

Development of Multiple Combined Regression Methods for Rainfall Measurement.

Nusrat Jahan Prottasha¹, Md. Jashim Uddin², Md. Kowsher³, Rokeya Khatun Shorna⁴, Niaz Al Murshed⁵, and Boktiar Ahmed Bappy⁶

¹ Daffodil International University Dhaka 1207, Bangladesh,
jahannusratprotta@gmail.com

² Noakhali Science and Technology University, 3814, Dhaka,
mdjaud12@gmail.com

³ Stevens Institute of Technology, Hoboken, NJ 07030 USA,
ga.kowsher@gmail.com

⁴ Daffodil International University, 1207, Dhaka,
rokeyashorna5@gmail.com

⁵ Jahangirnagar University, 1342, Dhaka,
niazalmurshed.ai@gmail.com

⁶ Jhenaidah polytechnic institute, 7300, Dhaka,
entbappy73@gmail.com

Abstract. Rainfall forecast is imperative as overwhelming precipitation can lead to numerous catastrophes. The prediction makes a difference for individuals to require preventive measures. In addition, the expectation ought to be precise. Most of the nations in the world is an agricultural nation and most of the economy of any nation depends upon agriculture. Rain plays an imperative part in agribusiness so the early expectation of rainfall plays a vital part within the economy of any agricultural. Overwhelming precipitation may well be a major disadvantage. It's a cause for natural disasters like floods and drought that unit of measurement experienced by people over the world each year. Rainfall forecast has been one of the foremost challenging issues around the world in the final year. There are so many techniques that have been invented for predicting rainfall but most of them are classification, clustering techniques. Predicting the quantity of rain prediction is crucial for countries' people. In our paperwork, we have proposed some regression analysis techniques which can be utilized for predicting the quantity of rainfall (The amount of rainfall recorded for the day in mm) based on some historical weather conditions dataset. we have applied 10 supervised regressors (Machine Learning Model) and some preprocessing methodology to the dataset. We have also analyzed the result and compared them using various statistical parameters among these trained models to find the bestperformed model. Using this model for predicting the quantity of rainfall in some different places. Finally, the Random Forest regressor has predicted the best r2 score of 0.869904217, and the mean absolute error is 0.194459262, mean squared error is 0.126358647 and the root mean squared error is 0.355469615...

Keywords: Rainfall, Supervised Learning, Regression, Random Forest Tree, AdaBoost Regressor, Gradient Boosting Regressor, XGBoost.

1 Introduction

This research paper proposed a scientific method to predict rainfall quantity based on some different weather conditions considering preceding weather records and present weather situations using some regression analysis techniques .[1] Rainfall determining is exceptionally vital since overwhelming and irregular rainfall can have numerous impacts on many other things like annihilation of river-bank, crops, agriculture, and farms. One of the very deleterious departures is flooding due to the over rain.[2] According to Wikipedia in late summer 2002, enormous storm downpours driven to gigantic flooding in eastern India, Nepal, and Bangladesh, killing over 500 individuals and clearing out millions of houses. Each year in Bangladesh approximately 26,000 square kilometers (10,000 sq mi) (around 18% of the country) is flooded, killing over 5,000 individuals and wrecking more than 7 million homes. On the other hand, Western Sydney is now the "greatest concern" from the worst floods in decades to have ravaged eastern Australia.[3] Jonh C, Rodda et al. presented a very rational method of the rainfall measurement problem. The application of science and innovation that predicts the state of the environment at any given specific period is known as climate determining or weather forecasting. There are many distinctive strategies for climate estimate and weather forecasting. But rainfall prediction is rare. Some of the research has shown some classification method to predict whether it would be rain tomorrow or not. But instead of a classification method for predicting rain, we need to the quantity of the rainfall in a particular place. There is numerous equipment implement for foreseeing rainfall by utilizing the climate conditions like temperature, humidity, weight. These conventional strategies cannot work productively so by utilizing machine learning procedures. we can create an exact comes about rain forecast. Ready to fair do it by having the historical information investigation of rainfall and can anticipate the precipitation for future seasons. In our paper, we presented some predictive regression analysis techniques to quantify rainfall quantity at a place. Here we used more than 10 years of historical data to train our model. The dataset contains various weather conditions of different places. This method can be utilized to predict the rainfall (The amount of rainfall recorded for the day in mm) and avoid the annihilation caused by it to life, agriculture, farm, and property. If we can quantify the rainfall most people can make some decisions before overwhelmed rain-affected. The contributions of this work are summarised as:

- We have assessed a pipeline of making choices for evaluating the finest reasonable rain prediction.
- We have utilized 10 supervised regressors (Machine Learning Model). Because different regressors give us different results. So, it's essential to find out the right model according to the requirements.

- We have discussed a big comparison among all trained models to figure out the best performer.

The paper is organized as takes after: Section II clarifies the related work of different classification strategies for the forecast of rain classification. Section-III depicts the technique and materials utilized. Section-IV depicts the experimental analysis including performance and result. Section V talks about the conclusion of this research work where section VI described about the plan of future.

2 Related Works

In this paper, through a systematic investigation Rodda et al. have presented the rainfall measurement problem, they claim there's an orderly mistake in the estimation of precipitation made in an ordinary way, a mistake which may influence any gauges utilizing these estimations.[3] Besides Prabakaran et al. proposed a method that speaks to a numerical strategy called Linear Regression to anticipate the rainfall in different areas in southern states of India.[4] To improvement Wang et al. showed a case study they proposed an application of generalized regression neural network (GRNN) model to anticipate yearly precipitation in Zhengzhou .[5] On the other hand, Sethi et al. presented an exploiting data mining technique for the early prediction of rainfall called multiple linear regression (MLR) .[6] Sunyoung Lee et al. presented a divide and conquer approach to predict the rainfall based on the locational information only .[7] Also, Bagirov, M Adil et al. developed the Clusterwise Linear Regression (CLR) technique for the prediction of monthly rainfall .[8] In addition, Mohammed Moulana et al. represented machine learning techniques to precipitation prediction the purpose of this project is to offer non-experts simple get to the methods, approaches utilized within the division of precipitation forecast and give a comparative think about among the different machine learning methods.[9] P, Asha, et al proposed a mutual neural classification model for predicting rainfall. [10] S, Sakthivel, et al described neural networks and the rapid miner-based rain prediction system.[11] Diwakar, Naidu, et al presented the changes in rainfall patterns in numerous agro-climatic zones using machine learning approaches.[12] Besides, Tuan Vu, Dinh, et al utilized an LSHADE-PWI-SVM method for the integration of machine learning classifiers conjointly metaheuristic optimization .[13] On the other hand, Malathi, R, et al showed a Information Gain based Feature Selection Method for Weather Dataset for the prediction of rainfall.[14] Also, Nor, SamsiahSani, et al. evaluated many machine learning classifiers based on Malaysian data for rainfall prediction.[15] David, Ahijevych, et al presented a random forest (RF) that is utilized to produce 2-h figures of the probability for the start of mesoscale convective frameworks (MCS-I).[16] John T, Allen, et al performed property and agribusiness, as well as handfuls of fatalities and Wonders related to extreme electrical storms.[15][17] Harold E, Brooks, et al, displayed the current dissemination of serious rainstorms as a work of large-scale natural conditions.[18] Pierre, Gentine, et al. Representing uncertain sodden convection in coarse-scale climate models remains one of the most bottlenecks of

current climate recreations.[19] McPhaden, et al. described the participation of the pivotal for agriculture-dependent.[20] Hazell et al, Represented to reduce the risk of life and also maintain the agriculture farms in a better way [21] Then, Mollinga et al. elucidates farmers to take early measurements of floods, and manage the water resources properly.[22] Shah et al, discussed to related this task to predict rain.[23]

3 Methodology

To perform the complete technique, we assume the four significant steps such as data collection, data pre-processing, training model using 10 supervised regressors, and execution examination. Within the information collection step, we have used a dataset⁷ from the Kaggle platform which has been split into two parts such as the training part and validation part. Here we have utilized one of the validation parts as the testing data to evaluate the models' performance. Each row has various weights for decision making to suggest the sensible best rain prediction. Afterward, gathering all raw data, firstly we would be made ready for the training model with the help of data pre-processing techniques and this has been used for outliers free and more rigid. It also assists to increase the performance of the models. As a result, we have applied six pre-processing methods such as cleaning data, missing value check, handling the categorical data, handling outliers, handling outliers, feature selection. Next, to establish supervised regressors models, we utilized the regressors such as Linear Regression, Ridge Regression, Polynomial Regression, and Lasso Regression. From all the training methods we have used a total of 10 regressors so that we can compare the performance and figure out the best model. Most of the regressors come up with a good performance. We have described the whole methodology in the figure 1:

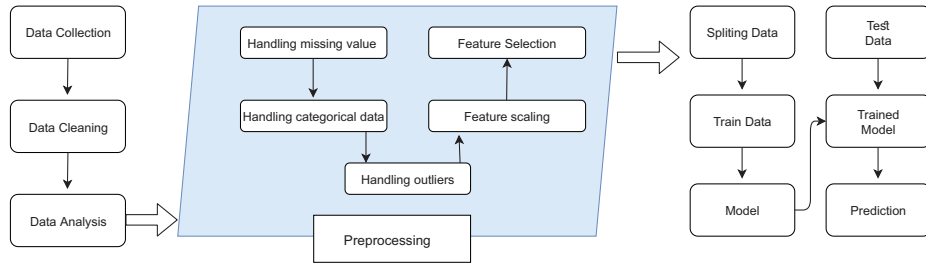


Fig. 1. The whole methodology of rainfall prediction including all important steps such as data collection, necessary preprocessing, and training model with performance prediction

⁷ <https://www.kaggle.com/jsphyg/weather-dataset-rattle-package>