INTERNATIONAL JOURNAL OF RESEARCH CULTURE SOCIETY ISSN: 2456-6683 Volume - 4, Issue - 6, June – 2020
Monthly, Peer-Reviewed, Refereed, Indexed Journal Scientific Journal Impact Factor: 5.245
Received on : 0506/2020 Accepted on : 20/06/2020 Publication Date: 30/06/2020

HEURISTIC PREDICTION OF RAINFALL USING MACHINE LEARNING

[1]Neha Appaji , [1]Chaitra B. R., [1]Poojashree, [2]Ramya S.
[1]Student, [2]Assistant Professor, Computer Science and Engineering,
The National Institute of Engineering, Mysuru, India
Autonomous under Visvesvaraya Technological University (VTU)
Email - 1nehaappaji@gmail.com 1chaitrabr98@gmail.com 1poojashree189pooja@gmail.com 2ramyas@nie.ac.in

Abstract: *India's main source of economy is agriculture. Rainfall prediction in advance is very necessary for agricultural purposes so that farmers can take important decisions on which crop to grow, how much quantity of fertilizers or other chemicals are needed for the proper cultivation of crops. Rainfall prediction is also necessary to avoid damages caused by natural disasters like flood etc. So in this paper, we have selected an actual dataset from 3 types of regions of Karnataka and we predict rainfall rate using machine learning technique by applying K-Nearest Neighbor supervised learning algorithm on dataset which we have collected manually from different regions weather forecasting departments. Results of our web based application will display the rate of rainfall in cm which will be helpful to agriculture and farmers thereby boosting the economy of our country.*

Keywords: *Rainfall prediction, agriculture, supervised, machine learning, K-Nearest Neighbour.*

1. INTRODUCTION:

Agriculture is the main source of Indian economy. Irrigation facility is still not so good in India and most of the agriculture based activities depends upon the amount of rainfall. Early prediction of rainfall is very necessary as too less or too heavy rainfall both the extremes are harmful and will affect the yield of crops thus affecting agriculture. Rainfall prediction in advance is very necessary for agricultural purposes so that farmers can take important decisions on which crop to grow, how much quantity of fertilizers or other chemicals are needed for the proper cultivation of crops. Rainfall prediction is also necessary to avoid damages caused by natural disasters like flood etc. Here we are using empirical approach to predict rate of rainfall which is based on analysis of past historical data of weather and its relationship to a variety of atmospheric variables over different parts of Karnataka such as North interior Karnataka, south interior Karnataka, and coastal Karnataka. We apply supervised machine learning technique to predict rainfall rate by applying K-Nearest Neighbor supervised learning algorithm on dataset which we have collected manually from different regions weather forecasting departments.

Through machine learning computers or machines will learn through experience and there is no need to explicitly program it. As it is evident from the name, it gives the computer that makes it more similar to humans: *The ability to learn.* Machine learning is actively being used today, perhaps in many more places than one would expect. Machine learning algorithms have the ability to improve themselves through training. In the proposed system we are using supervised machine learning technique. Supervised learning is a learning in which we train the machine by providing training dataset along with correct results. Then machine will analyse and learn through training dataset and predict the rainfall rate result for testing dataset.

In this paper we have made use of supervised learning algorithm KNN for classifying the input data and predicting the occurrence of rainfall by making use of various parameters such as wet bulb temperature, dry bulb temperature, humidity. Minimum temperature, maximum temperature, wind speed, average wind speed, wind direction. The K-Nearest-Neighbours (KNN) is a classification technique. It is also called lazy learning algorithm. This algorithm works as follows: We have to first find the value of K = number of nearest neighbors. Apply the Euclidean distance formula and find the distance between the respective parameters of testing dataset and all the training dataset. Sort the distances obtained in ascending order and find the 'K' nearest neighbors based on the value of K. After we obtained the K-Nearest Neighbors, we will gather their results and check the number of occurrences of each result. Result with maximum number of occurrences will be the final predicted result of testing dataset.

The proposed model tends to forecast rainfall based on the previous records of a particular geographic area, therefore, this prediction will prove to be much reliable. The performance of the model is more accurate when compared with traditional rainfall prediction systems.

2. LITERATURE SURVEY:

Choujun Zhan1, Fujian Wu1, Zhengdong Wu1, Chi K. Tse [1], "Daily Rainfall Data Construction and Application to Weather Prediction". In this paper various types of climatic data in time-series form is collected at observation sites or satellites through sensor technology. In this work, they have reconstructed a climate dataset including more than 30 climatic variables measured by more than 103,473 observation sites covering the world surface from 1800 to 2017. Then, they applied the state-of-the-art machine learning methods, including deep learning (CNN, RNN, and LSTM networks) and ensemble learning (Adaboost, GBDT, and XGBoost), to develop a short-term precipitation system. Experiments on real-world data show that incorporating multiple climate variables into a prediction system improves the prediction results. The best performance of their proposed method reached an accuracy of more than 80%.

Chandrasegar Thirumalai, M Lakshmi Deepak, K Sri Harsha, K Chaitanya Krishna [2], "Heuristic Prediction of Rainfall Using Machine Learning Techniques". Here they have selected the real dataset which consists of past years according to various crops seasons like rabi, Kharif, zaid and predicts the rainfall in future seasons. They have used linear regression method to predict the unknown value of season from known value of another season.

Tharun V.P, Ramya Prakash S, Renuga Devi[3], "Prediction of Rainfall Using Data Mining Techniques". In this paper, they are predicting the rainfall rate of Coonoor in Nilgiris district, Tamil Nadu. They have used regression techniques such as Support Vector Regression (SVR), Random forest (RF) and Decision Tree (DT) for predicting the rate of rainfall. They have implemented the proposed model in python and found out that RF technique was efficient compared to other two techniques.

3. PROPOSED SYSTEM:

Proposed system is an agriculture application which analyzes the previous data related to region and year on daily basis. Proposed system makes use of machine learning in agriculture for decision making. The research is conducted taking under consideration the various constraints such as temperature, humidity, region, wind speed, wind direction etc. System uses "*supervised learning*" technique - *KNN Algorithm* for rainfall prediction. Data collected from government sector (agriculture department) can be used to predict rainfall. The proposed system predicts rainfall for three types of region that is, North Interior Karnataka, South Interior Karnataka and Coastal Karnataka as per one district in each region.

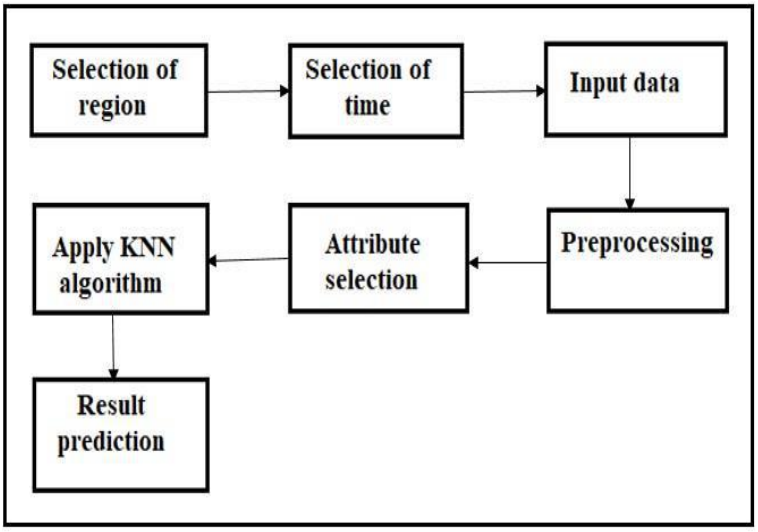

Figure 1: Proposed system architecture

Selection of region: Select the required region from the dropdown menu (Bidar, Shivmoga, Uttarakannada).

Selection of time: Select the time at which we want to view the rainfall rate (8.30AM, 5.30PM).

Input data: Data includes information regarding various attributes such as temperature, year, humidity, wind speed, etc, which is collected over some period of time.

Preprocessing: Data which is collected should be preprocessed. Redundant data, inconsistent should be taken care.

Attribute Selection: Important Features have to be extracted.

Apply KNN Algorithm: KNN algorithm is applied on the input data to predict rainfall rate.

Result Prediction: Rainfall rate can be provided to the farmers based on the results obtained.

4. METHODOLOGY:

Supervised machine learning

KNN Algorithm

Step 1: We have to first find the value of K = number of nearest neighbors.
Step 2: Apply the Euclidean distance formula and find the distance between the respective parameters of testing dataset and all the training dataset.
Step 3: Sort the distances obtained in ascending order and find the 'K' nearest neighbors based on the value of K.
Step 4: After we obtained the K-Nearest Neighbors, we will gather their results and check the number of occurrences of each result.
Step 5: Result with maximum number of occurrences will be the final predicted result of testing dataset.
Example:

Training dataset:

Attributes or Constraints: mintemp, maxtemp, drybulbtemp, wetbulbtemp, humidity, windspeed, avgwindspeed, winddirection.

SINo	mintemp	maxtemp	drybulbtemp	wetbulbtemp	humidity	windspeed	avewindspeed	winddirection	Result
1.	23.5	26	23	22.5	91	0.24	0	E	0
2.	24	28.5	23.5	23	96	2.3	0	E	6.2
3.	24	26	23.5	23	91	0.01	0	E	2.8
4.	23.5	26.5	23	22.5	91	0.13	0	SW	2.6
5.	23.5	26	23	22.5	91	0.24	0	E	0
6.	24	28.5	23.5	23	96	2.3	0	E	6.2
7.	23.5	27	23	22.5	95	0.13	0	S	1
8.	23.5	26	23	22.5	91	0.24	0	E	0
9.	23.5	32	23	22.5	91	0.23	0	E	5.8
10.	23.5	31	23	22.5	91	0.09	0	S	5.8
11.	23	30	23	22	95	0.75	0	N	7

TABLE 1: TRAINING DATASET

Sl No	mintemp	maxtemp	drybulbtemp	wetbulbtemp	humidity	windspeed	avewindspeed	winddirection	Result
12	23.5	29.5	23	22.5	91	0.22	0	SW	

TABLE 2: TESTING DATASET

Let K=5,

Euclidean distance formula:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Record Id	Euclidean Distance
1	3.5000
2	5.5746
3	3.6116

4	**3.0013**
5	**3.5000**
6	**5.5746**
7	**4.7178**
8	**3.5000**
9	**2.5000**
10	**1.5056**
11	**4.1268**

TABLE 3: EUCLIDEAN DISTANCE TABLE

After finding Euclidean distance we sort it in ascending order and list the first 5 nearest neighbours as K=5 along with their result.

Record Id	Distance	Result
10	1.5056	5.8
9	2.5000	5.8
4	3.0013	2.6
1	3.5000	0
5	3.5000	0.1

TABLE 4: K-NEAREST NEIGHBOURS TABLE

Here 5.8 has repeated maximum times than other results. So result for 12th record is 5.8.

5. RESULT:

SHIVMOGA JULY MONTH RAINFALL PREDICTION

RAINFALL PREDICTION (KNN ALGORITHM)!!!

SlNo	mintemp	maxtemp	drybulbtemp	wetbulbtemp	humidity	windspeed	avewindspeed	winddirection	Result
1.	23.5	28	23	22.5	91	0.24	0	E	0
2.	24	28.5	23.5	23	96	2.3	0	E	6.2
3.	24	28	23.5	23	91	0.01	0	E	2.8
4.	23.5	26.5	23	22.5	91	0.13	0	SW	2.6
5.	23.5	28	23	22.5	91	0.24	0	E	0
6.	24	28.5	23.5	23	96	2.3	0	E	6.2
7.	23.5	27	23	22.5	95	0.13	0	S	1
8.	23.5	28	23	22.5	91	0.24	0	E	0
9.	23.5	32	23	22.5	91	0.23	0	E	5.8
10.	23.5	31	23	22.5	91	0.09	0	S	5.8
11.	23	30	23	22	95	0.75	0	N	7
12.	23.5	29.5	23	22.5	91	0.22	0	SW	5.8
13.	23	30	22.5	22	95	0.75	0	N	7
14.	23	30	22.5	22	95	0.57	0	N	7
15.	22.5	27.5	22	21.5	91	0.16	0	E	6.2
16.	23.5	28	22.5	22	95	1.18	0	NW	5.8
17.	23	27.5	22.5	22	95	0.94	0	N	2.6
18.	23	30	22.5	22	95	0.75	0	N	7
19.	23	30	22.5	22	95	0.75	0	N	7
20.	23.5	28	23	22.5	91	0.24	0	E	0
21.	23	27.5	23	22	95	0.94	0	N	2.6
22.	23.5	26.5	23	22.5	91	0.41	0	SW	1.2
23.	24	28	23.5	23	96	0.33	0	E	3
24.	24.5	27	23.5	23	96	0.26	0	N	7
25.	24.5	28	24	23.5	91	1.1	0	SW	28
26.	24	28.5	23.5	23	96	2.3	0	E	6.2
27.	24	28.5	24	23	91	0.19	0	E	1.2
28.	23	27.5	22.5	22	95	0.94	0	N	2.6
29.	24	28.5	23.5	23	96	2.3	0	E	6.2
30.	23.5	27	23	22.5	95	0.13	0	S	1
31.	23.5	27	23	22.5	95	0.13	0	S	1

Constraint	KNN Algorithm
Accuracy	96.7741935483871%
Time (milli secs)	156
Correctly Classified	96.7741935483871%
InCorrectly Classified	3.2258064516129%

BIDAR JULY MONTH RAINFALL PREDICTION

RAINFALL PREDICTION (KNN ALGORITHM)!!!

SlNo	mintemp	maxtemp	drybulbtemp	wetbulbtemp	humidity	windspeed	avewindspeed	winddirection	Result
1.	33.5	36	33	32.5	98	0.24	0	E	0
2.	34	35.5	33.5	33	96	2.3	0	E	6.2
3.	34	36	33.5	33	98	0.01	0	E	2.8
4.	33.5	36.5	33	32.5	98	0.13	0	SW	2.6
5.	33.5	36	33	32.5	98	0.24	0	E	0
6.	34	35.5	33.5	33	96	2.3	0	E	6.2
7.	33.5	37	33	32.5	95	0.13	0	S	1
8.	33.5	36	33	32.5	98	0.24	0	E	0
9.	33.5	38	33	32.5	98	0.23	0	E	1.8
10.	33.5	31	33	32.5	98	0.09	0	S	1.8
11.	33	40	33	32	95	0.75	0	N	1
12.	33.5	35.5	33	32.5	98	0.22	0	SW	1.8
13.	33	40	32.5	32	95	0.75	0	N	1
14.	33	40	32.5	32	95	0.57	0	N	1
15.	22.5	35.5	32	31.5	98	0.18	0	E	6.2
16.	33.5	35	32.5	32	95	1.18	0	NW	1.8
17.	33	35.5	32.5	32	95	0.94	0	N	2.6
18.	33	40	32.5	32	95	0.75	0	N	1
19.	33	40	32.5	32	95	0.75	0	N	1
20.	33.5	36	33	32.5	98	0.24	0	E	0
21.	33	35.5	33	32	95	0.94	0	N	2.6
22.	33.5	36.5	33	32.5	98	0.41	0	SW	1.2
23.	34	36	33.5	33	96	0.33	0	E	3
24.	34.5	37	33.5	33	96	0.26	0	N	1
25.	34.5	35	24	33.5	98	1.1	0	SW	8
26.	34	35.5	33.5	33	96	2.3	0	E	6.2
27.	34	35.5	24	33	98	0.19	0	E	1.2
28.	33	35.5	32.5	32	95	0.94	0	N	2.6
29.	34	35.5	33.5	33	96	2.3	0	E	6.2
30.	33.5	37	33	32.5	95	0.13	0	S	1
31.	33.5	37	33	32.5	95	0.13	0	S	1

Constraint	KNN Algorithm
Accuracy	96.77419354838710%
Time (milli secs)	176
Correctly Classified	96.77419354838710%
InCorrectly Classified	3.225806451612900%

UTTARA KANNADA AUGUST MONTH RAINFALL PREDICTION

RAINFALL PREDICTION (KNN ALGORITHM)!!!

SlNo	mintemp	maxtemp	drybulbtemp	wetbulbtemp	humidity	windspeed	avewindspeed	winddirection	Result
1.	24	27.5	23.5	23	96	0.29	0	SW	0
2.	24	29	23.5	23	91	0.47	0	SW	0
3.	23	28	23	22	91	0.34	0	E	13.2
4.	24	27	23	22.5	91	1.14	0	S	13.4
5.	23	26	23.5	23	96	1.37	0	S	19.2
6.	23	25.5	23	22.5	91	0.38	0	SW	61
7.	22.5	25.5	23	22	91	0.45	0	N	61
8.	22.5	25	23	22.5	91	3.79	0	S	12.4
9.	22.5	25	22.5	23	95	0.92	0	SW	63
10.	23	24.5	23	22	91	0.86	0	N	67
11.	23.5	25	22.5	22	95	0.61	0	E	13.8
12.	23.5	25.5	23	22.5	91	1.55	0	S	10.2
13.	23	26	22.5	22	95	1.09	0	SW	15.8
14.	23.5	25.5	23	22.5	91	0.27	0	S	0
15.	23.5	26	23	22.5	91	0.87	0	N	2
16.	23.5	26	23	22.5	91	0.81	0	N	2
17.	23.5	25.5	23	22.5	95	0.39	0	SW	0
18.	24	26	23	22.5	91	0.33	0	E	0
19.	24	26	23.5	23	96	0.19	0	W	13.6
20.	24	25.5	23.5	23	96	0.09	0	S	0
21.	24	26	24	23.5	91	0.54	0	E	0
22.	24.5	27	23.5	23	96	1.84	0	SW	0
23.	24	27.5	23.5	23	96	0.68	0	N	0
24.	24.5	27	24	23.5	91	0.39	0	SW	0
25.	25	27	24	24	96	0.25	0	SW	6.2
26.	25	27.5	24.5	23	87	0.23	0	SW	0
27.	23.5	27.5	23.5	23	96	0.11	0	N	12.4
28.	24	28	24	23.5	91	0.07	0	S	0
29.	23.5	28.5	23.5	23	96	0.33	0	SW	0
30.	23.5	29	23	22.5	95	0.23	0	N	0
31.	24	29.5	24	23	91	0.21	0	S	0

Constraint	KNN Algorithm
Accuracy	93.54838709677420%
Time (milli secs)	155
Correctly Classified	93.54838709677420%
InCorrectly Classified	6.451612903225810%

6. CONCLUSION:

The proposed system is giving very high accurate results for rainfall prediction of various regions of Karnataka as we can see above. And we here by conclude that, this early prediction of rainfall is helpful to farmers to take important decisions on which crop to grow, how much quantity of fertilizers or other chemicals are needed for the proper cultivation of crops. Rainfall prediction will also help to avoid damages caused by natural disasters like flood etc.

REFERENCES:

1. Choujun Zhan1, Fujian Wu1, Zhengdong Wu1, Chi K. Tse, 2019: “Daily Rainfall Data Construction and Application to Weather Prediction”.
2. Chandrasegar Thirumalai, M Lakshmi Deepak, K Sri Harsha, K Chaitanya Krishna, 2017, “Heuristic Prediction of Rainfall Using Machine Learning Techniques”.
3. Tharun V.P, Ramya Prakash, S. Renuga Devi, 2018, “Prediction of Rainfall Using Data Mining Techniques”.
4. Nikhilkumar B. Shardoor, Mandapati Venkateswar Rao**,** 2018, “Analysis of Rainfall Prediction Using Machine Learning Data Mining and Satellite Techniques”.
5. CT Cheng, 2015, “Heuristic Methods for Reservoir Monthly Inflow Forecasting”.
6. J.Refonaa, M. Lakshmi, Raza Abbas, Mohammad
7. Raziullha, 2019, “Rainfall Prediction using Regression Model”.
8. Goutami Bandyopadhyay, 2011, “The Prediction of Indian Monsoon Rainfall: A Regression Approach”.
9. Kar, Kaveri, Neelima Thakur, and Prerika Sanghvi 2019, "Prediction of Rainfall Using Fuzzy Dataset".