Research Exhibition in Mathematics and Computer Sciences (REMACS 6.0)/
ISBN: 978-629-97440-5-4

RAINFALL PREDICTION BY USING MACHINE LEARNING APPROACH

Siti Nurin Syafiqah Binti Mohd Akmal and Norhayati Shafii
College of Computing, Informatics and Mathematics,
Universiti Teknologi MARA, Perlis Branch
Sitinurinsyafiqahbt.mohdakmal@gmail.com and norhayatishafii@uitm.edu.my

ABSTRACT- This study utilized machine learning to forecast rainfall levels and identify influential atmospheric elements. Python software was employed to analyze historical rainfall data, focusing on 16 selected attributes. The study aimed to predict rainfall amounts and determine the features that influence rainfall the most. Its significance lies in optimizing agricultural practices, enhancing preparedness for heavy rains, and supporting infrastructure planning, sustainable agriculture, and economic resilience. MLR algorithms were utilized and evaluated using the dataset. The data underwent preprocessing to handle missing values and create an organized dataset suitable for model building. The rainfall dataset was divided into a training set (70%) and a testing set (30%). The resulting model's predictive performance for rainfall was assessed by calculating the R-squared value. RFE was employed to select the most relevant two features, ranking them based on relevance and eliminating less important ones. The model developed using RFE demonstrated a reasonably good fit to the testing data, achieving an R-squared value of 0.1316. The RMSE indicated an average difference of around 0.11385 units between predicted and actual values. The errors between predicted and actual values were reasonably minimal, with a MSE of 0.01919 and a MAE of 0.07436. The model accounted for approximately 13% of the variation in rainfall. To conclude, the general aims of this study, where the factors that influence most rainfall have been found, by comparing the model and the Recursive Features Selection (RFE) process. The specific objectives of this study were fully addressed as the MSE and MAE value is low.

Keywords: Machine learning (ML), Multiple Linear Regression (MLR), Recursive Features Selection (RFE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean squared Error (MSE).

1. INTRODUCTION

In any country, especially one in the equatorial climate region, which lacks seasons, rainfall is a common meteorological occurrence (BBC, 2022). Rainwater may restore the soil, according to Selase et al. (2015), it is also necessary for vegetation, it provides a habitat for fish, and it fills reservoirs that hold drinking water. In a previous study, Dar (2017) employed a mathematical strategy known as multiple linear regression. This method allows for the running of several independent variables and the evaluation of the model's accuracy when the MSE, RMSE are, and high value of r squared. Including that, the goal of machine learning (ML), is to connect every relevant information point in order to draw highly precise judgements and ultimately alter the behavior of the model. Based on climatic variables like temperature, humidity, and pressure, rainfall can be anticipated. Thus, in this research study ML approach and MLR were used with more attributes.

2. METHODOLOGY

The dataset for this study will be obtained from Kaggle.com. This secondary data collection consists of 366 records with 18 attributes from the rainfall dataset. Phyton software will be used to assess historical rainfall data. MinTemp, MaxTemp, Evaporation, Sunshine, WindGustSpeed, WindSpeed9am, WindSpeed3pm, Humidity9am, Humidity3pm, Pressure9am, Pressure3pm, Cloud9am, Cloud3pm, Temp9am, Temp3pm, Rainfall, RainToday, and RainTomorrow are all included in this dataset. The prediction accuracy will be calculated using all of these attributes except for RainToday, and RainTomorrow. Machine Learning approach was use in this study by conducting several processes which are data collection, identify dependent variables and independent variables, data preparation, data preprocessing, data describe, build regression model which is Multiple Linear Regression Model, evaluation and deployment process. Recursive Features Elimination was also used in this study to select 2 features related to amount of rainfall.

3. RESULTS AND DISCUSSION

The value observed is the MAE, MSE and RMSE and adjusted R Squared. As 2 features were requested for ranking the relevant variables affecting amount of Rainfall, then the system generate it for closer observation. The features

given were 'Humidity9am', and 'Pressure9am'. Based on the machine learning methodology, the R2 value obtained for the test data is 0.1316, exhibiting a high degree of similarity to the R2 value observed for the RFE method. The difference between RMSE of training dataset and RMSE of testing dataset is just 0.0371 units, it can be concluded that the model we have developed is the most optimal in terms of fitting the data. The mean squared error (MSE) value of 0.01919 and mean absolute error (MAE) value of 0.07436 suggest that the discrepancies between the predicted and actual values are relatively minimal. The R-squared score of the MLR model is approximately 0.1316. This implies that the model explains approximately 13% of the variability observed in the target variable.

4. NOVELTY OF RESEARCH / PRODUCT

The World Health Organisation (2022) records that from 1998 to 2017, floods caused injury to over 2 billion people worldwide. Excessive rainfall may also have an effect on the agriculture system. Due to this issue, the desire to prevent unpredictable, heavy rainfall has increased. It is an honor to learn to analyse and solve problems by using Phyton software. The experience of frustration to get a good model for rainfall prediction gave full relief once realize this small step is a part of polishing my skill and knowledge.

5. CONCLUSION

To conclude, the general aims of this study, where the factors that influence most rainfall have been found which are 'Humidity9am', and 'Pressure9am' features by comparing the model and the Recursive Features Selection (RFE) process. The specific objectives of this study were fully addressed as the MSE and MAE value is low, the smaller the value implies higher accuracy of the regression model which means that all expected values matched with the prediction values.

REFERENCES

BBC Bitesize. (2023). Equatorial climate - Natural regions - National 5 Geography Revision - BBC Bitesize. https://www.bbc.co.uk/bitesize/guides/z8r6fg8/revision/7#:~:text=Rainforests%20are%20located%20in%20the%20equatorial%20climate%20region.,only%20a%20few%20degrees.%20There%20are%20no%20season

Dar, L. A. (2017). Rainfall-Runoff Modeling using Multiple Linear Regression Technique RAINFALL-RUNOFF MODELING OF JHELUM RIVER BASIN. SJ Impact Factor:6, 887. www.ijraset.com214

KGomathy, C., Bala Narasimha Reddy, A., Pavan Kumar, A., & LOKESH Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya, A. (2021). A study on rainfall prediction techniques. *International Journal of Scientific Research in Engineering and Management (IJSREM)*. https://www.researchgate.net/publication/357448842

Selase, A. E., Eunice, D., Agyimpomaa, E., Selasi, D. D., Melody, D., & Hakii, N. (2015). Business Management, Keta Business Senior High School. In Online) (Vol. 5, Issue 20). www.iiste.org

World Health Organization: WHO. (2019). Floods. www.who.int. https://www.who.int/health-topics/floods/#tab=tab_1