BALTIC JOURNAL OF LAW & POLITICS
A Journal of Vytautas Magnus University
VOLUME 15, NUMBER 4 (2022)
ISSN 2029-0454

Cite: *Baltic Journal of Law & Politics* 15:4 (2022): 424-430
DOI:10.2478/bjlp-2022-004045

Rainfall Accuracy Prediction using Machine Learning Technique based on Linear Regression over Logistic Regression

K.Nanda Kumar
Research Scholar, Department of Computer Science and Engineering,Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105

V Chandrasekar
Project Guide, Corresponding Author, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105

Received: August 8, 2022; reviews: 2; accepted: November 29, 2022.

Abstract

Aim: The main aim of the research is to predict rainfall using Linear Regression over Logistic Regression. **Materials and Methods:** Linear Regression and Logistic Regression are implemented in this research work. Sample size is calculated using G power software and determined as 10 per group with pretest power 80%, threshold 0.05% and CI 95%. **Result:** Linear Regression provides a higher of 91.18% compared to Logistic Regression algorithm with 87.05% in predicting rainfall. Two groups differ significantly from one another, as shown by a significance value of 0.003 ($p<0.05$). **Conclusion:** Linear Regression algorithm predicts the rainfall better than Logistic Regression algorithm.

Keywords:
Climate, Rainfall Prediction, Novel Linear Regression, Logistic Regression, Forecasting.

INTRODUCTION

The work is about Rainfall Prediction using Novel Linear Regression over Logistic Regression. The prediction using machine learning has succeeded in comparing Linear Regression over Logistic Regression Algorithm. The ability to predict rainfall helps prevent flooding, save human lives and propertRainfall is a crucial occurrence within a meteorological system, because its irregular nature immediately affects water-useful planning, agricultural, and biological processes.(Cramer et al. 2017).Precipitation is intimately linked to the vast majority of the pressing contemporary issues, including soil erosion, draughts, heat waves, and weather-related difficulties(Mohammed et al. 2020). Applying the logistic regression approach, which is completely dependent on retrospective Demeter forecasts, to monthly rainfall projections in India's meteorological divisions (Prasad, Dash, and Mohanty 2010). Predicting rainfall is a vital component of weather forecasting. In the past, attempts have been made to forecast the daily rainfall using certain approaches.The applications of Rainfall

downscaling represents precipitation and temperature scenarios based totally on weather move indices(Mahmood 2017).

In the last five years,Google Scholar identified almost 17000 articles on Rainfall Prediction using Machine Learning. To develop a prediction model for accurate rainfall, forecasting poses one of the greatest problems for academics in a variety of domains, including meteorology, environmental device learning, and numerical forecasting (Basha et al. 2020). An important component of the hydrological cycle is rainfall, and changes to its pattern have an immediate effect on water reserves.(Praveen et al. 2020).India is an agricultural nation, and crop production plays a significant role in the country's economy. Rainfall is a natural miracle defined as the outgrowth of commerce between several complicated atmospheric procedures (Barde and Patole 2016). The perpendicular histories of temperature and moisture are the best forecasters of rainfall, but new factors like the wind and company cleave or confluence also be important(Yang et al. 2019).Forecasting is a difficult task, and doing so for the climate is even more concerning(Rodrigues and Deshpande 2017). From all these research papers, the best study paper in my opinion is (Rodrigues and Deshpande 2017).

Previously our team has a rich experience in working on various research projects across multiple disciplines (Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020).The research gap identified from the existing system shows poor accuracy. Linear Regression is a direct approach for modeling the association between a scalar reply and one or further explanatory variables. Thus, the primary goal of the study is to forecast rainfall using linear regression rather than logistic regression

MATERIALS AND METHODS

This study setting was done in the Soft Computing Laboratory, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The number of required samples in research are two in which group 1 is Linear Regression compared with group 2 of Logistic Regression Algorithm. The samples were taken from the device and iterated 10 times to get desired accuracy with G power 80%, threshold 0.05% and CI 95% (Vannitsem, Wilks, and Messner 2018).

Linear Regression

Linear Regression was applied to forecast the value of the variable predicated on another variable. Linear Regression is extensively used in Machine Learning to make predictions.Linear Regression is often used in rainfall predictions to predict future rains. It has a big effect on rainfall prediction. So, the program predicts the rainfall.

Pseudocode for Linear Regression

Step1: Import packages.
Step2: Create an input dataset.
Step3: Analyze the size of taken input data.
Step4: Split the datasets for testing and training the dataset.
Step5: Apply Linear Regression
Step6: Predict the results.

Logistic Regression

A scientific analysis model called logistic regression is used to predict a data value based on prior data set observations. It is also used to measure the accuracy of a rainfall prediction.Logistic Regression also requires the required amount of input data to perform the assigned task.

Pseudocode for Logistic Regression

Step1: Import packages.
Step2: Create an input dataset.
Step3: Analyze the size of taken input data.
Step4: Split the datasets for testing and training the dataset.
Step5: Apply Logistic Regression.
Step 6: Predict the results.

Recall that the testing setup includes both hardware and software configuration choices. The laptop has an Intel Core i7 8th generation CPU with 12GB of RAM, x86-based processor, a 64-bit operating system, and a hard drive. Currently, the software runs on Windows 10 and is programmed in Python. Once the program is finished, the accuracy value will appear. Procedure: Wi-Fi laptop connected. Chrome to Google Collaboratory search Write the code in Python. Run the code. To save the file, upload it to the disc, and create a folder for it. Log in using the ID from the message. Run the code to output the accuracy and graph.

Statistical Analysis

SPSS is a software tool used for statistics analysis. The designed system utilized 10 iterations for each group with predicted accuracy noted and analyzed. Absolute samples t-test was done to obtain significance between two groups. Month and annual parameters are independent variables and Year is dependent variable. According to the hydrological cycle, the quantity of the face runoff, which occurs after a rainfall event, is determined by the amount of water lost in the river basin as well as the time and regional variations in rainfall patterns(Chan, Chen, and Lee 2018).

RESULTS

Table 1 shows the accuracy value of iteration of Novel Linear Regression and Logistic Regression.

Table 2 represents the Group statistics results which depicts Linear Regression with mean accuracy of 91.18%, and standard deviation is 2.63. Logistic has a mean accuracy of 87.05% and standard deviation is 1.21. Proposed Linear Regression algorithm provides better performance compared to the Logistic Regression algorithm.

Table 3 displays the independent samples T-test value for Linear Regression and Logistic Regression with Mean difference as 4.12, std Error Difference as 0.91. Significance value is observed as 0.003 ($p>0.05$).

Figure 1 displays a bar graph comparing the accuracy mean on Linear Regression and Logistic Regression algorithm. Mean accuracy of Novel Linear Regression is 91.18% and Logistic Regression is 87.05%.

DISCUSSION

In this study, predicting rainfall using Novel Linear Regression has significantly higher accuracy, approximately 91.18% in comparison to Logistic Regression 87.05%. Linear Regression appears to drop added coherent results with lowest standard deviation.

The similar findings of the paper (Imon et al. 2012) had an accuracy of 86% with Logistic Regression which was used to predict rainfall. The proposed work of reported Logistic Regression has 82% accuracy which is used to predict rainfall. The work proposed by (C, Srikantaiah, and Sanadi 2021) shows the Novel Linear Regression yields a 90% improvement in accuracy. Logistic Regression is a parameter to measure rainfall and climate conditions which is used in both traditional and modern methods (Chan, Chen, and Lee 2018) as per their research it opposes Linear Regression has highest accuracy and Logistic will get least accuracy compared to other machine learning techniques which ranges between 60% when

compared to other machine learning algorithms will get more accuracy than this. By using Logistic Regression for forecasting rainfall it will have key issues to pretend (Chhetri et al. 2020) in this paper shows Linear Regression has the least accuracy of 79%. Increasing the dataset's value only tends to get desired accuracy. Linear Regression performs better with a combination of other machine learning algorithms.

The drawback of this research is that it can't provide useful findings with lower amounts of data.This model can't take into account all of the specified feature variable parameters during training. The proposed work's future scope will include rainfall prediction based on categorization using class labels for less complicated time periods.

CONCLUSION

The results show that the performance of Novel Linear Regression in predicting accuracy of rainfall (91.18%) is better than that of the Logistic Regression(87.05%) in terms of accuracy. The Linear Regression method predicts rainfall more accurately than the Logistic Regression approach.. The prediction results denoted that the rainfall analysis proceeding showed exact forecasts.

DECLARATION

Conflict of Interests

No conflict of interests in this manuscript

Authors Contribution

Author KN was involved in data collection, data analysis, manuscript writing. Author VC was involved in conceptualization, data validation, and review of manuscript.

Acknowledgement

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary Infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Edify Techno Solutions, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

Barde, Nishchala C., and M. Patole. 2016. "Classification and Forecasting of Weather Using ANN, K-NN and Naïve Bayes Algorithms." https://doi.org/10.21275/v5i2.nov161672.

Basha, Cmak Zeelan, Nagulla Bhavana, Ponduru Bhavya, and V. Sowmya. 2020. "Rainfall Prediction Using Machine Learning & Deep Learning Techniques." *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*. https://doi.org/10.1109/icesc48915.2020.9155896.

Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.

Chan, Hsun-Chuan, Po-An Chen, and Jung-Tai Lee. 2018. "Rainfall-Induced Landslide Susceptibility Using a Rainfall–Runoff Model and Logistic Regression." *Water*. https://doi.org/10.3390/w10101354.

Chhetri, Manoj, Sudhanshu Kumar, Partha Pratim Roy, and Byung-Gyu Kim. 2020. "Deep BLSTM-GRU Model for Monthly Rainfall Prediction: A Case Study of Simtokha, Bhutan."

Remote Sensing. https://doi.org/10.3390/rs12193174.

Cramer, Sam, Michael Kampouridis, Alex A. Freitas, and Antonis K. Alexandridis. 2017. "An Extensive Evaluation of Seven Machine Learning Methods for Rainfall Prediction in Weather Derivatives." *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2017.05.029.

C, Srikantaiah K., K. C. Srikantaiah, and Meenaxi M. Sanadi. 2021. "Rainfall Prediction Using Logistic Regression and Support Vector Regression Algorithms." *Communications in Computer and Information Science*. https://doi.org/10.1007/978-3-030-81462-5_54.

Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.

Imon, A. H. M. Rahmatullah, A. H. Rahmatullah Imon, Manos C. Roy, and S. K. Bhattacharjee. 2012. "Prediction of Rainfall Using Logistic Regression." *Pakistan Journal of Statistics and Operation Research*. https://doi.org/10.18187/pjsor.v8i3.535.

Mahmood, Arshad. 2017. *Rainfall Prediction in Australia: Clusterwise Linear Regression Approach*.

Mohammed, Moulana, Roshitha Kolapalli, N. V. V. Golla, and S. Maturi. 2020. "Prediction Of Rainfall Using Machine Learning Techniques." https://www.semanticscholar.org/paper/Prediction-Of-Rainfall-Using-Machine-Learning-Mohammed-Kolapalli/9bf0a85ff6323559663cb45c89ae1e51684401dc#citing-papers.

Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.

Prasad, K., S. K. Dash, and U. C. Mohanty. 2010. "A Logistic Regression Approach for Monthly Rainfall Forecasts in Meteorological Subdivisions of India Based on DEMETER Retrospective Forecasts." *International Journal of Climatology*. https://doi.org/10.1002/joc.2019.

Praveen, Bushra, Swapan Talukdar, Shahfahad, Susanta Mahato, Jayanta Mondal, Pritee Sharma, Abu Reza Md Towfiqul Islam, and Atiqur Rahman. 2020. "Analyzing Trend and Forecasting of Rainfall Changes in India Using Non-Parametrical and Machine Learning Approaches." *Scientific Reports* 10 (1): 1–21.

Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.

Rodrigues, Jesleena, and Arti Deshpande. 2017. "Prediction of Rainfall for All the States of India Using Auto-Regressive Integrated Moving Average Model and Multiple Linear Regression." *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. https://doi.org/10.1109/iccubea.2017.8463914.

Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.

Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.

Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.

Vannitsem, Stéphane, Daniel S. Wilks, and Jakob Messner. 2018. *Statistical Postprocessing of Ensemble Forecasts*. Elsevier.

Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.

Yang, Junho, Mikyoung Jun, Courtney Schumacher, and R. Saravanan. 2019. "Predictive Statistical Representations of Observed and Simulated Rainfall Using Generalized Linear Models." *Journal of Climate* 32 (11): 3409–27.

Tables and Figures

Table 1. Accuracy Values for Linear and Logistic Regression

S.NO	Linear	Logistic
1	89.76	86.12
2	89.89	85.77
3	87.05	86.95
4	87.45	88.89
5	93.15	87.59
6	93.45	86.75
7	90.28	85.33
8	94.07	86.45
9	93.58	87.95
10	93.17	88.76

Table 2. Group Statistics Results-Linear Regression has an mean accuracy (91.18%), std.deviation (2.63), whereas for Logistic Regression has mean accuracy (87.05%), std.deviation (0.38).

Group Statistics					
	Groups	**N**	**Mean**	**Std deviation**	**Std. Error Mean**
Accuracy	Linear	10	91.18	2.63	0.83
	Logistic	10	87.05	1.21	0.38

Table 3. Independent Samples T-test - Linear Regression seems to be significantly better than Logistic Regression(p=0.003)

Accuracy	Independent Samples Test								
	Levene's Test for Equality of Variances		T-test for Equality of Means						
	F	Sig	t	df	Sig(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	11.643	0.003	4.496	18	0.00	4.12	0.91832	2.19967	6.05833
Equal variances not assumed			4.496	12.669	0.01	4.12	0.91832	2.13980	6.11820

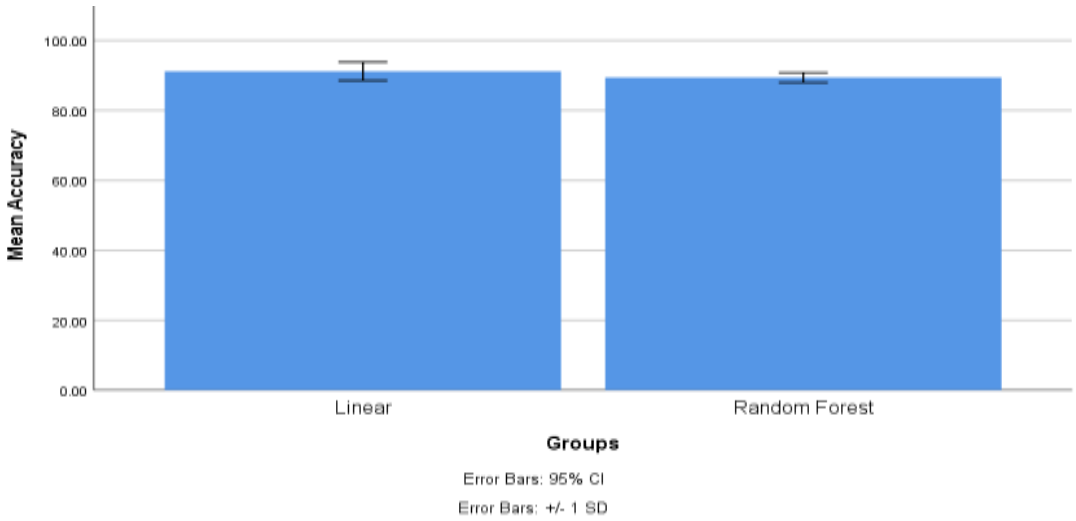

Fig. 1. Bar Graph Comparison on mean accuracy of Linear Regression (91.18%) and Logistic Regression (87.05%). X-axis: Linear,Logistic Regression, Y-axis: Mean Accuracy with ±1 SD.