Flood Prediction and Rainfall Analysis using Machine Learning

YashasAthreya[1], VaishaliBV[1], SagarK[1], SrinidhiHR[2]

[1]UG Student, Dept. of Information Science and Engineering, NIE Institute of Technology, Mysuru, Karnataka, India.
[2]Associate Professor ,Dept. of Information Science and Engineering, NIE Institute of Technology, Mysuru, Karnataka, India.

Abstract - *Flood acts as a major cause of destruction, loss of property and life. This results in sociological as well as economic loss. Prediction plays a major role to avoid such calamities .This paper presents the Flood prediction and Rainfall analysis using Machine Learning .The main goal of employing this application is to prevent immediate impacts of flood .This application can be easily used by the common people or government to predict the occurrence of flood beforehand .Prediction of flood is done by analyzing using previous data, map the flood concentration, and then provide them necessary help through helpline for further evacuation or any other necessary precautions .Machine Learning algorithms such as Linear Regression ,Gaussian Naïve bayes are used to build the model.*

Key Words*: *Flood Prediction, Rainfall Analysis , Machine Learning ,Linear Regression ,Gaussian Naïve bayes.

1. INTRODUCTION

Floods have large social consequences for communities and individuals. As most people are well aware, the immediate impacts of flooding include loss of human life, damage to property, destruction of crops, loss of livestock, and deterioration of health conditions owing to waterborne diseases.

Due to the unpredictability of global warming, there has been an increase in natural disasters all around the world .The nation of India has been hit with several of them, the most damaging being floods. Even though the northeastern states experience annual floods, they have been hit by some not so favorable deluges .Although flood risks cannot be completely eliminated, real time flood forecasting models, as an important and integral part of a flood warning service, can help to provide timely flood warnings with an adequate lead time for the public to minimize flood damages. Rainfall readings are valuable to local emergency situations, assessing flood conditions and taking appropriate actions. Advanced warning provided by early detection is critical to saving lives in flood prone areas. Advanced mathematical modeling can bring about enough difference in time .Different models are available which are used for forecasting variety of scenarios like flood, rainfall etc .Artificial intelligence techniques like Artificial Neural Network (ANN) and the Support Vector Machine (SVM)have been introduced which acts as an efficient and flexible computational tool giving good results. However, the inability to produce highly accurate results due to values beyond the range has proven to be a serious limitation.

We focus on time series data along with building a highly accurate model by comparing the results using different Machine Learning algorithms .The algorithms with most accurate outcome is then used to build the final model.

2. RELATED WORKS

According to the survey, we got some of the methodologies used in the previous studies for this purpose. Those are Flood prediction using Multi-layered artificial neural network using linear regression, Artificial intelligence techniques like Artificial Neural Network (ANN) and the Support Vector Machine (SVM), Synthetic Unit Hydrograph of ITS-2, two level hierarchical predictions for flood detection using Artificial Neural Networks. The dam opening is the first level of prediction. The water from the dam is released to canals and flow rate and level of water in one portion of canals is used to predict the possibility of a flood in the next portion of the canal. This constitutes the succeeding level of flood prediction.

The drawbacks of the above mentioned methodologies are accuracy, since there is a vast number of data available ,limiting the usage of data to train would lead to low accuracy .As well as limitations in certain topography due to results limited only to sheds. 70-30 split of training data to test data used in the experiment can be increased to provide better results.

The inability to produce highly accurate results due to values beyond the range has proven to be a serious limitation. Further work needs to be done in processing data since the samples used for training are not adequate to satisfy the consistency

Looking at the survey we have done, we can see that the disadvantages seem to be present in costs, datasets and software. So to counter this we decided on an extremely low cost approach by using Machine Learning Algorithms to tackle our problem statement. According to survey we can see the advantage of using the certain algorithms which provide better and accurate results.

Also by implementing a rainfall analysis module we can analyze the rainfall for a particular region .Regression algorithms play a beneficial role in rainfall prediction .Our model uses High Level architecture which makes it reliable

3. METHODOLOGY

We basically run our application in local host. There are two modules in our application:

1. **RAINFALL:** This module aims in prediction of rainfall of a particular region on a particular day using certain attributes such as temperature, humidity, Dew point, Sea level pressure etc.

2. **FLOOD:** This module aims in prediction of Flood of a particular division followed by entering the rainfall details of each month.

   The model uses high level architecture where in it is trained and fit using the Machine Learning algorithm and then the prediction is used to assess performance of model. Flask Jinja behaves as an interface between the user and the model.

HIGH LEVEL ARCHITECTURE:

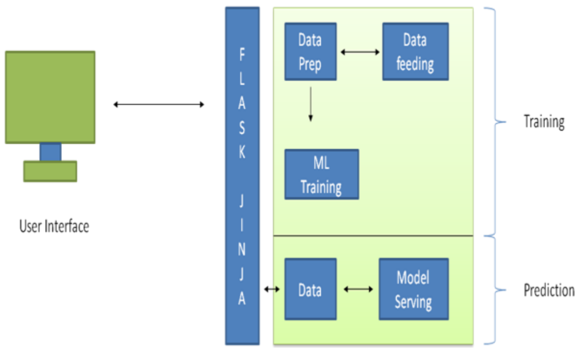


Fig -1: Architecture Diagram.

3.1 PROPOSED SYSTEM.

This application is used for predicition of flood and rainfall analysis which reduces the loss resulted due to them. We are using different ML Algorithms to firstly obtain an efficient one .We are using two different datasets for Flood and Rainfall .We tested the accuracy of regression algorithms for rainfall using dataset of one division and finally implemented the best one on the whole dataset.

Flood prediction module was built by testing the accuracy of algorithms on the dataset and finally the best one was used to build the model.

Advantages of the proposed system are as follows.

1. Huge data is used which results in accurate prediction.
2. Work done at less time.
3. This paper proposes a new method of flood prediction as well as analysis of rainfall.
4. This process is highly accurate.

Our proposed system overcomes the following drawbacks over the existing system.

1. Inadequate results from previous model.
2. Topographic issues.
3. The inability to produce highly accurate results.

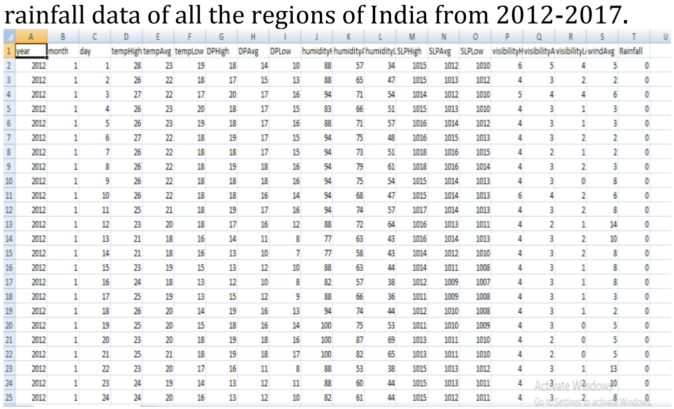
3.2 IMPLEMENTATION

We develop a application in two levels i.e. we build models using ML algorithms ,find the best one among them which also includes training and testing the data .Then the efficient modules are used to build a final model using Flask Jinja2.

Jinja2 is a template engine written in pure Python. It provides a non-XML syntax but supports inline expressions and an optional sandboxed environment. Flask is a Python-based micro web framework which allows you to write your web applications quickly and efficiently.

3.2.1 RAINFALL MODULE

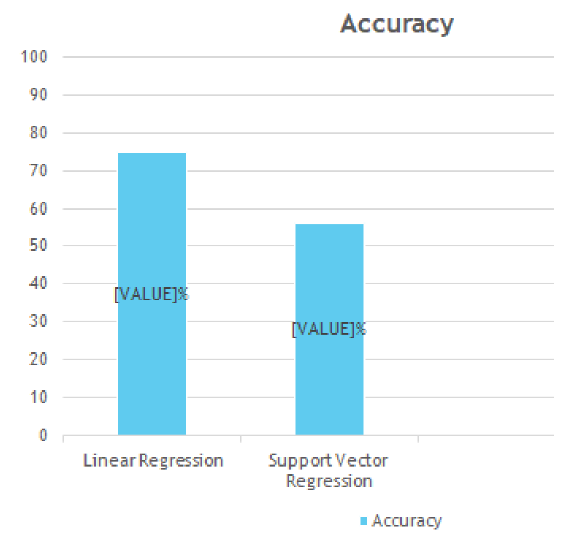
To build this module we have used the dataset from https://www.kaggle.com/. It is the world's largest data science community with powerful tools and resources to help you achieve your data science goals. The dataset has rainfall data of all the regions of India from 2012-2017.



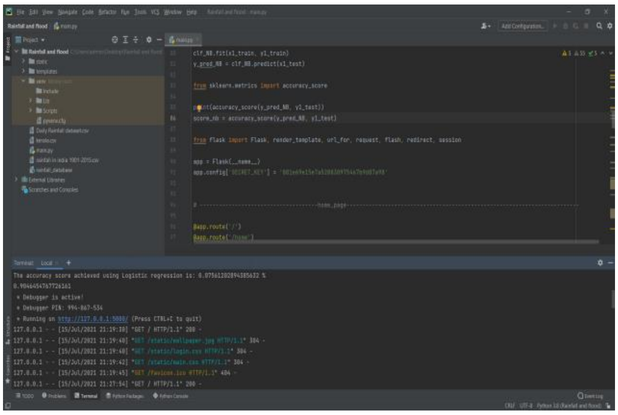
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	year	month	day	tempHigh	tempAvg	tempLow	DPHigh	DPAvg	DPLow	humidityH	humidityA	humidityL	SLPHigh	SLPAvg	SLPLow	visibilityH	visibilityA	visibilityL	windAvg	Rainfall	
2	2012	1	1	28	23	19	18	14	10	88	57	34	1015	1012	1010	6	5	4	5	0	
3	2012	1	2	26	22	18	17	15	13	88	65	47	1015	1013	1012	4	3	2	2	0	
4	2012	1	3	27	22	17	20	17	16	94	71	54	1014	1012	1010	5	4	4	6	0	
5	2012	1	4	26	23	20	18	17	15	83	66	51	1015	1013	1010	4	3	1	3	0	
6	2012	1	5	26	23	19	18	17	16	88	71	57	1016	1014	1012	4	3	1	3	0	
7	2012	1	6	27	22	18	19	17	15	94	75	48	1016	1015	1013	4	3	2	2	0	
8	2012	1	7	26	22	18	18	17	15	94	73	51	1018	1016	1015	4	2	1	2	0	
9	2012	1	8	26	22	18	19	18	16	94	79	61	1018	1016	1014	4	3	2	3	0	
10	2012	1	9	26	22	18	18	18	16	94	75	54	1015	1014	1013	4	3	0	8	0	
11	2012	1	10	26	22	18	18	16	14	94	68	47	1015	1014	1013	6	4	2	6	0	
12	2012	1	11	25	21	18	19	17	16	94	74	57	1017	1014	1013	4	3	2	8	0	
13	2012	1	12	23	20	18	17	16	12	88	72	64	1016	1013	1011	4	2	1	14	0	
14	2012	1	13	21	18	16	14	11	8	77	63	43	1016	1014	1013	4	3	2	10	0	
15	2012	1	14	21	18	16	13	10	7	77	58	43	1014	1012	1010	4	3	2	8	0	
16	2012	1	15	23	19	15	13	12	10	88	63	44	1014	1011	1008	4	3	1	8	0	
17	2012	1	16	24	18	13	12	10	8	82	57	38	1012	1009	1007	4	3	1	8	0	
18	2012	1	17	25	19	13	15	12	9	88	66	36	1011	1009	1008	4	3	1	3	0	
19	2012	1	18	26	20	14	19	16	13	94	74	44	1012	1010	1008	4	3	1	2	0	
20	2012	1	19	25	20	15	18	16	14	100	75	53	1011	1010	1009	4	3	0	5	0	
21	2012	1	20	23	20	18	19	18	16	100	87	69	1013	1011	1010	4	2	0	5	0	
22	2012	1	21	25	21	18	19	18	17	100	82	65	1013	1011	1010	4	2	0	5	0	
23	2012	1	22	23	20	17	16	11	8	88	53	38	1015	1013	1012	4	3	1	13	0	
24	2012	1	23	24	19	14	13	12	11	88	60	44	1015	1013	1011	4	3	2	10	0	
25	2012	1	24	24	20	16	13	12	10	82	61	44	1015	1012	1011	4	3	2	8	0	

Firstly the Dataset containing several attributes (temperature, humidity, Dew point, Sea level pressure etc.), is used hence helping in building a highly accurate model. Dataset is then analyzed using Support Vector Regression and Linear Regression.

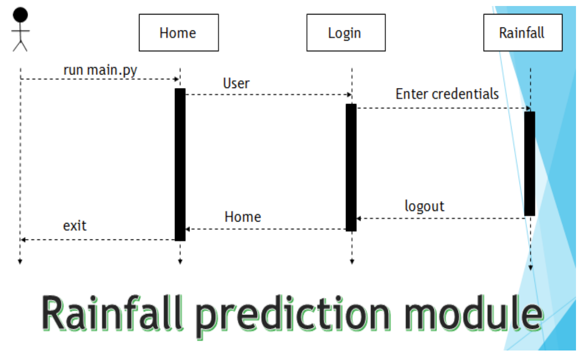
Linear Regression algorithm produced a higher accuracy of 75% (56% for Support Vector Regression).Hence we continued with Linear Regression for further process.



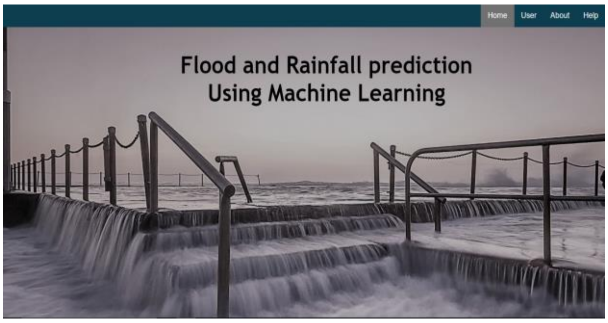
This module is implemented in PyCharm IDE using the linear regression algorithm, followed by the template implementation using Flask Jinja2.



The Sequence of events that take place when a user enters the rainfall module is as follows.

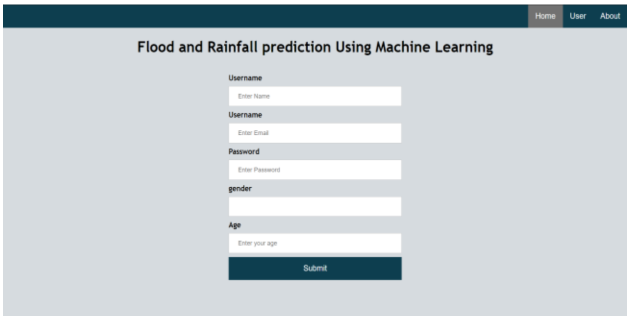


When the user runs the application he is redirected to the home page.

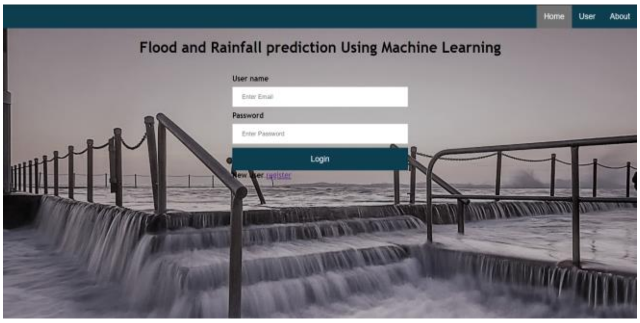


In the home page the user can access the modules by registering with necessary credentials and later login using the same.

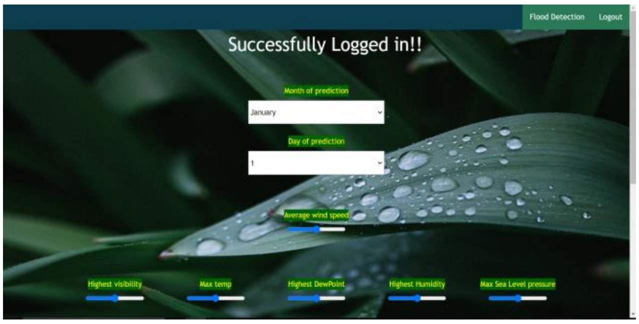
Registration Page



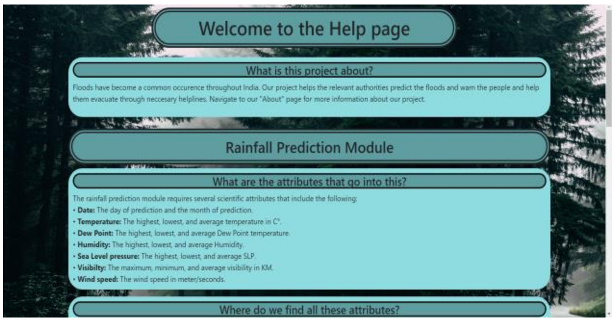
Login Page



On successful Login he will be redirected to the Rainfall module where on entering the necessary attributes the user can predict the rainfall in millimeters.



The attributes might seem quite confusing for a normal user, thus we have included a help page where the attributes are been defined for a clear picture



To reduce the hassle we have also provided a website link which has all the values for the attributes of a particular day. Using this data the user will be able to predict the rainfall.

3.2.2 FLOOD MODULE

Similar to rainfall module we have used the dataset from https://www.kaggle.com/. The dataset has data of all the subdivision's of India from 1901-2015.The dataset includes the data of 35 subdivision's of India.

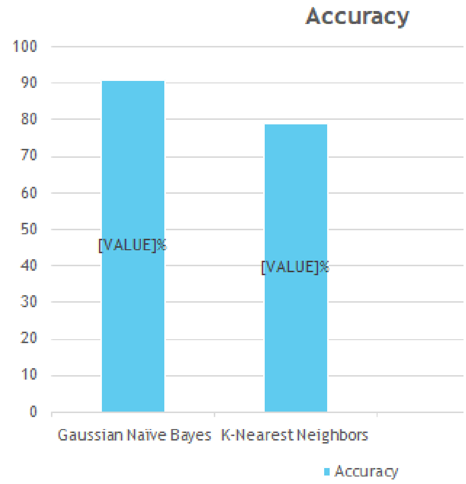
1	SUBDIVISI	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec	Flood
2	ANDAMAI	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.6	3373.2	136.3	560.3	1696.3	980.3	YES
3	ANDAMAI	1902	0	159.8	12.2	0	446.1	537.1	228.9	753.7	666.2	197.2	359	160.5	3520.7	159.8	458.3	2185.9	716.7	YES
4	ANDAMAI	1903	12.7	144	0	1	235.1	479.9	728.4	326.7	339	181.2	284.4	225	2957.4	156.7	236.1	1874	690.6	NO
5	ANDAMAI	1904	9.4	14.7	0	202.4	304.5	495.1	502	160.1	820.4	222.2	308.7	40.1	3079.6	24.1	506.9	1977.6	571	YES
6	ANDAMAI	1905	1.3	0	3.3	26.9	279.5	628.7	368.7	330.5	297	260.7	25.4	344.7	2566.7	1.3	309.7	1624.9	630.8	NO
7	ANDAMAI	1906	36.6	0	0	0	556.1	733.3	247.7	320.5	164.3	267.8	128.9	79.2	2534.4	36.6	556.1	1465.8	475.9	NO
8	ANDAMAI	1907	110.7	0	113.3	21.6	616.3	305.2	443.9	377.6	200.4	264.4	648.9	245.6	3347.9	110.7	751.2	1327.1	1158.9	YES
9	ANDAMAI	1908	20.9	85.1	0	29	562	693.6	481.4	699.9	428.8	170.7	208.1	196.9	3576.4	106	591	2303.7	575.7	YES
10	ANDAMAI	1910	26.6	22.7	206.3	89.3	224.5	472.7	264.3	337.4	626.6	208.2	267.3	153.5	2899.4	49.3	520.1	1701	629	NO
11	ANDAMAI	1911	0	8.4	0	122.5	327.3	649	253	187.1	464.5	333.8	94.5	247.1	2687.2	8.4	449.8	1553.6	675.4	NO
12	ANDAMAI	1912	583.7	0.8	0	21.9	140.7	549.8	468.9	370.3	386.2	318.7	117.2	2.3	2960.5	584.5	162.6	1775.2	438.2	NO
13	ANDAMAI	1913	84.8	0.5	1.3	2.5	190.7	530	280.8	205.8	580.1	288.8	133	67.5	2365.8	85.3	194.5	1596.7	489.3	NO
14	ANDAMAI	1914	0	0	0	37.7	298.8	383.3	792.8	520.5	310.8	139.8	184.4	289.7	2957.8	0	336.5	2007.4	613.9	NO
15	ANDAMAI	1915	45	56.7	33.3	40.9	170.2	334.7	269	317.2	429.8	468.1	258.4	318	2741.3	101.7	244.4	1350.7	1044.5	NO
16	ANDAMAI	1916	0	0	0	0.5	487.4	450.1	317.3	425	561.2	369.7	192.6	133.7	2937.5	0	487.9	1753.6	696	NO
17	ANDAMAI	1917	8	3.6	112	4.5	295.9	301.1	394.8	437.4	471.8	238.1	108.3	236.9	2612.4	11.6	412.4	1605.1	583.3	NO
18	ANDAMAI	1918	77.4	6.9	11.4	10.7	729.3	710.8	200.9	455.4	303.3	227	366.9	175	3275	84.3	751.4	1670.4	768.9	YES
19	ANDAMAI	1919	10.2	18	0	35.5	283.9	542.5	246.5	259.8	170.7	186.2	340.4	258.4	2352.1	28.2	319.4	1219.5	785	NO
20	ANDAMAI	1920	122.3	7.4	3.1	13	237.4	546.9	294.4	467.4	505.4	397.5	262.9	85.5	2943.2	129.7	253.5	1814.1	745.9	NO
21	ANDAMAI	1921	13.2	3.1	0	37.5	351.2	282.7	487.1	330	581.2	360.7	118.2	41.5	2606.4	16.3	388.7	1681	520.4	NO
22	ANDAMAI	1922	245.3	34.3	15.6	323.1	289.7	506.1	425.8	307.4	511.7	162	541	192.2	3554.2	279.6	628.4	1751	895.2	YES
23	ANDAMAI	1923	79.5	0	NA	91.3	293.5	808.4	636.9	182.2	560.5	131.9	197.4	70.6	NA	79.5	NA	2188	399.9	NA
24	ANDAMAI	1924	28.7	0	14.8	89.7	191.2	261.2	493.3	290.9	251.2	331.1	378.6	NA	NA	28.7	295.7	1296.6	NA	NA
25	ANDAMAI	1925	36.6	0	8.6	50.4	282.2	663.8	241.8	278.2	201.9	249.5	271.5	196	2480.5	36.6	341.2	1385.7	717	NO

Meteorological data from past 115 years is used for the module. Dataset tuning is performed to remove unwanted and null data.

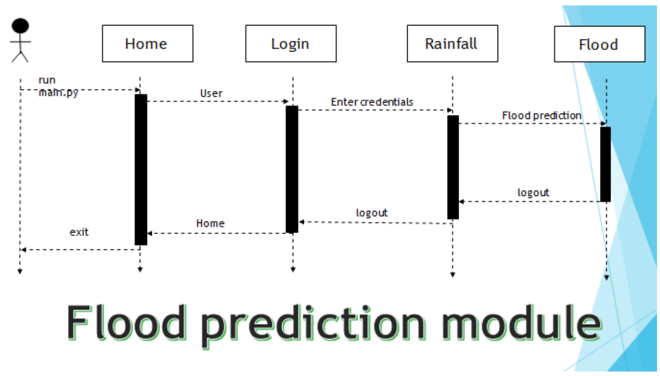
Classification algorithms are used for prediction module namely K-Nearest neighbors algorithm and Gaussian Naïve Bayes. As the dataset contained information from 35 subdivisions, we built a model for the subdivision of Kerala by using a separate dataset and implemented K-Nearest neighbors and Gaussian Naïve Bayes algorithm.

1	SUBDIVISI	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	FLOODS
2	KERALA	1901	28.7	44.7	51.6	160	174.7	824.6	743	357.5	197.7	266.9	350.8	48.4	3248.6	YES
3	KERALA	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205	315.8	491.6	358.4	158.3	121.5	3326.6	YES
4	KERALA	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2	341.8	354.1	157	59	3271.2	YES
5	KERALA	1904	23.7	3	32.2	71.5	235.7	1098.2	725.5	351.8	222.7	328.1	33.9	3.3	3129.7	YES
6	KERALA	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6	217.2	383.5	74.4	0.2	2741.6	NO
7	KERALA	1906	26.7	7.4	9.9	59.4	160.8	414.9	954.2	442.8	131.2	251.7	163.1	86	2708	NO
8	KERALA	1907	18.8	4.8	55.7	170.8	101.4	770.9	760.4	981.5	225	309.7	219.1	52.8	3671.1	YES
9	KERALA	1908	8	20.8	38.2	102.9	142.6	592.6	902.2	352.9	175.9	253.3	47.9	11	2648.3	NO
10	KERALA	1909	54.1	11.8	61.3	93.8	473.2	704.7	782.3	258	195.4	212.1	171.1	32.3	3050.2	YES
11	KERALA	1910	2.7	25.7	23.3	124.5	148.8	680	484.1	473.8	248.6	356.6	280.4	0.1	2848.6	NO
12	KERALA	1911	3	4.3	18.2	51	180.6	990	705.3	178.6	60.2	302.3	145.7	87.6	2726.7	NO
13	KERALA	1912	1.9	15	11.2	122.7	217.3	948.2	833.6	534.4	136.8	469.5	138.7	22	3451.3	YES
14	KERALA	1913	3.1	5.2	20.7	75.7	198.8	541.7	763.2	247.2	176.9	422.5	109.9	45.8	2610.8	NO
15	KERALA	1914	0.7	6.8	18.1	32.7	164.2	565.3	857.7	402.2	241	374.4	100.9	135.2	2899.1	NO
16	KERALA	1915	16.9	23.5	42.7	106	154.5	696.1	775.6	298.8	396.6	196.6	302.5	14.9	3024.5	YES
17	KERALA	1916	0	7.8	22	82.4	199	920.2	513.9	396.9	339.3	320.7	134.3	8.9	2945.3	YES
18	KERALA	1917	2.9	47.6	79.4	38.1	122.9	703.7	342.7	335.1	470.3	264.1	256.4	41.6	2704.8	NO
19	KERALA	1918	42.9	5	32.8	51.3	683	464.3	167.5	376	96.4	233.2	295.4	54.1	2501.9	NO
20	KERALA	1919	43	6.1	33.9	65.9	247	636.8	648	484.2	255.9	249.2	280.1	53	3003.3	YES
21	KERALA	1920	35.2	5.5	24.1	172	87.7	964.3	940.8	235	178	350.1	302.3	8.2	3303.1	YES
22	KERALA	1921	43	4.7	15	171.3	104.1	489.1	639.8	641.9	156.7	302.4	136.2	15.8	2719.9	NO
23	KERALA	1922	30.5	21.4	16.3	89.6	293.6	663.1	1025.1	320.6	222.4	266.3	293.7	25.1	3267.6	YES
24	KERALA	1923	24.7	0.7	78.9	43.5	80	722.5	1008.7	943	254.3	203.1	83.9	41.6	3484.7	YES
25	KERALA	1924	19.3	2.9	66.6	111	185.4	1011.7	1526.5	624	289.1	176.5	162.9	50.4	4226.4	YES

Gaussian Naïve Bayes shows a maximum accuracy of 91% and minimum accuracy of 88%.This Accuracy was higher than K-Nearest neighbors.

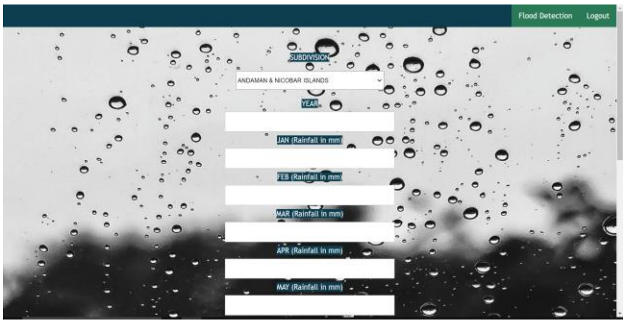


1. The Sequence of events that take place when a user enters the rainfall module is as follows.

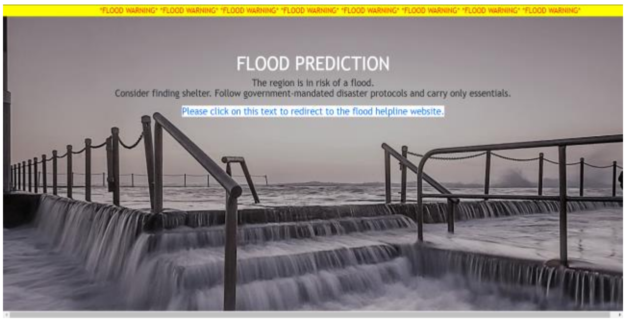


When the user runs the application he is redirected to home page wherein he logins or register's with necessary credentials and is redirected to the rainfall page.

Here on the top right corner he will able to find the flood module wherein entering the necessary attributes the user can predict the occurrence of Flood.



If the region is prone to flood occurrence the user is redirected to the helpline website with a Flood Warning message where further necessary precautionary measures can be taken or help can be received.



If the region is not in risk of flood he will be redirected to the page where a message saying "No risk of Flood" will be displayed with some guidelines of what has to be done in case of floods.

4. RESULTS

The presented results show the very high accuracy of flood prediction and rainfall analysis .Successful prediction of flood occurrence for the 10 years approximately can be obtained. It is cost effective and easy to handle by a common person.

5. CONCLUSION

The proposed Flood Prediction and Rainfall Analysis has been successfully implemented and tested with the available data. This is one of the best methods to predict the occurrence of flood. For future enhancements we can include Cloud Deployment of the application .Updating data utilizing live Information from sensors .Inserting more rescue groups as helpline .Also using advanced datasets taking into consideration climate change.

REFERENCES

- Dola Sheeba Rani, Dr. Jayalakshmi G.N, Dr. Vishwanath P Baligar (IEEE 2020)
- Chen Chen, Qiang Hui, Qingqi Pei, Yang Zhou, Bin Wang, Ning Lv, Ji Li (IEEE 2019)
- I Gede Tunas, Nadjadji Anwar (IEEE 2018)
- Riberio Alexandra, Cardoso Alberto, Marques Alfeu S., Simoes Nuno E (IEEE 2017)
- Jojene R. Santillan, Meriam Makinano-Santillan, Linbert C. Cutamora (IEEE 2016
- Metro Manila, Felan Carlo C. Garcia, Alvin E. Retamar, Joven C. Javier I(IEEE 2015)