RAINFALL PREDICTION BASED ON MACHINE LEARNING

PuttaSrivani1, Sanjeevini Harwalkar2, Arshiya Nooren Fatima3
Asso. Professor, Assistant professor 2,3,
CSE Department1,2,3,
Malla Reddy Engineering College for Women, Maisammaguda, Hyderabad-500100
KCT Engineering College, Gulbarga-585104

Abstract: There are currently no reliable ways to determine if it will rain today. Estimates made by the meteorological agency themselves may turn out to be inaccurate. In this post, we will learn how to create a machine-learning model that can forecast whether or not it will rain today based on various atmospheric variables. Because machine learning models typically outperform human beings at the previously known task—predicting rainfall—this issue is connected to rainfall prediction using machine learning.

1. INTRODUCTION

The project called Rainfall Prediction aims to create machine learning that can predict rainfall using previous weather data. The project's goals, the dataset used, and the model architecture are all explained in this documentation, which also offers a broad summary of the research.

In Indian culture, weather forecasting is significant because it affects people's lives greatly. The difficult task of predicting how often it will rain falls on the meteorological bureau. Accurate forecasting can be challenging when there is enough change. The amount of rain that will fall during the summer and rainy seasons is impossible to predict.

Data Pre-processing: Deal with missing values and fill in or get rid of them from the dataset. Scaling will be applied using normalized features, which will scale the numerical features to a common range. If necessary, it will also preprocess the encoded categorical features to turn the variables into numerical representations.

Feature selection: Examine the relationships between features and choose those that are pertinent to the model.

Model Training: The dataset should be divided into training and test sets. Choose the best machine learning algorithm (such as Random Forest, Gradient Boosting, etc.).

Model Evaluation: Use relevant metrics to assess the performance of the model, such as Mean Absolute Error (MAE) or Room Mean Squared Error (RMSE). Compared to the testing set's actual rainfall values, compare the model's predictions.

Model Deployment: When the model performs satisfactorily, it may be used to create predictions about the amount of rainfall based on the testing set's actual rainfall values.

We can't predict whether it will rain today with any certainty using current approaches. Sometimes even the forecast from the meteorological agency is off.

Since machine learning models frequently outperform human beings at previously recognized tasks that required high levels of expertise, this issue is connected to rainfall prediction using machine learning.

2. Literature survey on rainfall prediction

This chapter conducts a comprehensive analysis of pertinent literature to demonstrate the many approaches authors use to forecast rainfall. The review covers a variety of topics related to the input into, output from, and procedures employed in the many systems developed in the literature for this purpose. In particular, the review focuses on studies that apply supervised learning to classification and regression problems.

Using the keywords "machine learning" OR "deep learning" AND ("precipitation prediction" OR

"rainfall prediction" OR "precipitation nowcasting"), Google Scholar was used to gather works from 2016 to 2020. Only supervised rainfall prediction studies using meteorological data from, for example, radar, satellites, and stations were chosen among the over 1240 results that were produced; however, publications using data from regular cameras were also considered. Examples include the exclusion of images. The techniques employed to achieve this can be expanded upon and used for other geophysical characteristics like temperature and wind, even though the focus of this review is on the prediction of rainfall. The results and arguments of this chapter can therefore be modified to fit different conditions. Except for one article that was published in 2015 and is a seminal work on this subject, 66 publications have been reviewed in total, distributed among conference and journal papers published from 2016 to 2020.

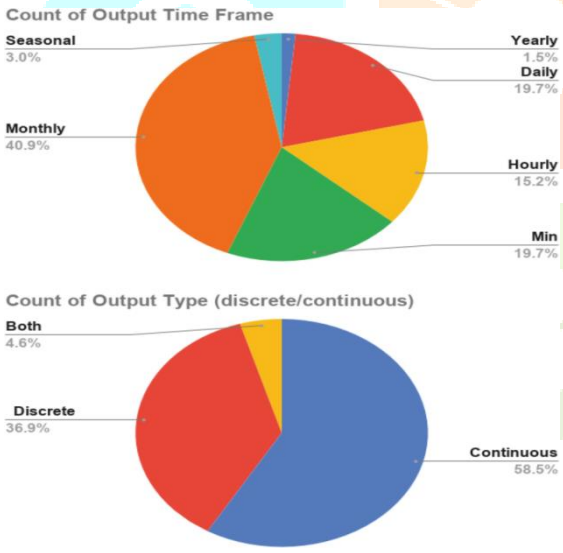


Figure 1

Prediction of rainfall is a subject of study that has a significant impact on many areas of our daily lives. Machine learning has been widely applied to the prediction of rainfall as a result of advancements in computer technology. However, several studies contend that machine learning applications in various disciplines have some shortcomings.

The studies' data source, output goal, input features, the pre-processing, model employed, and results are all examined. The review reveals that many studies have dubious components. Many studies, in particular, lack a baseline predictor to compare results to. Additionally, a lot of references don't give error bars for prediction mistakes, making it impossible to assess the importance of variations in prediction techniques. Furthermore, several references employ techniques that allow data

A few benchmark tests were carried out by Szilard Pafka to evaluate how well XGBoost performed in comparison to various gradient-boosting strategies and bagged decision trees. The classification datasets that XGBoost works on are structured and tabular. depending on its predictive modeling issues, datasets on classification, and regression. The algorithm used to determine contest winners on the Kaggle competitive data science platform is the outcome. Gradient boosting is a method where new models are made to forecast the errors or residuals of earlier models, which are then added together to form the final prediction. It gets its name from the gradient descent process it employs to reduce loss when new models are introduced. Both regressions and classification predictive modeling are supported by this method.

3. DEFINE THE PROBLEM

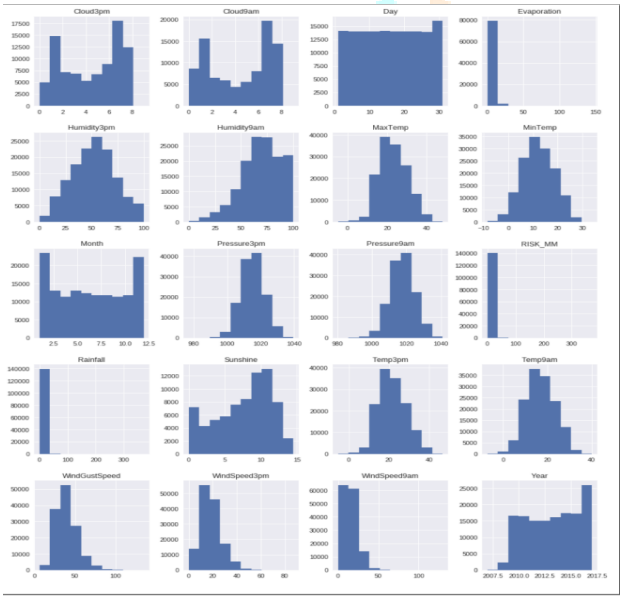
Forecasting rainfall is essential because it can have a variety of effects, including the destruction of crops and farms and property damage. A better forecasting model is needed for an early warning that can lower the risks to life and property and also aids in managing agricultural farms more effectively. Each year, people all around the world are affected by natural disasters like floods and droughts that are brought on by excessive rainfall. To assess rainfall and gauge its likelihood, numerous models have been devised. These models are based on machine learning techniques that are both supervised and unsupervised. We cannot determine whether it rains in particular circumstances by considering the amount of rainfall overall. Most important Numerous models are created to assess rainfall and forecast the likelihood of rain. Both supervised and unsupervised machine learning algorithms constitute the foundation of these models. Considering total rainfall won't enable us to determine if it rains under particular circumstances. The main issue with machine learning is accuracy. To predict whether it will rain given certain conditions, we will first understand the data and train the algorithm accordingly.

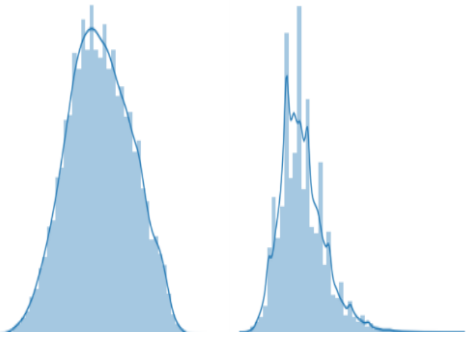
4. MODULES AND FUNCTIONALITIES

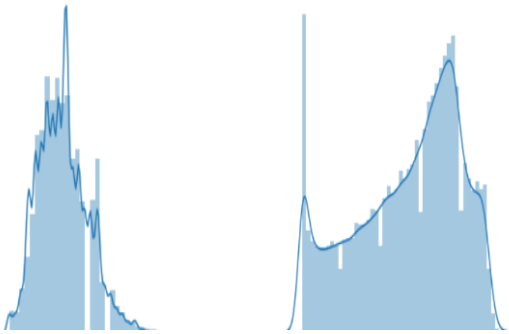
Three local SVR or/and ANN were used in the modular models. Two methods of data preparation, MA and SSA, were investigated using a three-layer feed-forward ANN. Results indicate that the MA outperformed the SSA. Four rainfall records from India, China, Zhongxian, Wuxi, and Zhenwan were chosen as test cases.

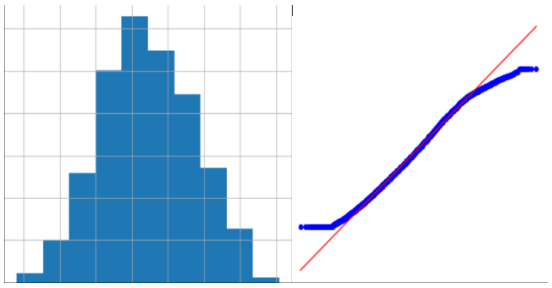
Decision trees, k-nearest neighbors, linear regression, and rule-based approaches are some of the machine-learning techniques used to predict rainfall.

5. RESULTS

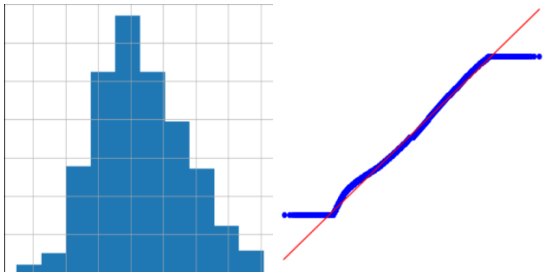








MLRPredictionResult



MLR Accuracy After Prediction

CONCLUSION

The general goal is to define several machine learning (ML) algorithms that can be used to forecast rainfall. By using fewer features and tests, this research seeks to develop reliable and effective models. The data is first pre-processed before being used in the model. The most effective classification algorithms are the Random Forest classifier with about 88% efficiency and K-Nearest Neighbour with 87% efficiency. The Decision Tree classifier, however, has the lowest accuracy (73%). This research can be expanded to include additional ML methods, such as time series, clustering and association rules, and other ensemble methods. Given the limits of this study, more complicated models need to be combined to increase the accuracy of rainfall forecast systems. Study This kind of model can be developed for a large dataset employing more extensive particulate monitoring for a specific area, which will enhance the calculation rate while improving precision and accuracy.

FUTURE SCOPE

The ability to predict rainfall is crucial because it can reduce risks to life and property and help manage agricultural farms more effectively. Heavy or erratic rainfall can have a variety of negative effects, including the destruction of crops and farms and property damage.So, by making early predictions, we can also take precautions.

REFERENCES

[1] ManojitChattopadhyay, Surajit Chattopadhyay, "Elucidating the role of topological pattern discovery and support vector machine in generating predictive models for Indian summer monsoon rainfall", Theoretical and Applied Climatology, pp. 1-12, July 2015, DOI: 10.1007/s00704-015-1544-5.

[2] Kumar Abhishek, Abhay Kumar, Rajeev Ranjan, Sarthak Kumar," A Rainfall Prediction Model using Artificial Neural Network", 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC 2012), pp. 82-87, 2012.

[3]MinghuiQiu, Peilin Zhao, Ke Zhang, Jun Huang, Xing Shi, Xiaoguang Wang, Wei Chu, "A Short-Term Rainfall Prediction Model using Multi-Task Convolutional Neural Networks", IEEE International Conference on Data Mining, pp. 395-400, 2017, DOI 10.1109/ICDM.2017.49.

[4] Aswin S, Geetha P, and Vinayakumar R, "Deep Learning Models for the Prediction of Rainfall", International Conference on Communication and Signal Processing, April 3-5, 2018, India, pp. 0657-0661.

[5] Xianggen Gan, Lihong Chen, Dongbao Yang, Guang Liu, "The Research Of Rainfall Prediction Models Based On Matlab Neural Network", Proceedings of IEEE CCIS2011, pp. 45- 48.

[6] Sam Cramer, Michael Kampouridis, Alex A. Freitas and AntonisAlexandridis, "Predicting Rainfall in the Context of Rainfall Derivatives Using Genetic Programming", 2015 IEEE Symposium Series on Computational Intelligence, pp. 711 – 718.

[7] Mohini P. Darji, Vipul K. Dabhi, HarshadkumarB.Prajapati, "Rainfall Forecasting Using Neural Network: A Survey", 2015 International Conference on Advances in Computer Engineering and Applications (ICACEA) IMS Engineering College, Ghaziabad, India, pp.706 – 713

[8] Sandeep Kumar Mohapatra, Anamika Upadhyay, Channabasava Gola, "Rainfall Prediction based on 100 years of Meteorological Data",2017 International Conference on Computing and Communication Technologies for Smart Nation, pp.162 – 166.

[9] Sankhadeep Chatterjee, Bimal Datta, Soumya Sen, Nilanjan Dey, "Rainfall Prediction using Hybrid Neural Network Approach", 2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom), pp. 67 – 72.

[10] Mr. Sunil Navadia, Mr. Pintukumar Yadav, Mr. Jobin Thomas, Ms. Shakila Shaikh, "Weather Prediction: A novel approach for measuring and analyzing weather data", International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC 2017), pp. 414.