iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IV **Month of publication:** April 2022

DOI: https://doi.org/10.22214/ijraset.2022.41297

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

Flood Prediction Using ML Classification Methods on Rainfall Data

Thegeshwar Sivamoorthy[1], Asif Mohammed Ansari[2], Dr. B. Sivakumar[3], V. Nallarasan[4]
[1, 2, 3, 4]*Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India*

Abstract: Floods are among the most destructive, most complex natural disasters to mimic. Research on the development of flood forecast models has contributed to risk reduction, a policy proposal, reduction of human lives, and mitigation of flood-related property damage. To mimic the complex statistical manifestations of natural flood processes, over the past two decades, neural network approaches have contributed significantly to the development of predictive systems that provide better performance and cost-effective solutions. To prevent this problem predict the occurrence of floods or not with a rain database by investigating neural network-based strategies. Database analysis by Multi-Layer Perceptron Classifier (MLP) to capture a number of details such as dynamic identification, deficit treatment, data validation, and data cleaning/preparation will be done across the given database. To apply flood forecasts for or without accurate calculation in the class division report, find the confusion matrix and the result shows the efficiency of the python frame-based flask based on the given attributes.
Keywords: Flood forecast, decision tree, Random Forest, Descent, SVM, Flasks.

I. INTRODUCTION

The flood problem is as old as time. However, although the natural floods of large areas did not create the most dangerous conditions in the prehistoric world, the increase in human activity and cities has led to the prevention of flood damage. Since the end of the eighteenth century, with the advent of the industrial era, there have been two phases of action: hydraulic activities in the area, such as land reclamation activities, which often undermine global balance-based, river flow, especially in mountainous and mountainous areas, leading to flooding at first.

II. LITERATURE SURVEY

The timely ongoing flood alarms classification in industrial monitoring systems is essential to be able to provide a safe and efficient operation. It can provide online support so that plant operators can take action in time, without having to wait for the alarm. A data-driven approach is proposed this article to address the issue of premature classification of non-label history data. In order to give foremost importance to pre-set alarms and take utmost benefit of the alarm time alert information, a vector representation exponentially attenuated component (EAC) is used to forecast alarm levels. In this article, a slow-moving approach to GMM-based data was planned to address the problem of premature segregation of continuous alarm floods with not documented historic data. Also, a vector presentation called the EAC was used to modify the sequence of alarm levels into vectors, which abridged the computational complexity they ran into in online mining methods. It cannot better determine the frequency of rainfall data and obtain more accurate flood prediction results due to a lack of results analysis in the form of confusion matrix

III. PREDICTION TECHNIQUES USED

1) *Logistic Regression:* This predicts the outcome of a class-based variance. Therefore, the result will be a grouping or different value. Either yes or no or 0 or 1, true or false, etc. but instead of giving a straight forward value such as 0 and 1, it provides probable values between 0 and 1.
2) *Decision Tree:* Decision Tree is a supervised learning method which can be used for classification and regression problems, both. It is a tree-shaped divider, where the internal nodes signify the elements of the dataset, the branches signify the rules of decision and each leaf node signifies the result.
3) *Random Forest:* This is a unique method in machine learning used to resolve classification and regression difficulties. It uses ensemble (integrated) learning. Ensemble learning is a multidisciplinary approach to providing answers to multifaceted problems. This algorithm contains many decision trees.
4) *SVM:* One of the most popular algorithms to solve Classification and Regression problems which is built with Supervised Learning in mind, it is mainly used for Classification Problems in Machine Learning.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

5) *KNN:* This is a non-parametric supervised learning method. It is used for classification and regression. The input consists of the *k* closest training instances in a data set for both Classification and Regression. Whether K-NN is used for classification or regression depends on the output. The output is a class membership in K-NN classification.
6) *Naïve Bayes:* It is a division of "probabilistic classifiers" which is based on relating Bayes' theorem with strong independence expectations between the structures. They belong to basic Bayesian network models, however it is combined with kernel density approximation, High accuracy level can be achieved.
7) *Standard Scaler (Enhancement Algorithm):* It is a technique which can be used to regulate the range of variables or features which are not dependent of data. As the collection of values of fresh data differs, in few machine learning algorithms, objective functions do not work correctly without normalization.

IV. DATA DESCRIPTION

In this project we took rain data from a well-known database site called Kaggle. The database (.CSV) has a size of 597KB and contains monthly rainfall details of less than 36 parts of India's meteorological metrological data. The data consists of 641 lines and 21 columns indicating the amount of rainfall in each region in India received in 1951-2000. Each column has a data parameter such as the name of the region, the month of data collection, the total rainfall of the year, the occurrence of floods and the estimates of specific months. We select this dataset to analyze and predict flood events. A small section of the dataset is displayed in Table (i).

	STATE_UT	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec	flood	Avg_june:	maytojune
0	ANDAMAI	107.3	57.9	65.2	117	358.5	295.5	285	271.9	354.8	326	315.2	250.9	2805.2	165.2	540.7	1207.2	892.1	0	98.5	63
1	ANDAMAI	43.7	26	18.6	90.5	374.4	457.2	421.3	423.1	455.6	301.2	275.8	128.3	3015.7	69.7	483.5	1757.2	705.3	0	152.4	82.8
2	ANDAMAI	32.7	15.9	8.6	53.4	343.6	503.3	465.4	460.9	454.8	276.1	198.6	100	2913.3	48.6	405.6	1884.4	574.7	0	167.7667	159.7
3	ARUNACH	42.2	80.8	176.4	358.5	306.4	447	660.1	427.8	313.6	167.1	34.1	29.8	3043.8	123	841.3	1848.5	231	0	149	140.6
4	ARUNACH	33.3	79.5	105.9	216.5	323	738.3	990.9	711.2	568	206.9	29.5	31.7	4034.7	112.8	645.4	3008.4	268.1	1	246.1	415.3
5	ARUNACH	28	48.3	85.3	101.5	140.5	228.4	217.4	182.8	159.8	75.9	20.9	11.6	1300.4	76.3	327.3	788.4	108.4	0	76.13333	87.9
6	ARUNACH	42.2	72.7	141	316.9	328.7	614.7	851.9	500.6	418.3	218.7	42.9	22.9	3571.5	114.9	786.6	2385.5	284.5	0	204.9	286
7	ARUNACH	42.2	80.8	176.4	358.5	306.4	447	660.1	427.8	313.6	167.1	34.1	29.8	3043.8	123	841.3	1848.5	231	0	149	140.6
8	ARUNACH	83.7	153.9	303.5	383.6	268	374.2	272	160.5	266.7	167.2	64	56	2553.3	237.6	955.1	1073.4	287.2	0	124.7333	106.2
9	ARUNACH	70.3	170.9	367.9	554.4	334.2	526.2	460.8	291.5	353.6	275	64.9	74.2	3543.9	241.2	1256.5	1632.1	414.1	0	175.4	192
10	ARUNACH	33.5	67.8	106.1	226.9	453	640.5	609.5	503.4	492.3	214.7	19.2	11.3	3378.2	101.3	786	2245.7	245.2	0	213.5	187.5
11	ARUNACH	97.5	109.3	92.4	204.3	266.2	284.1	248.9	270.5	192.7	78.5	49.5	27.2	1921.1	206.8	562.9	996.2	155.2	0	94.7	17.9
12	ARUNACH	74.3	176.7	362.6	397.5	408.7	801.9	653	417.9	686	264.9	86.9	71.7	4402.1	251	1168.8	2558.8	423.5	1	267.3	393.2
13	ARUNACH	26	66.7	76.8	229.2	239.5	416.6	592.4	312.4	291.1	126.8	33.7	29.5	2440.7	92.7	545.5	1612.5	190	0	138.8667	177.1
14	ARUNACH	83.7	153.9	303.5	383.6	268	374.2	272	160.5	266.7	167.2	64	56	2553.3	237.6	955.1	1073.4	287.2	0	124.7333	106.2
15	ARUNACH	35.2	43.5	58.9	134.3	341.1	665.3	749.9	579.1	490.9	233.9	40.3	27	3399.4	78.7	534.3	2485.2	301.2	1	221.7667	324.2
16	ARUNACH	49	74.4	96.5	156.9	208	345.7	368.5	256.2	275.9	138.2	34.4	27.2	2030.9	123.4	461.4	1246.3	199.8	0	115.2333	137.7
17	ARUNACH	35.2	43.5	58.9	134.3	341.1	665.3	749.9	579.1	490.9	233.9	40.3	27	3399.4	78.7	534.3	2485.2	301.2	1	221.7667	324.2
18	ARUNACH	82.7	70	128.2	245.7	271.4	292.7	404	276.3	283.5	92.3	32.3	42.4	2221.5	152.7	645.3	1256.5	167	0	97.56667	21.3
19	ASSAM	13.3	50.2	168.3	262.5	386.4	532.1	526.2	470.8	360.8	182.4	34.8	11.4	2999.2	63.5	817.2	1889.9	228.6	0	177.3667	145.7
20	ASSAM	13.1	21.4	53.5	168.8	320	419.7	345.8	272.1	221.5	95.4	17.2	9.3	1957.8	34.5	542.3	1259.1	121.9	0	139.9	99.7
21	ASSAM	12.7	20.4	51.1	196.6	399.8	567.8	502.8	334.6	304.9	157.7	21.7	5.2	2575.3	33.1	647.5	1710.1	184.6	0	189.2667	168
22	ASSAM	12	20.8	58.6	151.7	293.4	365.5	345.1	248.7	188.4	106.6	15.1	7.5	1813.4	32.8	503.7	1147.7	129.2	0	121.8333	72.1

Table (i) Sample Dataset

V. LIST OF MODULES

A. Data Pre-processing
B. Data Analysis of Visualization
C. Implementation of Logistic Regression
D. Implementation of Random Forest
E. SVM Implementation
F. Decision Tree Implementation
G. KNN Implementation
H. Naïve Bayes Implementation
I. Standard Scalar Random Forest (Enhancement) Implementation
J. Deployment Using Flask

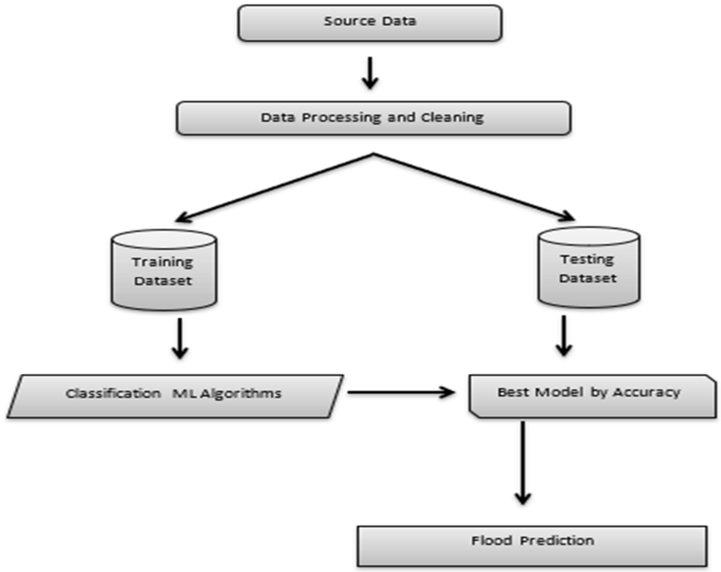


Fig (i) Block Diagram

VI. RESULT AND DISCUSSION

We have considered previous rainfall data where rainfall patterns for the months January and February, March to May, June to September and October to December are the characteristics used to predict future flood emergencies. We used four flood prediction methods such as Linear Regression Prediction, Decision Tree, Random Forest and SVM and collected the accuracy and effectiveness of these four methods. We then compared which algorithm had the most accuracy and best performance, and used the algorithm with the best performance in the machine learning model.

Model	Accuracy
Logistic Regression	71.3
Random Forest	80.8
SVM	60.5
Decision Tree	69.6
KNN	60.6
Naïve Bayes	62.2
Standard Scalar Logistic Regression	71.3
Standard Scalar Random Forest	99.5

Table (ii) Accuracy Description

Model	Sensitivity	Specificity
Logistic Regression	0.61	0.79
Random Forest	0.95	0.95
SVM	0.15	1.0
Decision Tree	0.96	0.93
KNN	0.47	0.95
Naïve Bayes	0.82	0.71
Standard Scalar Logistic Regression	0.61	0.79
Standard Scalar Random Forest	0.97	0.96

Table (iii) Performance Description

Model	TP	TN	FP	FN
Logistic Regression	46.63	16.06	11.92	25.39
Random Forest	56.48	1.04	2.07	40.41
SVM	58.55	35.23	0.00	6.22
Decision Tree	55.96	1.04	2.59	40.41
KNN	55.96	21.76	2.59	19.69
Naïve Bayes	41.97	7.25	16.58	34.20
Standard Scalar Logistic Regression	46.63	16.06	11.92	25.39
Standard Scalar Random Forest	56.48	1.04	2.07	40.41

VII. CONCLUSION AND THE FUTURE WORK

The systematic process initiated by the process of data cleaning, processing the missing values, exploratory analysis and in the end building the model for evaluation. The performance and accuracy on test data is taken into consideration and the model with highest performance and accuracy is implemented in the machine learning model. This application can help to predict future floods due to rainfall. The Random Forest algorithm has been implemented in the project website using Flask as it has the best accuracy and performance out of all the four algorithms tested.

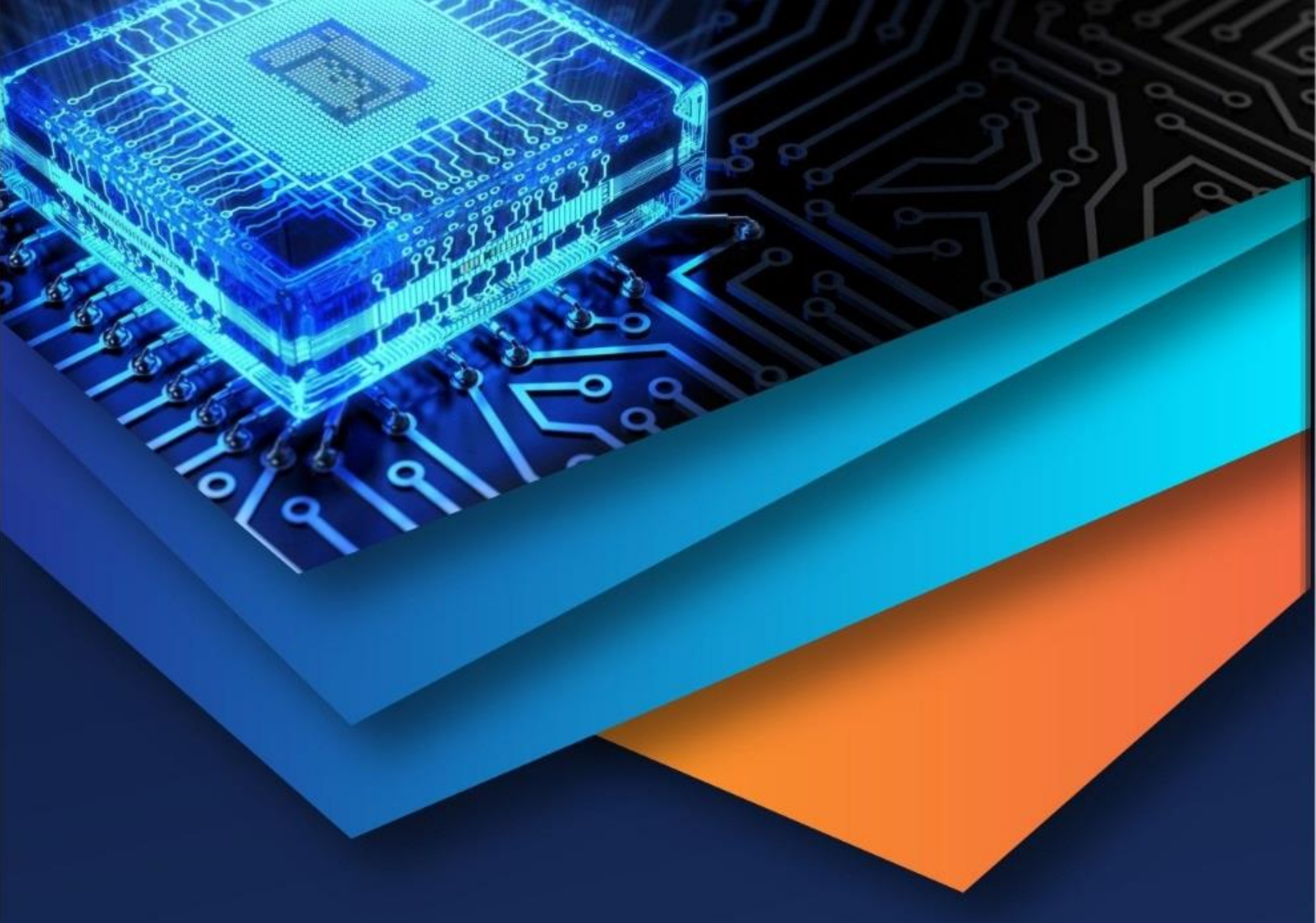
A. Future Work

Predicting flood using data from other natural occurrences such as:

1) Water level data
2) Seismic data

REFERENCES

[1] Pengzhan Cui, Yeqing Guan, Ying Zhu. "Flood Loss Prediction of Coastal City Based on GM-ANN" International Conference on Grey Systems and Intelligent Services (GSIS):2017
[2] Swapnil Bande, Virendra V. Shete. "Smart flood disaster prediction system using IOT & Neural Networks" International Conference On Smart Technologies For Smart Nation (SmartTechCon):2017
[3] Febus Reidj G. Cruz, Matthew G. Binag, Marlou Ryan G, Francis Aldrine A. "Flood Prediction Using Multi-Layer Artificial Neural Network in Monitoring System with Rain Gauge, Water Level, Soil Moisture Sensors" TENCON 2018 - 2018 IEEE Region 10 Conference:2018
[4] Indrastanti R. Widiasari, Lukito Edi Nugoho, Widyawan, Rissal Efendi. "Context-based Hydrology Time Series Data for A Flood Prediction Model Using LSTM" 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE):2018
[5] Haniyeh Seyed Alinezhad, Jun Shang and Tongwen Chen. ”Early Classification of Industrial Alarm Floods Based on Semi-Supervised Learning” IEEE :2021
[6] T Gurleen Kaur, Anju Bala. ” An Efficient Automated Hybrid Algorithm to Predict Floods in Cloud Environment” IEEE Xplore :2019
[7] Manomy K V, Meghna C S, Renuka Jayan1, Raghi R Menon. ”Flood Prediction and Tracking Trapped” IJERT Vol 9,Issue 6 :2020
[8] J. E. Reynolds, S. Halldin, J. Seibert, C.Y. Xu & T. Grabs .”Flood prediction using parameters calibrated on limited discharge data and uncertain rainfall scenarios” Hydrological Sciences Journal, 65:9 :2020
[9] Xiaodong Ming, Qiuhua Liang, Xilin Xia, Dingmin Li, and Hayley J. Fowler . ”Real-Time Flood Forecasting Based on a High- Performance 2-D Hydrodynamic Model and Numerical Weather Predictions” AGU Water Resources Research Vol 56 Issue 7 :2019
[10] Sharad Kumar Jain, Pankaj Mani, Sanjay K. Jain, Pavithra Prakash, Vijay P. Singh, Desiree Tullos, Sanjay Kumar, S. P. Agarwal & A. P. Dimri. ”A Brief review of flood forecasting techniques and their applications.” DOI:10.1080/15715124.2017.1411920 :2017



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY