

Rainfall Prediction Using Machine Learning Techniques

Dr. Mercy Paul Selvan¹, Tanneru Navanish², Kancharla Narayan Pavan³, Dr. Viji Amutha Mary⁴, Dr. S. Jancy⁵, Dr. Joshila Grace⁶, Dr. Suji Helen⁷

^{2,3}Student, ^{1,4,6,7}Associate Professor, Sathyabama Institute of Science and Technology, Chennai, India

⁵Associate Professor, Sathyabama Institute of Science and Technology, Chennai, India

Article Info

Page Number: 3553-3562

Publication Issue:

Vol. 71 No. 4 (2022)

Article History

Article Received: 25 March 2022

Revised: 30 April 2022

Accepted: 15 June 2022

Publication: 19 August 2022

Abstract

India is a farming nation and its economy is to a great extent dependent on rainforest creation. Downpour estimates are vital and fundamental for all ranchers to examine crop yields. Unsurprising rainfall is the capacity to foresee the climate with the assistance of science and innovation. It is essential to know how much rainfall to utilize water assets, horticultural creation and water arranging proficiently. Various strategies for information mining can foresee rainfall. Information extraction is utilized to appraise rainfall. This article features probably the most well-known rainfall forecast calculations. Logistic Regression, KNeighborsClassifier, Random Forest Classifier and Certificate Tree are a portion of the calculations contrasted with this record. According to a relative perspective, it is feasible to break down how rainfall is accurately anticipated.

Keywords: Rainfall Prediction, model evaluation, oversampling, under sampling, Logistic Regression, KNeighborsClassifier, Random Forest Classifier.

INTRODUCTION

Rainfall arranging is perhaps the most troublesome task. Albeit numerous calculations have been set up, it is truly challenging to precisely anticipate rainfall. In a nation like India, the attention is on crop achievement, disappointment, and water deficiencies at whatever year. A slight change in the stormy season will contrarily affect horticulture. Appropriate rainfall arranging is fundamental to forestall debacle. Legitimate climate anticipating sometimes, like floods and dry seasons, can assist with overseeing farming and forestall fiascos. This article investigates various calculations. Information mining procedures are all around used to recognize rainfall.

LITERATURE REVIEW

[1] Measurable investigation shows the idea of ISMR, which can't be precisely anticipated by insights or factual information. Hence, this review exhibits the utilization of three techniques: object creation, entropy, and artificial neural network (ANN). In view of this innovation, another technique for anticipating ISMR times has been created to address the idea of ISMR. This model has been endorsed and supported by the studio and exploration data. Factual examination of different information and near investigations showing the presentation of the normal technique The primary impact of this movement is

to exhibit the advantages of AI calculations, just as the more prominent degree of clever framework than the advanced rainfall determining methods. We analyze and think about the momentum execution (Markov chain stretched out by rainfall research) with the forecasts of the six most notable AI machines: Genetic programming, Vector relapse support, radio organizations, M5 organizations, M5 models, models - Happy. To work with a more itemized appraisal, we led a rainfall overview utilizing information from 42 metropolitan urban communities.

[2] RF was utilized to anticipate assuming that it would rain in one day, while SVM was utilized to foresee downpour on a blustery day. The limit of the Hybrid model was fortified by the decrease of day-by-day rainfall in three spots at the rainfall level in the eastern piece of Malaysia. Crossover models have likewise been found to emulate the full change, the quantity of days straight, 95% of the month-to-month rainfall, and the dispersion of the noticed rainfall.

[3] In India, farming is the backbone. Downpour is a significant plant. These days, climate is a major issue. Climate gauging gives data on rainfall estimating and crop security. Numerous strategies have been created to recognize rainfall. Machine Learning calculations are significant in foreseeing rainfall.

The reason for the framework is to anticipate the climate sooner or later. Climatic still up in the air utilizing various sorts of factors all over the place. [5] Of these, main the main highlights are utilized in climate conjectures. Picking something like this relies a great deal upon the time you pick. Underlying displaying is utilized to incorporate the fate of demonstrating, AI applications, data trade, and character examination.

[6] Contrasted with different spots where rainfall information isn't accessible, it consumes a large chunk of the day to build up a solid water overview for a long time. Improving complex neural organizations is intended to be a brilliant instrument for anticipating the stormy season. This downpour succession was affirmed utilizing a complex perceptron neural organization. Estimations like MSE (Early Modeling), NMSE (Usually Early Error), and the arrangement of informational collections for transient arranging are clear in the examination of different organizations, like Adanaive. AdaSVM.

[7] In this paper, Artificial Neural Network (ANN) innovation is utilized to foster a climate anticipating strategy to distinguish rainfall utilizing Indian rainfall information. Along these lines, Feed Forward Neural Network (FFNN) was utilized utilizing the Backpropagation Algorithm. Execution of the two models is assessed dependent on emphasis examination, Mean Square Error (MSE) and Magnitude of Relative Error (MRE). This report likewise gives a future manual for rainfall determining.

[8] This page features rainfall investigation speculations utilizing Machine Learning. The principle motivation behind utilizing this program is to secure against the impacts of floods. This program can be utilized by conventional residents or the public authority to anticipate what will occur before the flood. The flood card, then, at that point, furnish them with the vital help by moving versatile or other important measures.

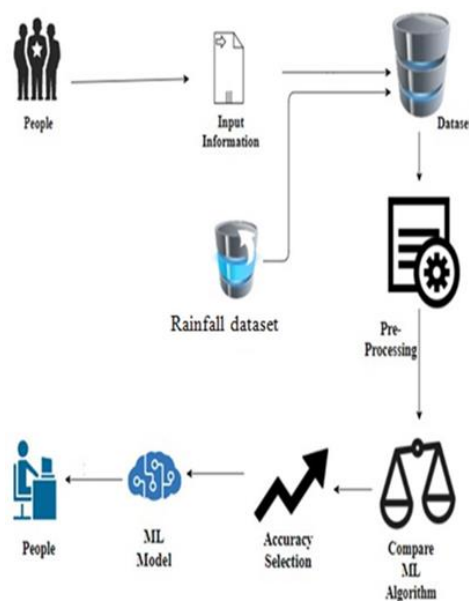
EXISTING SYSTEM

Agribusiness and the Economic Power of India. Ranchers possibly plant when its downpours. To get a decent collect, you really want great soil, composts and a decent environment. The climate is vital for each rancher. Unexpected changes in the climate are harming the populace monetarily and genuinely. The climate is one of the most troublesome issues today. The principle motivation behind this climate gauge page is to utilize various techniques for information mining. Isolating, combining, getting trees and nets. Meteorological data is additionally called meteorological data. The most generally utilized boundaries in this article arerainfall, wind speed, temperature, and temperature.

PROPOSED SYSTEM

Rainfall is significant for food conveyance, water assets the board, and every single ecological action. Long haul dry spells or substantial rainfall during extreme development and advancement can altogether diminish yields. India is a rural nation and its economy is to a great extent dependent on farming creation. In this way, rainfall determining is turning out to be an ever-increasing number of significant in farming nations like India. Rainfall gauging has become one of the world's most squeezing science and innovation issues in ongoinghundreds of years.

SYSTEM ARCHITECTURE

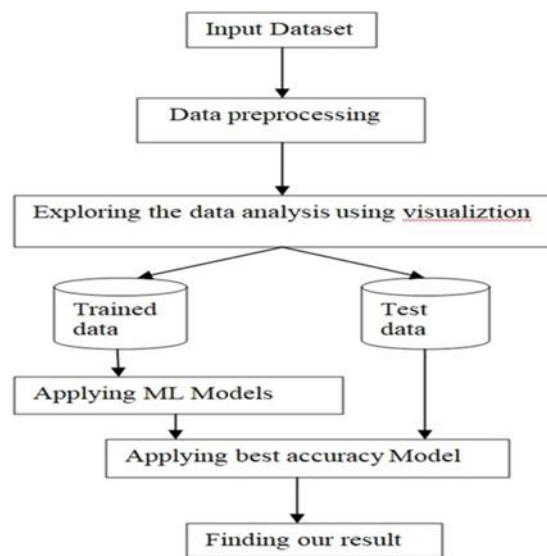


IMPLEMENTATION

This proposed method is been developed using various AI strategies like Random Forest, LogisticRegression, and, KNeighborsClassifier to predict and recommend for predication of rainfall.

In Dataset the key attributes we are using as follow: Maximum Temperature, Minimum Temperature,Wind Speed, Pressure etc.,

WorkFlow Diagram



The above are the module we are doing for preprocessing our dataset:

MODULES

Data Collection
 Data Cleaning
 Data Selection

Data Transformation
 Data Mining Stage

DataSet

MinTemp	MaxTemp	Rainfall	Evaporatic	Sunshine	WindGust1	WindGust2	WindDir1s	WindDir3p	WindSpeed	WindSpeed	Humidity1	Humidity3	Pressure1	Pressure3	Cloud1am	Cloud3pm	Temp1am	Temp3pm	RainToday	RainTomorrow
13.4	22.9	0.6	NA	NA	W	44	W	WNW	20	24	71	22	1007.7	1007.1	8	NA	16.9	21.8	No	No
7.4	25.1	0	NA	NA	WNW	44	NNW	WSW	4	22	44	25	1010.6	1007.8	NA	NA	17.2	24.3	No	No
12.9	25.7	0	NA	NA	WSW	46	W	WSW	19	26	38	30	1007.6	1008.7	NA	2	21	23.2	No	No
9.2	28	0	NA	NA	NE	24	SE	E	11	9	45	16	1017.6	1012.8	NA	NA	18.1	26.5	No	No
17.5	32.3	1	NA	NA	W	41	ENE	NW	7	20	82	33	1010.8	1006	7	8	17.8	29.7	No	No
14.6	29.7	0.2	NA	NA	WNW	56	W	W	19	24	55	23	1009.2	1005.4	NA	NA	20.6	28.9	No	No
14.3	25	0	NA	NA	W	50	SW	W	20	24	49	19	1009.6	1008.2	1	NA	18.1	24.6	No	No
7.7	26.7	0	NA	NA	W	35	SSE	W	6	17	48	19	1013.4	1010.1	NA	NA	16.3	25.5	No	No
9.7	31.9	0	NA	NA	NNW	80	SE	NW	7	28	42	9	1008.9	1003.6	NA	NA	18.3	30.2	No	Yes

Data Collection

The data utilized in this action was gathered by the meteorological association. The conversation covers the period from past years. This segment of the review talks about the accompanying systems: Cleaning the dad, picking the dad, changing the dad, cleaning the dad.

Data Cleaning

The primary components of coordinated media in this class are the quest for missing data, the quest for bogus data, and the obliteration of weeds. At last, the information cleaning

framework was changed into an effective information mining framework.

Data Selection

At this stage, the data identified with the examination tree is the choice tree and is extricated from the informational index. The meteorological informational index had ten attributes that pre-owned two elements later on. Because of the idea of cloud information, all qualities are the equivalent, and the vast majority of the qualities that are absent in sunlight-based information are not utilized in the investigation.

Data Transformation

It basically came to our notification then, at that point. This is the phase of changing over the chose data into an effective information mining framework. The information document is put away as a Comma Separated Value (CSV) record, and the information is standardized to limit the effect of the information estimation.

Data Mining Stage

The information mining stage is separated into three sections. All calculations were utilized to investigate the information at each stage. The exploratory strategy utilized in this review is separated into the level of trains in the informational collection, their cross-approval, and the excess rate.

Summarize the Data Visualization :

The data set consists of various attributes like Temperature, Wind Speed, Pressure etc., Depending up on these attributes the following are differ:

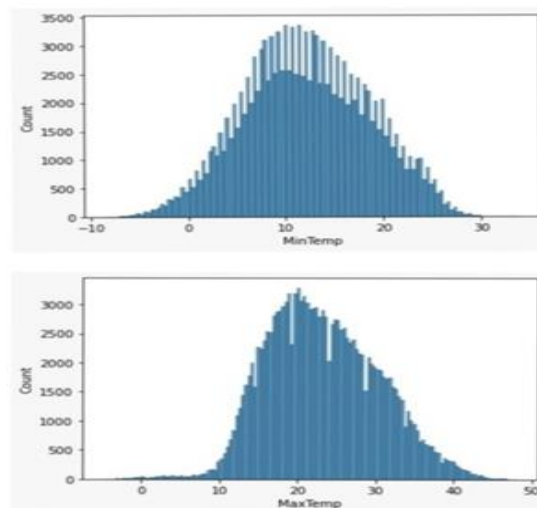


Fig1: Relation on Count of the Temperature For Rainfall

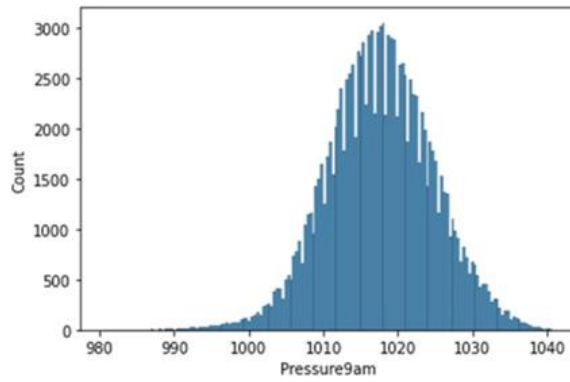


Fig2: Relation on Count of the Pressure for Rainfall

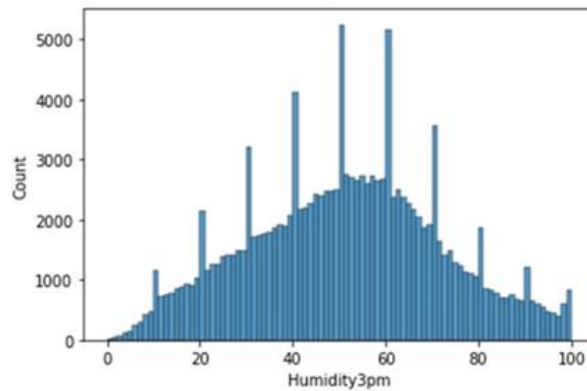


Fig3: Relation on Count of the Humidity for Rainfall

Figure 1,2,3 Tells the relationship Among the rainfall and their Count For the Given DataSet and below Figure 4 is the Heat map.

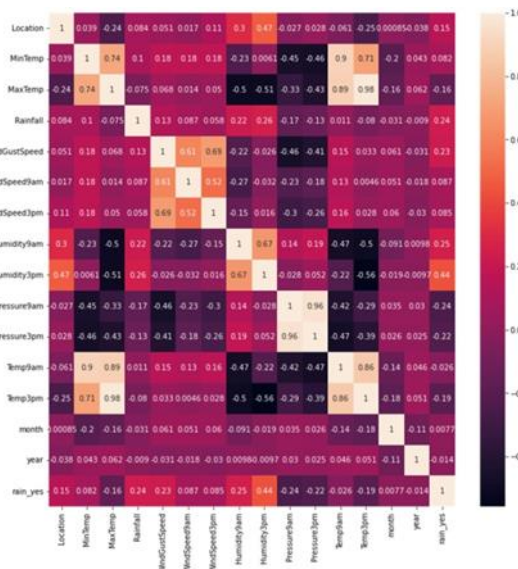


Fig4: Correlation between all the features

Machine learning techniques used:

Because more than one class may be given to a single instance, multi-label classification

(MLC) is the best and optimum option. The MLC is embedded into the algorithms used for categorization and model building, which are as follows:

1. Logistic Regression
2. Random Forest
3. KNeighborsClassifier Logistic regression:

This is a AI strategy used to predict and describe the connection between independent and dependent data values. It predicts the dependent variable by analyzing the relation between the other independent variables. Here is the Accuracy and F1_Score the particular machine learning algorithm.

Accuracy	F1_Score
0.83672635411	0.49511774842

Random Forest:

This is a AI strategy used to solve problems like regression and classification. The algorithm gives an outcome that is based on the

predictions of the decision trees by taking the average or mean of the output from different trees. It can be used to predict what will happen in the future. Here is the Accuracy and F1_Score for the particular machine learning algorithm.

Accuracy	F1_Score
0.85578423359	0.60700546669

KNeighborsClassifier:

This is a AI strategy used to solve problems like regression and classification. The algorithm takes nearest values around the unknown variable. Here is the Accuracy and F1_Score for the particular machine learning algorithm.

Accuracy	F1_Score
0.8375808009	0.5655802861

RESULT

With the created Random Forest model we acquired an accuracy of 85.57% with a voting classifier where we created an array called Random forest. which consists of our models and we made the voting parameter set to hard, which tells that our model to make the

predictions by the highest votes now, we can implement Random Forest model and we predicted the accuracy score with our test data and here is the individual accuracy of the algorithms.

Algorithm	Accuracy
LR	83.67%
RF	85.57%
KNN	83.75%

CONCLUSION

This record tells the best way to concentrate on how much rainfall checked utilizing a machine preparing machine to decrease rainfall data. We utilized different calculations to decide the normal rainfall. We looked at the logistic regression, Random Forest, and KNeighborsClassifier classifications. From above table, we can presume that normal timberland is a reasonable AI calculation for anticipating rainfall in India.

At present, AI is utilized for. industry. As howmuch data builds, how much data increments, so we use machines to more readily comprehend the data. The climate conjecture assists a ton with getting a decent score, and the rainfall gives a decent figure. Later on, we intend to build the workon harvest and yield estimating and downpour gauging.

REFERENCES

- [1] Singh, P., 2018. Indian summer monsoon rainfall (ISMR) forecasting using time series data: A fuzzy-entropy-neuro based expert system. *Geoscience Frontiers*, 9(4), pp.1243-1257.
- [2] Cramer, S., Kampouridis, M., Freitas, A. and Alexandridis, A., 2017. An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Systems with Applications*, 85, pp.169-181.
- [3] Pour, S., Shahid, S. and Chung, E., 2016. A Hybrid Model for Statistical Downscaling of Daily Rainfall. *Procedia Engineering*, 154, pp.1424-1430.
- [4] manjunath n, muralidhar b r, sachin kumar s, vamshi k and savitha p, 2021. rainfall prediction using machine learning and deep learning techniques. [online] *irjet.net*. available at: <<https://www.irjet.net/archives/v8/i8/irjet-v8i850.pdf>> [accessed 20 january 2022].
- [5] Tanvi Patil and Dr. Kamal Shah, 2021. Weather Forecasting Analysis using Linear and Logistic Regression Algorithm. [online] *Irjet.net*. Available at: <<https://www.irjet.net/archives/V8/i6/IRJET-V8I6454.pdf>> [Accessed 20 January 2022].
- [6] n. Divya prabha and p. Radha, 2019. Prediction of weather and rainfall forecasting using classification techniques. [online] *irjet.net*. Available at:

<<https://www.irjet.net/archives/v6/i2/irjet-v6i2154.pdf>> [accessed 20 January 2022].

- [7] Waghmare, D., 2021. Machine Learning Technique for Rainfall Prediction. *International Journal for Research in Applied Science and Engineering Technology*, 9(VI), pp.594-600.
- [8] YashasAthreya, VaishaliBV, SagarK and SrinidhiHR, 2021. Flood Prediction and Rainfall Analysis using Machine Learning. [online] Irjet.net. Available at: <<https://www.irjet.net/archives/V8/i7/IRJET-V8I7432.pdf>> [Accessed 20 January 2022].
- [9] Noone, David, and Harvey Stern. "Verification of rainfall forecasts from the Australian Bureau of Meteorology's Global Assimilation and Prognosis(GASP) system." *Australian Meteorological Magazine* 44.4 (1995): 275-286.
- [10] Hornik, Kurt, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators." *Neural networks* 2.5 (1989): 359-366.
- [11] Haykin, Simon. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.
- [12] Rajeevan, M., Pulak Guhathakurta, and V. Thapliyal. "New models for long range forecasts of summer monsoon rainfall over North West and Peninsular India." *Meteorology and Atmospheric Physics* 73.3-4 (2000): 211-225.
- [13] Xiong, Lihua, and Kieran M. OConnor. "An empirical method to improve the prediction limits of the GLUE methodology in rainfallrunoff modeling." *Journal of Hydrology* 349.1-2 (2008): 115-124.
- [14] Schmitz, G. H., and J. Cullmann. "PAI-OFF: A new proposal for online flood forecasting in flashflood prone catchments." *Journal of hydrology* 360.1-4 (2008): 1-14.
- [15] Riordan, Denis, and Bjarne K. Hansen. "A fuzzy casebased system for weather prediction." *Engineering Intelligent Systems for Electrical Engineering and Communications* 10.3 (2002): 139-146.
- [16] Guhathakurta, P. "Long-range monsoon rainfall prediction of 2005 for the districts and sub- division Kerala with artificial neural network." *Current Science* 90.6 (2006): 773-779.
- [17] Pilgrim, D. H., T. G. Chapman, and D. G. Doran. "Problems of rainfall-runoff modelling in arid and semiarid regions." *Hydrological Sciences Journal* 33.4 (1988): 379-400.
- [18] Lee, Sunyoung, Sungzoon Cho, and Patrick M. Wong. "Rainfall prediction using artificial neural networks." *journal of geographic information and Decision Analysis* 2.2 (1998): 233-242
- [19] Charaniya, Nizar Ali, and Sanjay V. Dudul. "Committee of artificial neural networks for monthly rainfall prediction using wavelet transform." *Business, Engineering and Industrial Applications (ICBEIA), 2011 International Conference on. IEEE, 2011*

- [20] French, Mark N., Witold F. Krajewski, and Robert R. Cuykendall. "Rainfall forecasting in space and time using a neural network." *Journal of hydrology* 137.1-4 (1992): 1-31.