

PREDICTION OF FLOOD BY RAINFALL USING MACHINE LEARNING

Dr.P.M.Joe Prathap¹, P.Devanesan²,G.Dinesh³,V.Manoj Prabha⁴

¹Professor, Department Of Information Technology ,R.M.D Engineering College

^{2,3,4}Student, Department Of Information Technology ,R.M.D Engineering College

Abstract - Floods are one of the most common natural disasters, which is highly complex to model. The research on flood prediction using ML algorithms leads to risk reduction, policy suggestion, minimization of the loss of human life, and reduction of property damage associated with floods. To mimic the complex mathematical expressions of physical processes of flood, Machine Learning provides better performance and cost-effective solutions. The analysis of a dataset by Machine Learning Algorithms to capture information like variable identification, missing value treatments, data validation and data cleaning/preparing will be done on the entire given dataset. To perform prediction of flood happen or not by accuracy calculation with evaluation classification report, identify the confusion matrix and the result shows the effectiveness of the GUI based application by given attributes.

its performance. Data scientists use many various sorts of machine learning algorithms to get patterns in python that cause actionable insights. At a high level, these different algorithms are often classified into two groups supported by the way they “learn” about data to form predictions: supervised and unsupervised learning. Classification is the process of predicting the category of given data points. Classes are sometimes called targets/ labels or categories. Classification predictive modeling is the task of approximating a mapping function from input variables(X) to discrete output variables(y). In machine learning and statistics, the classification

Key Words: Dataset, Python, Preprocessing, ML algorithms

1.INTRODUCTION

Machine learning is to predict the longer term from past data. Machine learning (ML) could be a sort of AI (AI) that gives computers the power to find out without being explicitly programmed. Machine learning focuses on the event of Computer Programs which will change when exposed to new data and therefore the basics of Machine Learning, implementation of an easy machine learning algorithm using python. The Process of coaching and prediction involves the use of specialized algorithms. It feeds the training data to an algorithm, and the algorithm uses this training data to give predictions on new test data. Machine learning can be roughly separated into three categories. There is supervised learning, unsupervised learning, and reinforcement learning. A Supervised learning program is both given the input file and therefore the corresponding labeling to find out data has got to be labeled by a person's being beforehand. Unsupervised learning has no labels. It provided the learning algorithm. This algorithm has got to find out the clustering of the input file. Finally, Reinforcement learning dynamically interacts with its environment and it receives positive or feedback to enhance

2. RELATED WORK

The classification is done on an image dataset that consists of two subgroups one is aerial images with flood-affected areas and the other is aerial images without any flooded affected areas. The downloaded dataset is randomly partitioned then the classification of the image is done. This classification involves the hybridization of the SVM classifier along with k-means clustering. The SVM classifier is used to train various categories of the dataset. Comparative studies show that the SVM classifier is one of the preferable classifiers among all the others that are widely used in many remote sensing applications and provides much better accuracy when compared to other algorithms like decision trees. It was then used for solving multi-class classification problems. It builds a hyperplane that separates both the categories and is used for categorizing new images. Classifiers are evaluated based on various parameters; some of them being accuracy, prediction speed, and training time. The confusion matrix is built to represent the performance of the classifier. This matrix represents the number of correctly and incorrectly classified examples from the dataset. It is capable of classifying the flooded areas with good accuracy. This model classified 90% of flooded images correctly. A comparison is done with the accuracy of the model by training it with different environments and the results have been plotted. Kernel functions are also changed for the SVM classifier to plot the difference in the prediction and training time.

3. LITERATURE SURVEY

Title: A Hybrid Machine Learning Approach for Classifying Aerial Images of Flood-Hit Areas

Author: Akshya .J, P.L.K.Priyadarsini

Year: 2019

The classification is done on an image dataset which consists of two subgroups one is aerial images with flood-affected areas and other is aerial images without any flooded affected areas. The downloaded dataset is randomly partitioned then the classification of image is done. This classification involves hybridization of SVM classifier along with k-means clustering. The SVM classifier is used to train various categories of the dataset. Comparative studies show that SVM classifier is one of the preferable classifiers among all the others that are widely used in remote sensing applications and provides a much better accuracy when compared to other algorithms like decision trees. It was then used for solving multi-class classification problems. It builds a hyperplane that separates both the categories and is used for categorizing new images. Classifiers are evaluated based on various parameters; some of them being accuracy, prediction speed and training time. The confusion matrix is built to represent the performance of the classifier. This matrix represents the number of correctly and incorrectly classified examples from the dataset. The performance of this methodology is evaluated. It is capable of classifying the flooded areas with good accuracy. This model classified 90% of flooded images correctly. A comparison is done with the accuracy of the model by training it with different environments and the results have been plotted. Kernel functions are also changed for the SVM classifier to plot the difference in the prediction and training time.

Title: Flood Prediction Using Multi-Layer Artificial Neural Network in Monitoring System with rain gage, Water Level, Soil Moisture Sensors

Author: Febus Reidj G. Cruz, Matthew G. Binag, Marlou Ryan G, Francis Aldrine A.

Year: 2018

This study was about to implement a real-time monitoring system capable of measuring parameters such as rainfall intensity, soil moisture, water level and rate of increase in water level. The different sensor was integrated into a system, wherein the data was logged and stored. A prediction model based on a multi-layered artificial neural network was

developed and was tested using an actual setup. This study validated the response of the multi-layered network model. The flood prediction model had a small deviation with reference to the actual water level, with RMSD of 2.2648. In the Philippines that was a major issue as it leads to damage of properties, damage to infrastructures or even loss of lives. Current systems adhere to unravel issues to stop devastating disasters caused by floods. Multi-layered artificial neural network with the help of MATLAB was utilized in the development of the prediction model. In the training, test, validation and overall dataset, the network showed a really good goodness-of-fit specifically 0.99889 for the training dataset, 0.99362 for the test data set, 0.99764 for the validation dataset and 0.99795 considering all the info within the dataset. The network was programmed and integrated within the system within the actual setup. The model is validated by running trials with certain inputs and predicted flood level because the output is compared to the particular flood level after a particular period of your time. The flood prediction system showed an RMSD value of two .2648 which signifies a little overall difference between the anticipated flood level and actual flood level across the whole dataset tested in the actual setup.

Title: A Hybrid Machine Learning Approach for Classifying Aerial Images of Flood-Hit Areas

Author: Akshya .J, P.L.K.Priyadarsini

Year: 2019

The classification is done on an image dataset which consists of two subgroups one is aerial images with flood-affected areas and other is aerial images without any flood affected areas. The downloaded dataset is randomly partitioned then the classification of image is done. This classification involves hybridization of SVM classifier along with k-means clustering. The SVM classifier is used to train various categories of the dataset. Comparative studies show that SVM classifier is one of the preferable classifiers among all the others that are widely used in many remote sensing applications and provides a much better accuracy when compared to other algorithms like decision trees. It was then used for solving multi-class classification problems. It builds a hyperplane that separates both the categories and is used for categorizing new images. Classifiers are evaluated based on various parameters; some of them being accuracy, prediction speed and training time. The confusion matrix is built to represent the performance of the classifier. This matrix represents the number of correctly and incorrectly classified examples from the dataset. The performance of this methodology is evaluated. It is capable of classifying the flooded areas with good accuracy. This model classified 90% of flooded images correctly. A comparison is done with the accuracy of the model by training it with different environments and the results have been plotted. Kernel functions are also changed for the SVM classifier to plot the difference in the prediction and training time.

4. EXISTING SYSTEM

The prosaic strategies of flood forecasting are expensive and highly complex. Weather and rainfall forecasting may be a major task behind the prediction of a flood. Weather forecasting involves simulations supported by physics and differential equations. The rainfall forecast is completed using radars and satellite imaging. A Doppler radar is employed to locate the precipitation and detect the motion of rain droplets. Dedicated meteorological satellite provides images using which information about rainfall is often deduced and it's for short-term flash flood prediction in urban areas is to determine a theoretical model incorporating the factors influencing flood and use the facility of machine learning techniques to estimate flood before time. It is important to predict the occurrence of a flood so that its aftermath is often minimized. As the name suggests, an urban flash flood occurs during a populated area in a very short span of your time. To reduce the impact of those events, short-term forecasting or nowcasting is employed for the prediction of the very near future incident. In orthodox methods of flood forecasting, current weather is examined using conventional methods like the utilization of radar, satellite imaging, and calculations involving complicated mathematical equations. However, recent developments in Information and Communication Technology (ICT) and Machine Learning (ML) have helped us to review this hydrological problem from a special perspective. The aim is to style a theoretical model considering the parameters causing the urban flash flood and predict the event.

Drawbacks:

1. To implement a model for an urban area in which short-term forecasting of a flood and its complex scenario, including multiple cells, is imagined to undertake a possible real-world scenario.
2. It can't determine the regularity of rainfall data and achieve more accurate results.

5. PROPOSED SYSTEM

Floods are one of the biggest natural disasters causing many lives as well as damages. Different types of floods like river floods, urban floods, coastal floods, and flash floods have been observed over the years. A flash flood is a direct response to a rainfall having very high intensity in a small time. This kind of flood is seen typically in urban areas where the underlying ground cannot cope, or drain excess water away fast enough via the sewage system and drainage canals in a short amount of time. In recent years, we have seen the impact of floods in cities such as Mumbai, Chennai, Ahmedabad, and Kerala. Inadequate flood mitigation systems are the main reasons behind it. Given a dataset from different sources would be to form a generalized dataset, and applied to extract patterns and to obtain results with maximum accuracy. This section of the report will load in the data, check for cleanliness, and then trim and clean the given dataset for analysis. Make sure that the document steps carefully and is justified for cleaning decisions. The data set collected for predicting given data is split into a Training set and Test set. Generally, 7:3 ratios are applied for splitting. The Data Model

which was created using Machine learning algorithms will be applied to the Training set and based on the test result accuracy, the Test set prediction is done. Among the natural disasters, floods are the most destructive to human life, infrastructure, agriculture, and the socioeconomic system. Flood prediction models are of serious importance for hazard assessment. Robust and accurate prediction contributes highly to water resource management strategies, policy suggestions and analysis, and further evacuation modeling. Thus, the importance of advanced systems for both short-term and long-term prediction for flood and other hydrological events are strongly emphasized to alleviate the damage. However, the prediction of flood time interval and occurrence location is fundamentally complex thanks to the dynamic nature of climate conditions.

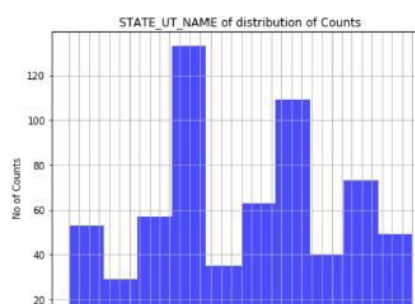
Advantages: These reports are to investigate the applicability of machine learning techniques for air quality forecasting in operational conditions. The results produced are of more accuracy

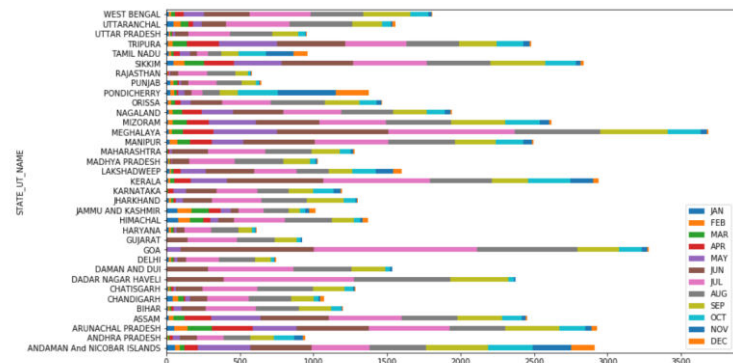
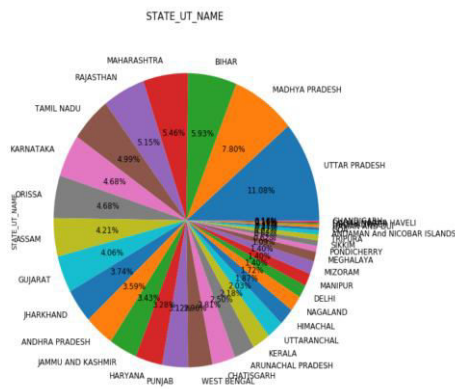
6. MODULE DESCRIPTION

- Data validation and pre-processing technique (Module-01)
- Create a predicted variable by rainfall range (Module-02)
- Performance measurements of ML algorithms (Module-03)
- GUI based prediction of flood by rainfall (Module-04)

Data validation and pre-processing technique:

Validation strategies in ML are wont to get the mistake fee of the Machine Learning model, If the information extent is massive sufficient to be a consultant of the populace, you could no longer want the validation strategies. However, in actual-global scenarios, to determine with samples of know-how with a purpose to now no longer be an actual consultant of the populace of a given dataset. To locate the lacking value, replica value, and outline of records kind whether or not it's a far waft variable or integer. The pattern of know-how wont to offer an impartial assessment of the dataset at the same time as tuning model hyperparameters. The assessment will become extra biased as a talent at the validation dataset is included into the version configuration. The validation set is used to evaluate a given version. Machine studying engineers used these records to fine-music the version hyperparameters. Data collection, records analysis, and for that reason the approach of addressing records content, quality, and shape can upload as much as a time-ingesting to-do list. During the technique of records identification, it allows you to recognize your records and its properties; this know-how will assist the system select which algorithms to construct the model.



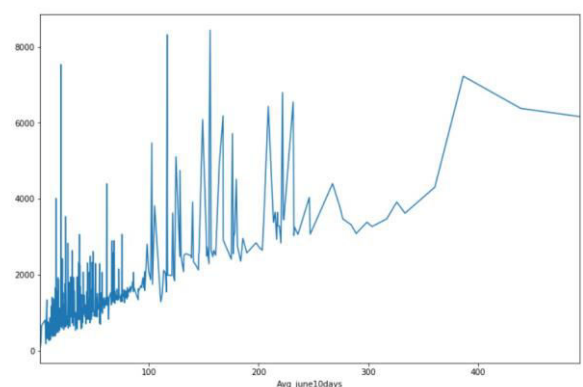


Create a predicted variable by rainfall range:

Data visualization is an essential talent in machine learning. Data visualization presents an essential suite of equipment for gaining a qualitative understanding. This may be beneficial while exploring and studying a dataset and might assist with figuring out patterns, corrupt data, outliers, and plenty extra. With a touch area knowledge, data visualizations may be used to specify and display key relationships in plots and charts which might be extra visceral and stakeholders than measures of affiliation or significance. Data visualization and exploratory statistics analysis are complete fields themselves and it'll advise a deeper dive into a number of the books stated on the end. Sometimes data does now no longer make sense till it could be checked out in a visible form, inclusive of charts and plots. Being capable of fast visualizing statistics samples and others is an essential talent in implementing facts and in implementing machine learning. It will find out the numerous kinds of plots that you'll want to recognize while visualizing statistics in Python and a way to use them to higher apprehend your personal statistics.

Performance measurements of ML algorithms:

Many machine learning algorithms are sensitive to the variety and distribution of characteristic values inside the input records. Outliers in input records can skew and deceive the training technique of machine learning algorithms ensuing in longer education times, much less accurate fashions, and in the end poorer results. Even earlier than predictive models are organized on training data, outliers can bring about deceptive representations and flip deceptive interpretations of collected records. Outliers can skew the precise distribution of characteristic values in descriptive records like imply and well-known deviation and in plots consisting of histograms and scatter plots, compressing the frame of the records. Finally, outliers can constitute examples of record times that can be applicable to the problem consisting of anomalies in the case of fraud detection and laptop security. For this, we ought to guarantee that the model was given the correct styles from the records. Cross-validation is a method wherein we teach our version the use of the subset of the records-set after which we evaluate the use of the complementary subset of the records-set.



Performance of Machine learning parameters

PARAMETER S	LR	SVC	KNN	RF	NB
Precision	0.99	0.96	0.97	0.98	0.99
Recall	0.96	1	0.98	0.98	0.99
F1-score	0.98	0.98	0.98	0.98	0.99
Sensitivity	0.96	1	0.98	0.97	0.98
Specificity	0.75	0	0.37	0.95	0.87
Accuracy(%)	95.33	95.85	95.85	97.85	98.44



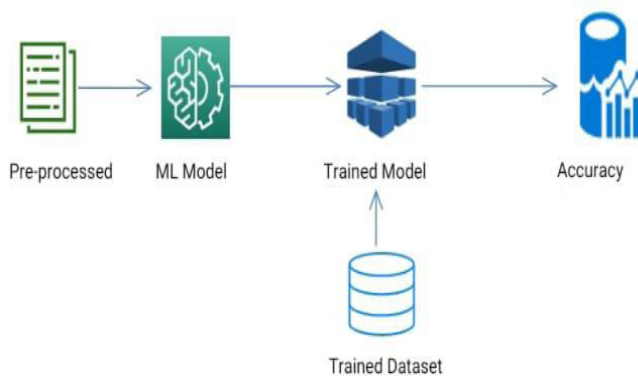
9. FUTURE ENHANCEMENT

Disaster management wants to automate detecting if the flash flood happened or not from the eligibility process (real time). To automate this process by showing the prediction result in web application or desktop application. To optimize the work to implement in Artificial Intelligence environment

GUI based prediction of flood by rainfall :

Tkinter is a python library that is used for developing GUI (Graphical User Interfaces). We use the Tkinter library for creating software of UI (User Interface), to create home windows and all different graphical consumer interfaces and Tkinter will include Python as a preferred package, it could be used for the safety motive of every customer or accountant.

7.SYSTEM ARCHITECTURE



8.RESULTS



10. CONCLUSION

There have been confined attempts in the prediction of Flash Flood. Radar changed into the most effective alternative used for the prediction of Flood. So, the primary concept of this undertaking is to offer prediction the use of ML algorithms. For which the analytical manner changed into use which began out from information cleansing and processing, lacking value, exploratory analysis, version constructing and evaluation. Finally, we are expecting floods, the use of exceptional devices gaining knowledge of algorithms, every providing distinct results. Logistic Regression, Support Vector Machine, Random forest, Naive Bayes are used. This brings a number of the subsequent insights to approximate flood prediction. This indicates an extensive version for the prediction of flash flood using Machine Learning.

REFERENCES

1. Akshaya J and Priyadarsini K, "A Hybrid Machine Learning Approach for Classifying Aerial Images of Flood-Hit Areas," in Second International Conference on Computational Intelligence in Data Science, 2019.
2. Amir Mosavi., Pinar Ozturk and Kwok-wing Chau, "Flood Prediction Using Machine Learning Models: Literature Review," MDPI, p. 40, 2018.
3. Amitkumar B and Durge P, "Different Techniques of Flood Forecasting and their Applications," 2018.
4. Analyn N., Charmaine C., Arnold C., Glenn O., Jose Angelo C., JanRalley P., and Jesse Dave S., "Real-Time Flood Water Level Monitoring System with SMS Notification," 2017.
5. Bipendra Basnyat., Nirmala Roy., and Aryya Gangopadhyay., "A Flash Flood Categorization System using Scene-Text Recognition," in the International Conference on Smart Computing, 2018.
6. Fazlina Ahmat Ruslan., Abd Manan Samad and Ramli Adnan, "4 Hours NNARX Flood Prediction Model Using "traingd" and "trainoss" Training Function: A Comparative Study," in 2018 IEEE 14th International Colloquium on Signal Processing & its Applications (CSPA 2018), 9 -10 March 2018, Penang, Malaysia, 2018.

7. Singh, K.K. and Singh, A., (2017), "Identification of flooded area from satellite images using Hybrid Kohonen Fuzzy C-Means sigma classifier", The Egyptian Journal of Remote Sensing and Space Science, 20(1), pp.147-155.

8. M. Ancona, A. Dellacasa, G. Delzanno, A. La Camera and I. Rellini, "An "Internet of Things" Vision of the Flood Monitoring Problem", The Fifth International Conference on Ambient Computing, Applications, Services And Technologies, 2015.

9. S. Gangopadhyay and M. Mondal, "A Wireless Framework for Environmental Monitoring and Instant Response Alert", International Conference on Microelectronics, Computing and Communications (MicroCom), 23-25 Jan, 2016.

10. P. Mitra, R. Ray, R. Chatterjee, R. Basu, P. Saha, S. Raha, R. Barman, S. Patra, S. Saha Biswas and S. Saha, "Flood forecasting using Internet of things and Artificial Neural Networks", Information Technology, Electronics and Mobile Communication Conference, IEEE, 2016.