . They achieved high accuracy than conventional machine learning algorithms. DNN is the part of neural networks that have at least three or four input and output layers. Mahsa Salehi et al. developed a Context-Based Fire Risk (CBFR) model for fire risk detection using ensemble learning techniques with high accuracy . They used weather data to determine the temporal variation of the wildfire danger prediction in Blue Mountains, Australia. Recently Malik et al. proposed two machine learning approaches based on random forest (RF) models to predict the wildfire risk in Monticello and Winters, California . They used fire history, weather, vegetation, powerline, and weather data in their study and obtained an accuracy of 92%. RF is an ensemble learning method that consists of multiple decision trees and used for classification and regression tasks.

Introduction

A wildfire is an uncontrolled fire that burns in the wildland vegetation, often in rural areas. Wildfires can burn in forests, grasslands, savannas, and other ecosystems, and have been doing so for hundreds of millions of years. They are not limited to a particular continent or environment.

Wildfires can burn in vegetation located both in and above the soil. Ground fires typically ignite in soil thick with organic matter that can feed the flames, like plant roots. Ground fires can smolder for a long time—even an entire season—until conditions are right for them to grow to a surface or crown fire. Surface fires, on the other hand, burn in dead or dry vegetation that is lying or growing just above the ground. Parched grass or fallen leaves often fuel surface fires. Crown fires burn in the leaves and canopies of trees and shrubs.

A myriad of research papers has been published addressing wildfire detection and prediction by using mathematical and statistical methods. Still, these models have a lot of limitations such as limited parameters, low accuracy of risk prediction, the complexity of equations, and lack of real-time decision-making processes. According to the recent wildfire survey, most of the wildfire emergency systems still use conventional wildfire detection and prediction approaches. In statistical and mathematical methods, we infer the relationship among the variables, while in the machine learning (ML) models, the focus is to make the most accurate predictions possible. Hence with the advancement in Machine Learning and Neural Networks, we can leverage the advanced algorithms to improve the lagging outcomes of wildfire risk prediction and detection systems. Many types of research are focused on investigating the probability of the burning, while others focus on the intensity and effects of the wildfires . Moreover, earlier studies also show the deployment of a limited number of paraments with limited accuracy. Therefore, we aim to include different parameters such as fire history, weather, remote sensing, and satellite data to improve the accuracy of fire risk prediction and detection models.

2.PROJECT

2.1 Tools and Technologies

2.1.1Python

Python is a high-level, general-purpose and a very popular programming language. Python programming language (latest Python 3) is being used in web development, Machine Learning applications, along with all cutting edge technology in Software Industry. Python Programming Language is very well suited for Beginners, also for experienced programmers with other programming languages like C++ and Java.

Below are some facts about Python Programming Language:

1. Python is currently the most widely used multi-purpose, high-level programming language.
2. Python allows programming in Object-Oriented and Procedural paradigms.
3. Python programs generally are smaller than other programming languages like Java. Programmers have to type relatively less and indentation requirement of the language, makes them readable all the time.
4. Python language is being used by almost all tech-giant companies like – Google, Amazon, Facebook, Instagram, Dropbox, Uber… etc.
5. The biggest strength of Python is huge collection of standard library which can be used for the following:

- Machine learning
- GUI Applications (like Kivy, Tkinter, PyQt etc.)
- Web frameworks like Django (used by YouTube, Instagram, Dropbox)
- Image processing (like OpenCV, Pillow)
- Web scraping (like Scrapy, BeautifulSoup, Selenium)
- Test frameworks
- Multimedia
- Scientific computing

2.1.2 Machine Learning

Machine learning (**ML**) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so.Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

A subset of machine learning is closely related to computational statistics, which focuses on making predictions using computers, but not all machine learning is statistical learning. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through unsupervised learning. Some implementations of machine learning use data and neural networks in a way that mimics the working of a biological brain. In its application across business problems, machine learning is also referred to as predictive analytics.

2.1.3 Deep Learning

Deep Learning is a machine learning technique that constructs artificial neural networks to mimic the structure and function of the human brain. In practice, deep learning, also known as deep structured learning or hierarchical learning, uses a large number hidden layers -typically more than 6 but often much higher - of nonlinear processing to extract features from data and transform the data into different levels of abstraction (representations).

As an example, assume the input data is a matrix of pixels. The first layer typically abstracts the pixels and recognizes the edges of features in the image. The next layer might build simple features from the edges such as leaves and branches. The next layer could then recognize a tree and so on. The data passing from one layer to the next is considered a transformation, turning the output of one layer into the input for the next. Each layer corresponds with a different level of abstraction and the machine can learn which features of the data to place in which layer/level on its own. Deep learning is differentiated from traditional “shallow learning” because it learns much deeper levels of hierarchical abstraction and representations.

2.2 Case Study

2.2.1 Data Collection

As discussed above,we have kept our focus on an perticular region(Uttrakhand in India). We collected an MODIS Dataset for South Asia region.This Data Set consists of parameters which influence the wildfire in a region.

1	latitude	longitude	brightness	scan	track	acq_date	acq_time	satellite	confidence	version	bright_t31	frp	daynight
2	7.37478	81.37296	318.39	1.01	1.01	11-05-2022	501	T	70	6.1NRT	292.44	11.15	D
3	28.47017	80.96015	324.15	1.46	1.19	11-05-2022	455	T	19	6.1NRT	299.9	12.9	D
4	27.89231	81.54675	326.51	1.33	1.14	11-05-2022	455	T	72	6.1NRT	300.69	10.63	D
5	26.63741	82.95482	331.01	1.1	1.05	11-05-2022	457	T	72	6.1NRT	305.98	11.3	D
6	26.64724	80.03714	325.69	1.58	1.24	11-05-2022	457	T	58	6.1NRT	302.43	14.25	D
7	25.32475	88.43656	314.04	1.32	1.14	11-05-2022	457	T	33	6.1NRT	295.92	11.37	D
8	25.53183	84.8814	335.01	1	1	11-05-2022	457	T	77	6.1NRT	307.02	17.06	D
9	25.51976	84.83902	331.55	1	1	11-05-2022	457	T	71	6.1NRT	306.25	11.7	D
10	25.01264	85.69677	333.04	1.02	1.01	11-05-2022	457	T	66	6.1NRT	308.86	15.28	D
11	25.08903	84.37562	333.09	1	1	11-05-2022	457	T	62	6.1NRT	307.36	9.72	D
12	24.81248	81.52876	334.23	1.22	1.1	11-05-2022	457	T	69	6.1NRT	307.81	15.1	D
13	24.53427	80.28177	333.47	1.43	1.18	11-05-2022	457	T	68	6.1NRT	308.56	17.23	D
14	24.1651	82.68256	331.28	1.07	1.03	11-05-2022	457	T	54	6.1NRT	308.06	9.01	D
15	24.17755	80.74178	336.73	1.33	1.14	11-05-2022	457	T	76	6.1NRT	307.82	26.41	D
16	24.17596	80.75475	330.97	1.33	1.14	11-05-2022	457	T	60	6.1NRT	307.88	14.24	D
17	23.95268	79.55893	335.01	1.57	1.23	11-05-2022	457	T	70	6.1NRT	311.02	22.01	D
18	23.38871	77.71944	331.06	2.08	1.4	11-05-2022	457	T	54	6.1NRT	310.09	22.74	D
19	22.4134	80.20012	335.94	1.36	1.16	11-05-2022	457	T	75	6.1NRT	314.81	16.74	D
20	22.41179	80.21321	334.19	1.36	1.15	11-05-2022	457	T	68	6.1NRT	313.29	12.61	D
21	20.82056	79.21906	334.15	1.5	1.21	11-05-2022	457	T	69	6.1NRT	304.57	25.73	D
22	14.7558	101.94987	317.12	1.27	1.12	11-05-2022	629	A	65	6.1NRT	295.29	7.8	D
23	17.37389	96.47832	313.26	2.82	1.6	11-05-2022	629	A	0	6.1NRT	289.02	25.39	D
24	39.7626	67.28103	309.9	1.18	1.08	11-05-2022	631	T	29	6.1NRT	295.01	5.83	D

MODIS_C6_1_South_Asia_7d

2.2.2 Machine Learning Approach

Our approach consists of using the data collected through Moderate Resolution Imaging Spectroradiometer (MODIS) data which is available through LANCE generally within 60 to 125 minutes after a satellite observation .

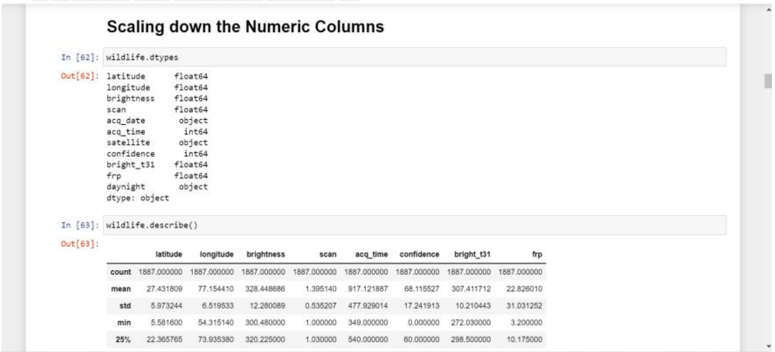
Our model uses the following parameters:

- **Descriptors:** Latitude,Longitude,Brightness,FRP,satellite,scan_binned,track,Day-night;
- **Target:** Confidence (Ranges from 0 to 100)

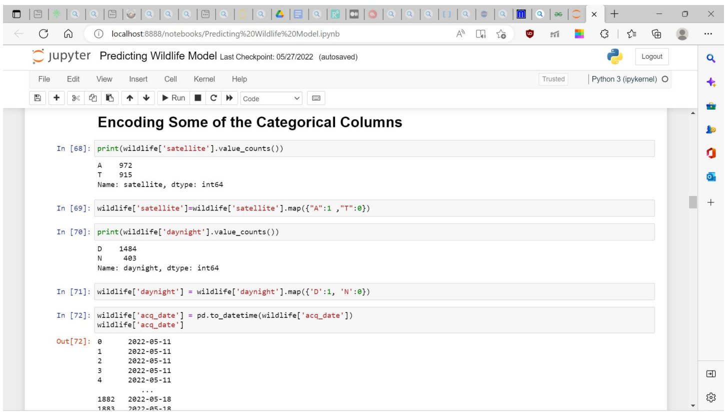
The idea is to analyze the time series defined by these descriptors and use them to predict the time series of the target.

Data Pre-Processing:

Feature Scaling:Feature scaling in machine learning is one of the most important steps during the preprocessing of data before creating a machine learning model. This can make a difference between a weak machine learning model and a strong one.In this Dataset,we had scaled down some of the faetures for efficient modelling.

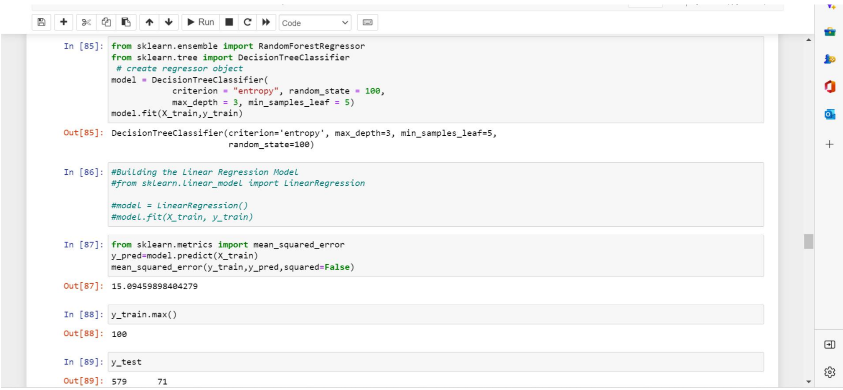


Encoding of Categorical columns:The categorial columns are encoded into numerical data for efficient modelling.



Modeling

Our problem consists in predicting the evolution of the of the fires using the data collected through the satellite .This Model results us with an confidence value which ranges from [0 to 100]. The confidence can be further classified as low confidence,High confidence.



We applied the following steps for machine learning. We:

- Pre-processed the data for efficient modeling ;
- Extracted features from Preprocessed Data;
- Apply a regression algorithm to these data.

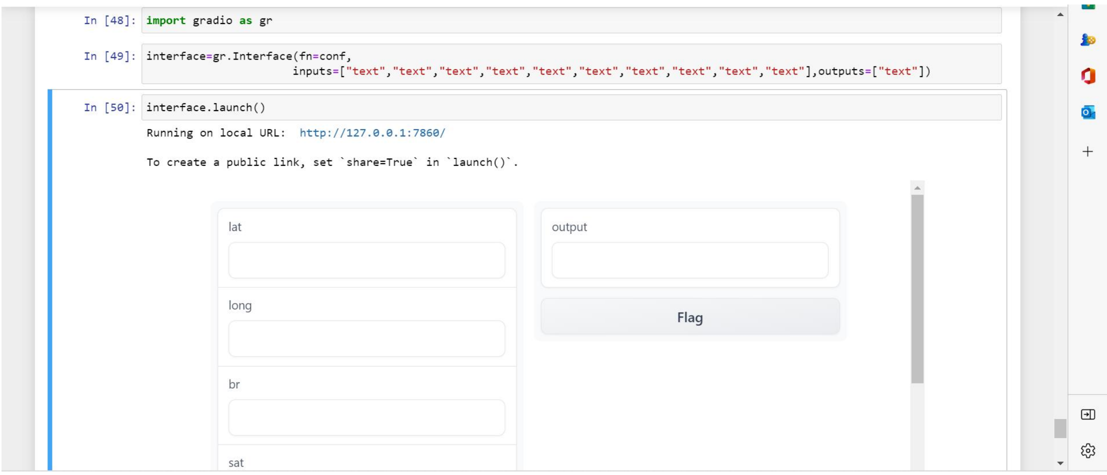
We have applied different regression algorithms in the model.Random forest regression model came as the most efficient model in training the Data-Set.

Results and Outcomes: We have used mean square error score as our evaluation metric to compare the performance of the algorithms. Each of the models was hyper-tuned and regularized to obtain the best evaluation metric score. From our observation we can conclude that Random Forest regressor gave the highest accuracy of all the other models.

Conclusion

In our research, we have explored the effect of various types of data to study fire risk prediction and detection using machine learning approaches. Unlike other research that examines either fire risk detection or fire risk prediction with limited data and parameters, our work focuses on understanding these concepts using past fire events, weather, remote sensing, and satellite data. We were able to develop a regression model for analyzing the wildfire of an perticular region. In the future, we plan to develop a real-time intelligent fire system that can provide information about the fire risk prediction, detection, and fire spread pattern.

SNAPSHOT OF THE PROJECT



REFERENCES

N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321-327, June 2002.

B. Ayhan, C. Kwan, B. Budavari, L. Kwan, Y. Lu, D. Perez, J. Li, D. Skarlatos, M. Vlachos, "Vegetation Detection Using Deep Learning and Conventional Methods," Remote Sens., vol. 12, pp. 2502, August 2020.

P. Jain, S. C. P. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, M. D. Flannigan, "A review of machine learning applications in wildfire science and management," Environ. Rev., vol. 28, pp. 73, December 2020.

G. Tiziano, A. Bachisio, P. Grazia, and D. Pierpaolo, "An Improved Cellular Automata for Wildfire Spread," Procedia Comput. Sci., vol. 51, pp. 2287-2296, June 2015.

G. L. Hung, M. S. B. Sahimi, H. Samma, T. A. Almohamad, B. Lahasan, "Faster R-CNN Deep Learning Model for Pedestrian Detection from Drone Images," SN Comp. Sci., vol. 1(116), April 2020.