AI based model for prediction of heavy/high impact rain events

Aviral Srivastava[1], Aruneema Joshi[2], Brahmansh Dwivedi[3], Neetu Kumari Rajput[4]

[1]*Dept. of Information Technology, Noida Institute of Engineering and Technology. Greater Noida, Uttar Pradesh,*
0201ite040@niet.co.in

[2]*Dept. of Information Technology, Noida Institute of Engineering and Technology. Greater Noida, Uttar Pradesh,*
0201ite129@niet.co.in

[3]*Dept. of Information Technology, Noida Institute of Engineering and Technology. Greater Noida, Uttar Pradesh,*
0201ite041@niet.co.in

[4]*Assistant Professor Dept. of Information Technology, Noida Institute of Engineering and Technology. Greater Noida, Uttar Pradesh, neetu.rajput@niet.co.in*

---***---

Abstract - India is a farming country whose economy is heavily reliant on the development of rainforests. In order to analyse agricultural yields, rainfall estimates are essential and fundamental for all ranchers. The ability to predict the climate with the aid of science and innovation is unsurprising rainfall. To effectively use water resources, horticulture production, and water arranging, it is important to determine the amount of rainfall. Rainfall can be predicted using a variety of information mining techniques. The process of information extraction is used to evaluate rainfall. Probably the most well-known rainfall forecast calculations may be found in this article. Some of the computations compared with this record include Logistic Regression, K-Neighbors Classifier, Random Forest Classifier, and Certificate Tree. It is possible to decompose the process of precisely predicting rainfall from a relative standpoint. To do this, we gathered satellite data. We have taken a number of variables into account in order to predict (temperature, dew point, humidity, wind pressure, wind speed, wind direction, etc.).

Key Words: Random Forest Classifier, KNeighbors Classifier, Logistic Regression, Oversampling, Under sampling, and Rainfall Prediction.

1.INTRODUCTION

Planning for rainfall is arguably the most difficult endeavor. Even with all the computations in place, it is quite difficult to predict rainfall with precision. In a country such as India, the focus is on crop performance, crop failure, and water scarcity in each given year. Conversely, horticulture will be negatively impacted by even a small shift in the stormy season. Planning for the right amount of rainfall is essential to preventing disasters. Accurately anticipating certain climate events, such as droughts and floods, can help manage agriculture and prevent disasters. This paper explores many kinds of computations. Rainfall detection is accomplished everywhere by means of information mining techniques.

The purpose of this study is to investigate the creation and use of an artificial intelligence (AI) model designed especially for forecasting periods of heavy or highly significant rainfall events. The suggested model will analyse enormous volumes of atmospheric and environmental data and produce probabilistic forecasts of the occurrence of heavy rain by utilising sophisticated machine learning algorithms, such as deep learning neural networks, ensemble approaches.

2. Literature Review

A. Indian Summer Monsoon Rainfall (ISMR) Forecasting using Time Series Data: A Fuzzy-Entropy-Neuro based Expert System Pritpal Singh/30 July 2018

Measurable research demonstrates the concept of ISMR, which cannot be accurately predicted by facts or insights. As a result, the three approaches used in this review are object creation, entropy, and artificial neural networks (ANN). A new method has been developed to address the concept of ISMR in light of this innovation for predicting ISMR times. The studio and exploration data have validated and supported this approach. Factual analysis of various data and close examinations demonstrating the standard procedure's presentation.

B. An Extensive Evaluation of Seven Machine Learning Methods for Rainfall Prediction in Weather Derivatives
Sam Cramera, Michael Kampouridis a Alex A. Freitas a Antonis K. Alexandridis b/2017
This movement's main effect is to demonstrate the benefits of Al computations as well as the greater degree of intelligent framework compared to sophisticated techniques for calculating rainfall. The momentum execution (Markov chain extended by rainfall study) is analyzed and considered in relation to the predictions of the six most prominent Al machines using genetic programming. Radio organizations, M5 organizations, MS models, happy models, vector relapse support. In order to work with a more detailed assessment, we conducted a rainfall overview using data from 42 large urban areas.

C. P A Hybrid Model Downscaling Sahar Hadi of Poura for Daily Statistical Rainfall Shamsuddin Shahida, Eun-Sung Chungb/2019

While SVM was used to predict precipitation on a windy day, RF was used to predict the assumption that it would rain one day. The decrease in daily rainfall in three locations at the rainfall level in the eastern part of Malaysia strengthened the boundaries of the Hybrid model. Additionally, it has been discovered that crossover models can replicate the entire variation, the number of consecutive days, 95% of the monthly rainfall, and the distribution of the observed rainfall.

D. RAINFALL PREDICTION USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES: REVIEW, 2021

India's primary industry is farming. Rainfall is an important type of plant. These days, a big concern is the climate. Climate gauging provides information on crop security and rainfall estimation. Many techniques have been developed to detect rainfall. Calculations using machine learning are important for predicting rainfall.

3. EXISTING SYSTEM

India's Economic Power and Agribusiness. Ranchers may plant during periods of precipitation. You actually need excellent soil, composts, and a nice habitat to have a decent collection. Every rancher needs to be aware of the climate. The population is suffering real and financial hardship as a result of unexpected changes in the climate. One of the most problematic topics of the day is the climate. The main goal of this climate gauge page is to use different information mining techniques. separating, merging, obtaining nets and trees. Another name for meteorological data is meteorological data. In this article, the boundaries for temperature, wind speed, rainfall, and temperature are most frequently used.

4. PROPOSED SYSTEM

Rainfall is essential for the transportation of food, the general management of water resources, and all biological processes. Extended periods of drought or heavy precipitation during periods of rapid development and progress might completely reduce harvests. India is a rural country whose economy depends heavily on agricultural production. In this sense, rainfall determination is becoming more and more important in agricultural countries like India. Over the course of several centuries, rainfall measurement has emerged as one of the most challenging scientific and innovative problems in the world.

## 5. SYSTEM ARCHITECTURE
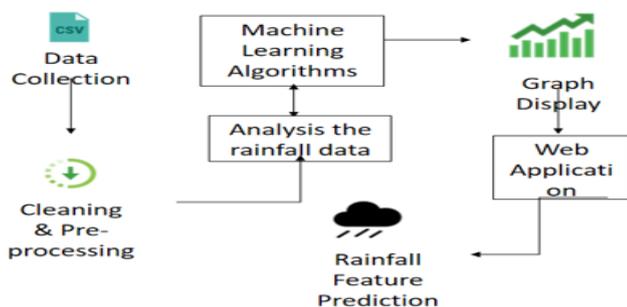
6. IMPLEMENTATION

This suggested approach was created to forecast and suggest rainfall utilizing a variety of AI techniques, including Random Forest, Logistic Regression, and KNeighbors Classifier.

The following are the primary attributes we are using in the dataset: Pressure, wind speed, minimum and maximum temperatures, and so forth.

7. WORKFLOW
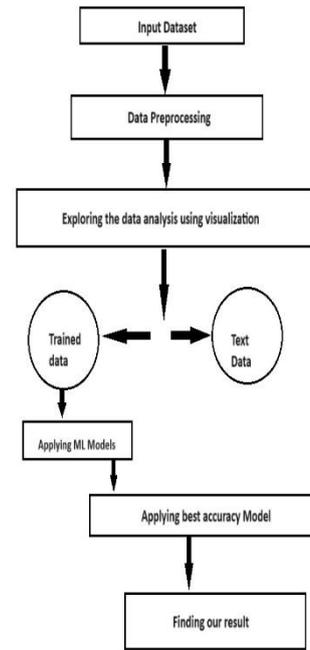
The above are the module we are doing for preprocessing our dataset:
1. Data collection
2. Data Cleaning
3. Data Selection
4. Data Transformation
5. Data Mining stage

DATA COLLECTION

The meteorological association collected the data that was used in this action. The discussion encompasses the years 2012–2015. This section of the review discusses the supporting systems: data cleaning, data selection, data alteration, and data cleaning.

DATA CLEANING

In this class, the search for fake data, the eradication of weeds, and the search for missing data are the main elements of coordinated media. Finally, an efficient information mining framework was created by modifying the information cleansing framework.

DATA SELECTION

Currently, the choice tree is the data identified by the examination tree, and it has been extracted from the informational index. There were 10 tributes in the meteorological informative index that were afterward owned by two elements. All attributes are similar due to the concept of cloud information, and the great majority of qualities missing from information based on sunlight are not used in the study.

DATA TRANSFORMATION

At that moment, it essentially reached our notice. This stage involves converting the selected data into a useful framework for information mining. To reduce the impact of the

information estimation, the information is standardized and stored in a Commas Separated Value (CVS) record.

DATA MINING STAGE

Three sections make up the information mining stage. Every computation was used to examine the data at every level. The degree of trains in the informative collection, their cross-approval, and the excess rate are the three distinct components of the exploratory technique used in this review.
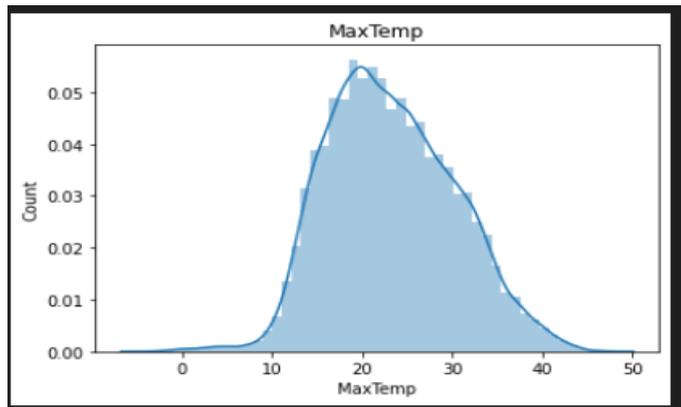
8. SUMMARIZE THE DATA VISUALIZATION

The data set includes a variety of attributes, such as pressure, temperature, wind speed, and others. The following vary depending on these attributes:
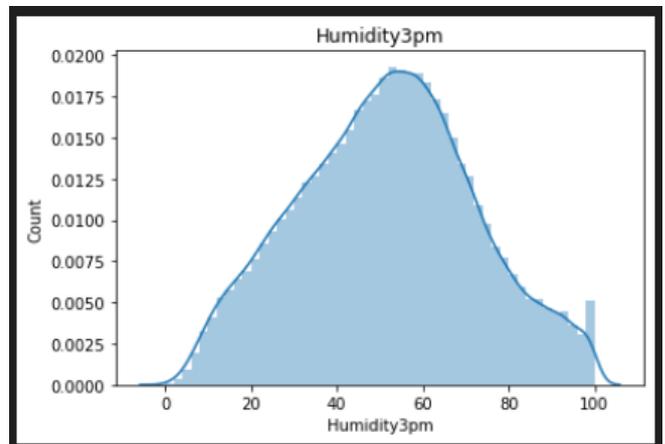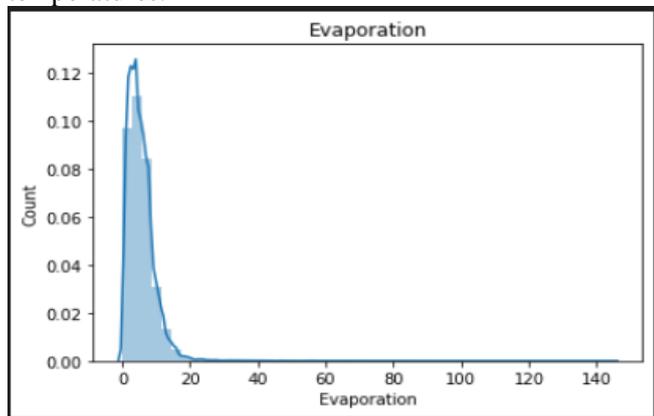

Relation on count of minimum and maximum temperatures.

Relation on count of evaporation.
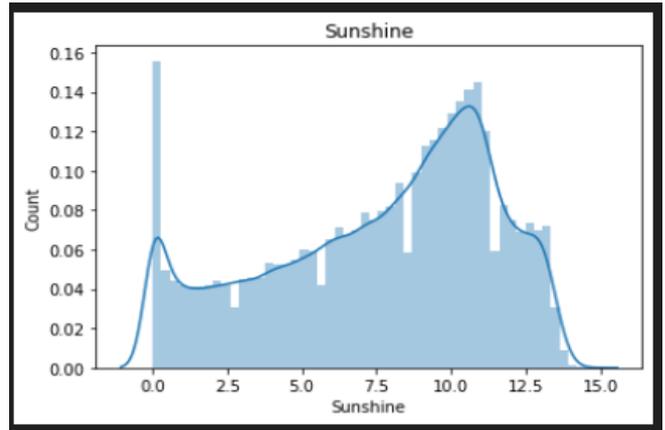
Relation on count of Sunshine.
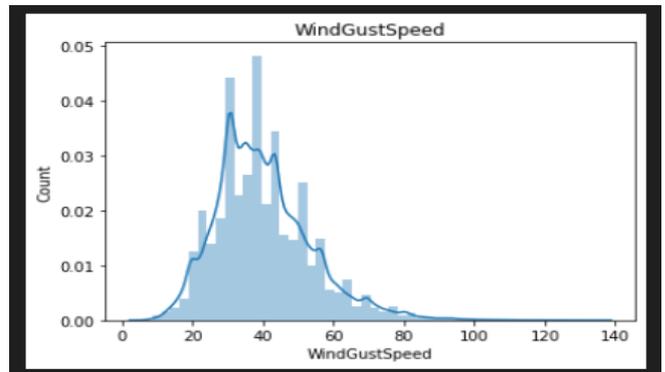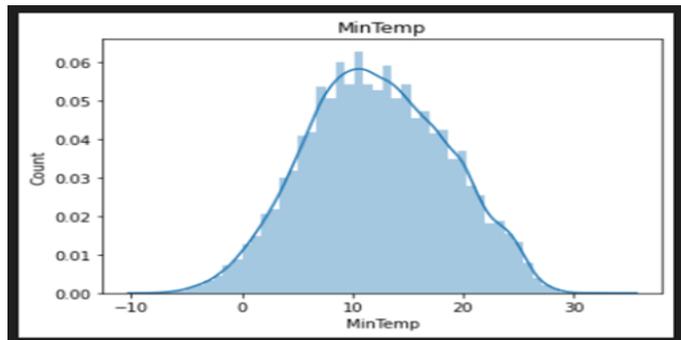
Relation on count of Wind Gust Speed.

Relation on count of Humidity.
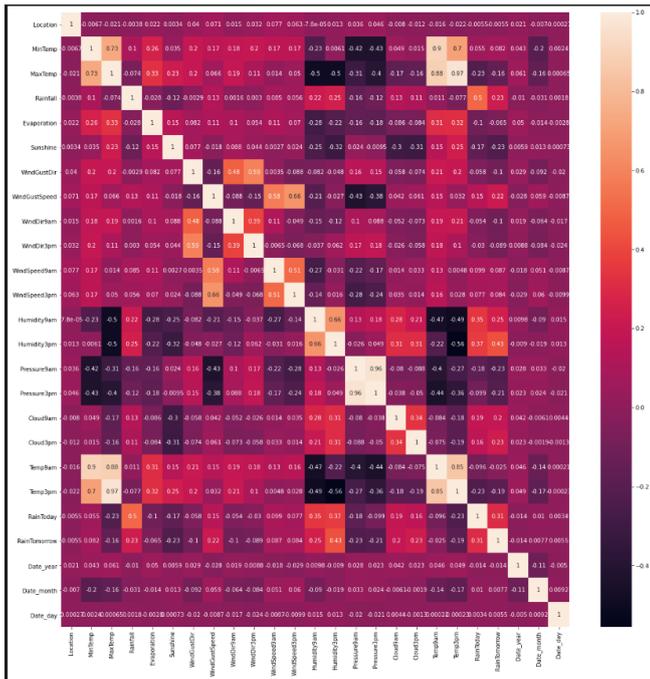
The figure shows the relation among the rainfall and there count for the given dataset and it is heat map.

9. Methods of machine learning employed:

Multi-level classification (MLC) is the ideal choice since a single instance may be assigned to multiple classes. The following algorithms are used model development and categorization, and they incorporate the MLC:

1. Logistic regression
2. The random forest
3. K neighbour classification

Logistic regression:

The relationship between the independent and dependent data values is predicted and described using a AI approach. Buy examining the relationship between the other independent variables 8 forecast the dependent variable. Here are the specific machine learning algorithms F1 score and accuracy.

Accuracy	F1_SCORE
0.83672635411	0.49511774842

Random Forest:

This Algorithm is employed to address issues related to regression and clarity. By averaging or meaning the output from several trees, the algorithm produces a result based on the decision trees' predictions. It can be applied to forecast future events.

For the specific machine learning algorithm, the Accuracy and F1-Score are shown here.

Accuracy	F1_SCORE
0.85578423359	0.60700546669

K-Neighbors Classifier:

This Algorithm is applied to address classification and regression problems. The algorithm uses the values that are closest to the unknown variable. Here are the specific machine learning algorithm's accuracy and F1 score.

Accuracy	F1_Score
0.8375808009	0.5655802861

10. RESULT

We developed an array called Random Forest, which consists of our models, and we set the voting parameter to hard, which indicates that our model should make predictions based on the highest votes. With this setup, we were able to implement the Random Forest model and predict the accuracy score using our test data. The individual accuracy of each algorithm is shown here. With this Random Forest model, we were able to achieve an accuracy of 85.57% with a voting classifier.

Algorithm	Accuracy
LR	83.67%
RF	85.57%
KNN	83.75%

11. CONCLUSIONS

This record indicates how to focus on measuring rainfall using a machine preparation machine in order to reduce rainfall data. Several computations were employed to determine the average precipitation. We examined the classifications using SVM, Good Forest, Navie Bayes, and MLP (Multilayer Perceptron). We can assume that typical timberland is a fair Al computation for predicting rainfall in India .

Al is being used in industries. We employ technology to help us understand the data more easily as it grows and increases in volume. The rainfall provides a decent figure, and the climatic

conjecture greatly aids in obtaining a respectable score. We plan to expand the work on yield and harvest estimation as well as rainfall gauge later.

ACKNOWLEDGEMENT

REFERENCES

[1] 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.

[2] Schmitz, G. H., and J. Cullmann. "PAI-OFF: A new proposal for online flood forecasting in flash flood prone catchments." Journal of hydrology 360.1-4 (2008): 1-14.

[3] 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.

[4] 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.

[5] 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,

[6] 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..

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

[8] 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.

[9] Noone, David, and Harvey Stern. "Verification of rainfall forecasts from the Australian Bureau of Meteorology's Prognosis(GASP) Global Assimilation system." and 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.