EARLY PREDICTION OF RAINFALL USING XGBOOST ALGORITHM IN COMPARISON WITH LOGISTIC REGRESSION

Shaik Arshad Hussain[1], Terrance Frederick Fernandez[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The objective of the work is to evaluate the accuracy and precision in predicting the rainfall using machine learning algorithms novel tree specific XGBoost (XGB) classification and Logistic Regression (LR) algorithms.
Materials and Methods: Novel Tree Specific XGBoost classifier is applied on a weatherAUS dataset that consists of 145461 records. A framework for rainfall prediction machine learning algorithms comparing XGBoost and Logistic Regression classifiers has been proposed and developed. The sample size was measured as 10 per group. Sample size was calculated using clinical analysis, with alpha and beta values 0.05 and 0.5, 95% confidence, pretest power 80% and enrolment ratio 1. The significance value (p) obtained for both accuracy and precision is 0.019, which is less than 0.05. The accuracy and the precision of the classifiers were evaluated and recorded.
Results: The machine learning algorithm Logistic Regression classifier produces 79.37% accuracy and 78.00% precision in predicting the rainfall on the dataset used whereas the another machine learning algorithm novel tree specific XGboost classifier predicts the same at the rate of 94.89% accuracy and 94.37% precision. **Conclusion:** The study proves that novel tree specific XGboost classifier algorithm exhibits better accuracy and precision than Logistic Regression algorithm in rainfall prediction.

Keywords: Novel Tree Specific XGBoost, Logistic Regression, Agriculture, Rainfall, Prediction, Machine Learning Algorithm.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India: 602105.
[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India: 602105.

1. Introduction

The primary occupation of India has been agriculture so the economy of the country highly depends on it. Therefore, prediction of rainfall is important for the economic growth of the country. Rainfall prediction has been one of the most challenging and difficult tasks in the world. India has a population of over a billion people, agriculture is the primary income of more than 60% of the population (Brownlee 2020). The advancement of technology and resources in today's world is increasing day by day. Weather of a particular location can be predicted using the applications of science and technology. Many attempts were made to predict the rainfall in the past to help the farmers to do agriculture . Informally these attempts were started way back in the eighteenth century and formally started in the nineteenth century (Dhyani 2020). Weather forecasting basically refers to the analyzing and predicting the climatic conditions of a specific region in the future. Rainfall prediction comes under the category of weather forecasting (Yuan and Forshay 2021). Interest in weather prediction started in the earliest days to help farmers for effective agriculture, and many rainfall prediction techniques were introduced to effectively predict the rainfall. Over time all those techniques evolved and each of the techniques has its own accuracy and efficiency (Kuradusenge, Kumaran, and Zennaro 2020). The proposed study helps in effectively predicting the rainfall that helps the overall development of the country.

There are 105 research articles published on the rainfall prediction in IEEE xplore and 165 articles on google scholar and 34 articles were found in sciencedirect. (Fadilah, Wigena, and Djuraidah 2020) presents a new predictor algorithm based on Bayesian Enhanced Approach (BEA) for long-term chaotic time series using Artificial Neural Networks (ANN). In many cases, the simplest representations of prior information in forecasting models are hard to surpass; the Bayes technique gives another approach to include prior knowledge in forecasting models. Predicting conditions, either because previous knowledge isn't available or because prior knowledge isn't useful. As a result, the situation appears to be stable. (Wen et al. 2018) Compared various types of rainfall models which have been used in modeling rainfall occurrence, rainfall amount or combination of both rainfall occurrence and rainfall amount. (Narejo et al. 2021) Proposed a multiple linear regression model in order to predict the rate of precipitation (PRCP). It is based on some weather parameters, such as temperature, wind speed, and dew point. The data used in this research has been provided from the website of the National Climatic Data Center. A Python code using the Pytorch library has been written to develop the model, which applies to Artificial Neural Networks. (Mahmood 2017) Proposed an algorithm for rainfall prediction through the random forest. It achieves an accuracy of 80.56% and outperforms all other feature representations, including decision tree and support vector machine classification algorithms using the same classifier on the same dataset.Our team has extensive knowledge and research experience that has translated into high quality publications(Pandiyan et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022) (Pandiyan et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022)

The research gap identified from the survey is that there are many methods proposed for rainfall prediction but most of those methods have less accuracy rate. Several works have demonstrated that the performance of machine learning algorithm Logistic Regression is poor and provides less accuracy in prediction of the rainfall. A study by (Mohammed et al. 2020) compares the accuracy of various mining classification algorithms in predicting rainfall. It is important to analyze and compare the various classification algorithms that provide better accuracy. Hence, the work aims at comparing the accuracy of Logistic Regression and XGBoost algorithms in predicting the rainfall.

2. Materials and Methods

The research work was carried out at the Data analytics lab, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, where the laboratory facilitates high configuration systems to obtain the experimental results. The number of groups identified for the study were two with the sample size used for experimenting the low false alarm rate is 52 and group1 consists of 10 sample sizes and group2 consists of 10 sample sizes. The computation is performed using G-power 0.95 with alpha value is 0.005 and beta value is 0.95 with a confidence interval 95%. The dataset was downloaded from Kaggle website (kerneler 2019).

Logistic Regression(LR) - Group 1
Inputs: WeatherAUS data set
Output: Selected features and Accuracy.
1. Load the dataset
2. Split the dataset randomly into training (80%) and testing (20%) dataset

3. Set the target variable
4. Generate the Logistic Regression classifier based on the training set
5. Train the classifier using rbf kernel parameter
6. Predict the testing set based on training dataset
7. Evaluate the classifier.
8. Return Accuracy.

Logistic Regression(LR) is a regulated machine learning algorithm which can be utilized for both classification and regression challenges. In this study, to train the Logistic Regression the Logistic Regression class of sklearn.linear_model library was used. Import the weatherAUS.csv dataset and load the dataset. The dataset is split randomly into training (80%) and testing (20%) sets. The target variable is selected. Then, the 3 of 15 Logistic Regression classifiers based on the training set is generated. Rbf was used as the value of the kernel parameter. The testing set is predicted based on the training set. The machine learning algorithm Logistic Regression classifier is evaluated and the accuracy is calculated.

XGBoost(XGB) - Group 2
Input: weatherAUS dataset
Output: Accuracy
1. Import and read the dataset
2. Select the features randomly from the dataset
3. Generate the XGB classifier criterion as a parameter.
4. Gini was used as a parameter value.
5. Construct a decision tree using XGB classifiers and predict the result for every sample.
6. Voting was performed for every predicted result.
7. Most voted prediction results were selected as the final outcome.
8. Return precision.

In this study, the XGBoost class of the sklearn ensemble library is used. The dataset is splitted randomly into training(80%) and testing(20%). It selects samples randomly and the decision trees were collected for every sample to predict the result. Voting was performed for every predicted result and the most voted result was selected as the final result. The algorithm uses a Novel Tree Specific XGBoost Classifier (NTS XGB).

The various parameters for the analysis can be calculated as follows:
Equation (1) - Accuracy : It identifies the number of instances that were correctly classified.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad (1)$$

Here “TN” means True Negative, “TP” means True Positive, “FP” means False Positive and “FN” means False Negative.

Equation (2) - Precision is used to calculate which part of prediction data is positive.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Equation (3) - Recall is also called sensitivity which calculates the relevant instances that are selected.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Equation (4) - F-measure measures model accuracy on a dataset.

$$F - measure = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall}\right) \quad (4)$$

This study was implemented using Google collab and SPSS software, and the hardware configuration required is an intel i3 processor, 50GB HDD, 4GB RAM, and the software configuration required is a windows OS.

Statistical Analysis

Besides experimental analysis, the work was evaluated statistically using Statistical Package for Social Sciences (SPSS). The analysis was done to obtain Mean, Standard Deviation and Standard Error Mean. The independent variables are temperature, humidity, wind speed and pressure and the dependent variable is accuracy(Dattalo 2013). An independent variable T-Test was carried out to compare the parameters on both the groups.

3. Results

Table 1 shows the comparison of accuracy and precision achieved during the evaluation of the machine learning algorithms Logistic Regression and XGBoost models for classification with different iterations. Table 2 depicts the various parameters of both groups. The accuracy, Precision, Recall, F1 Score and support has been calculated for Logistic Regression and novel tree specific XGBoost algorithm. The analysis of two groups shows that the XGB algorithm has higher accuracy (94.89%) and Precision (94.37%) compared to Logistic Regression. From Figure. 1, ROC graph shows the performance of Logistic Regression classification models at various classification thresholds. From Figure. 2, ROC graph shows the performance of novel tree specific XGBoost classification model at various classification thresholds. Table 3 shows the statistical analysis of Logistic Regression and XGBoost with different test datasets. The mean Accuracy of the XGBoost model appears to be higher than the Logistic Regression model. Also, the precision of XGBoost is much higher than the Logistic Regression. The performance of the novel tree specific XGBoost algorithm is superior to the Logistic Regression algorithm. Table 4 depicts the statistical analysis of significant levels for both groups. The significance difference obtained for both accuracy and precision of both the groups is 0.019. Hence XGBoost is better than Logistic Regression. Figure. 3 and 4 inferred the mean accuracy and mean precision of Logistic Regression and novel tree specific XGBoost classifiers. The statistical analysis of two independent groups shows that the novel tree specific XGBoost algorithm has higher accuracy mean(94.89%) and precision mean (94.37%) compared to Logistic Regression accuracy mean(79.37%) and precision mean(78.00%). The standard mean error for the precision of the XGBoost algorithm is a little higher than the standard mean error for the precision of the Logistic Regression algorithm.

4. Discussion

Agriculture has been the backbone of India and prediction of rainfall is a major issue in the agriculture sector. Experimental work was done among two groups: Logistic Regression(LR) and novel tree specific XGBoost(XGB) by varying the test size. From the experimental results (Figure 3 & 4) done in SPSS, the accuracy and the precision of XGB are 94.89% and 94.37% respectively, whereas Logistic Regression provides the accuracy and the precision 79.37% and 78.00% respectively. This depicts that XGB is better than Logistic Regression. The various parameters like TP rate, FP rate, Precision, Recall, F-measures are also compared. From the SPSS graph proposed XGBoost classifier performs better in terms of accuracy and precision compared with the logistic Regression algorithm. The analysis depicts that the mean error of precision of xgb is found to be little higher than the mean error of precision of Logistic Regression, which has to be minimized.

The two most important factors in rainfall prediction are accuracy and precision. In the study by (Yen et al. 2019) a machine learning based rainfall prediction by using the weather dataset was proposed. The research proved that the XGBoost algorithm is producing better accuracy than the other algorithms. In the study by (Anwar et al. 2021) XGBoost, catboost and support vector machine classification algorithms were compared and proved that the XGBoost algorithm is producing the best accuracy level of 93.59% compared to which catboost produced the accuracy level of 85.29% and SVM produced the accuracy level of 78.05%. In the study by (Dash, Mishra, and Panigrahi 2019) different algorithms were used to predict the accurate rainfall. Among all those machine learning algorithms, XGBoost stands out by resulting in higher accuracy. In the study by (Osman et al. 2021) a machine learning based rainfall prediction is done using the XGBoost algorithm and the SVM algorithm, XGBoost is concluded as the best in terms of accuracy. (Alexandridis and Zapranis 2014) has done research on the accurate rainfall prediction for agriculture purposes in the large area using genetic programming in application with neural networks. This technique can also be implemented and tested if it can produce better accuracy than the XGBoost algorithm for larger regions. The research from (Zhang et al. 2020) proves that the catboost algorithm is producing the same accuracy as the XGBoost algorithm in predicting the accurate and effective rainfall. Also, The precision value is higher for the precision of XGBoost algorithms compared with the Logistic Regression algorithms.

Although the proposed methodology attained satisfactory results, there are certain limitations in the work. The evaluation of accuracy cannot provide a better outcome on larger data sets. Moreover in XGB, the mean error appears to be higher for the precision than the mean error for the precision of LR. It would be better if the mean error can be reduced to a considerable extent. However, the work can be enhanced by applying optimization algorithm techniques, to achieve better accuracy and less mean error. Feature selection algorithms can be used before classification to improve the classification accuracy of classifiers.

5. Conclusion

XGBoosting is a classification technique that uses averaging to improve the accuracy and precision. The work shows that the accuracy and precision for rainfall prediction using XGBoost(XGB) appears to be better than the Logistic Regression(LR). It is found that XGB performs significantly better than LR in predicting the rainfall accurately, but the mean error in XGB is found to be little higher than LR. Hence, it is concluded that the XGBoost (XGB) classifier results in acceptable accuracy(94.89%) and precision(78.00%) than Logistic Regression(LR) accuracy(79.37%) and precision(78.00%).

DECLARATIONS

Conflicts of Interests

No conflicts of interest in this manuscript.

Author Contributions

Author SAH was involved in data collections, data analysis, algorithm framing, implementation and manuscript writing. Author TFF was involved in designing the workflow, guidance, and reviewing the manuscript.

Acknowledgements
The authors 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. Saveetha University.
2. Saveetha Institute of Medical And Technical Sciences.
3. Saveetha School of Engineering.

6. References

Alexandridis, Antonios K., and Achilleas D. Zapranis. 2014. *Wavelet Neural Networks: With Applications in Financial Engineering, Chaos, and Classification*. John Wiley & Sons.

Anwar, M. T., E. Winarno, W. Hadikurniawati, and M. Novita. 2021. “Rainfall Prediction Using Extreme Gradient Boosting.” *Journal of Physics. Conference Series* 1869 (1): 012078.

Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. “Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review.” *Chemosphere*. https://doi.org/10.1016/j.chemosphere.2022.134596.

Brownlee, Jason. 2020. “How to Use XGBoost for Time Series Forecasting.” August 4, 2020. https://machinelearningmastery.com/xgboost-for-time-series-forecasting/.

Dash, Yajnaseni, Saroj K. Mishra, and Bijaya K. Panigrahi. 2019. “Predictability Assessment of Northeast Monsoon Rainfall in India Using Sea Surface Temperature Anomaly through Statistical and Machine Learning Techniques.” *Environmetrics*. https://doi.org/10.1002/env.2533.

Dattalo, Patrick. 2013. *Analysis of Multiple Dependent Variables*. Oxford University Press.

Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. “Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review.” *Bioresource Technology*. https://doi.org/10.1016/j.biortech.2022.127234.

Dhyani, Sandeep. 2020. “Predicting Rainfall for Agriculture in India Using Regression.” Dublin Business School. https://esource.dbs.ie/handle/10788/4230.

Fadilah, Putri Maulidina, Aji Hamim Wigena, and Anik Djuraidah. 2020. “Extreme Rainfall Prediction Using Bayesian Binary Quantile Regression.” *International Journal of Scientific and Research Publications (IJSRP)*. https://doi.org/10.29322/ijsrp.10.11.2020.p10702.

Karpagam, M., R. Beaulah Jeyavathana, Sathiya Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. “A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims.” *Soft Computing*. https://doi.org/10.1007/s00500-022-06931-1.

kerneler. 2019. “Starter: weatherAUS 536c1115-4.” Kaggle. November 29, 2019. https://kaggle.com/kerneler/starter-weatheraus-536c1115-4.

Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. “Ultrasonication Time Optimization for Multi-Walled Carbon Nanotube Based Therminol-55 Nanofluid: An Experimental Investigation.” *Journal of Thermal Analysis and Calorimetry*. https://doi.org/10.1007/s10973-022-11298-4.

Kuradusenge, Martin, Santhi Kumaran, and Marco Zennaro. 2020. “Rainfall-Induced Landslide Prediction Using Machine Learning Models: The Case of Ngororero District, Rwanda.” *International Journal of Environmental Research and Public Health* 17 (11). https://doi.org/10.3390/ijerph17114147.

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.

Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. “Combined Economic Emission Dispatch of Microgrid with the

Incorporation of Renewable Energy Sources Using Improved Mayfly Optimization Algorithm." *Computational Intelligence and Neuroscience* 2022 (April): 6461690.

Nagaraju, V., B. R. Tapas Bapu, P. Bhuvaneswari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. "Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency Bands." *Silicon*. https://doi.org/10.1007/s12633-022-01825-1.

Narejo, Sanam, Muhammad Moazzam Jawaid, Shahnawaz Talpur, Rizwan Baloch, and Eros Gian Alessandro Pasero. 2021. "Multi-Step Rainfall Forecasting Using Deep Learning Approach." *PeerJ Computer Science* 7. https://doi.org/10.7717/peerj-cs.514.

Osman, Ahmedbahaaaldin Ibrahem Ahmed, Ali Najah Ahmed, M. F. Chow, Yuk Feng Huang, and A. El-Shafie. 2021. "Extreme Gradient Boosting (Xgboost) Model to Predict the Groundwater Levels in Selangor Malaysia." https://doi.org/10.1016/J.ASEJ.2020.11.011.

Pandiyan, P., R. Sitharthan, S. Saravanan, Natarajan Prabaharan, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. "A Comprehensive Review of the Prospects for Rural Electrification Using Stand-Alone and Hybrid Energy Technologies." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102155.

Venu, Harish, Ibham Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. "Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends." *Energy*. https://doi.org/10.1016/j.energy.2022.123806
.

Wen, Huaiyu, Sufang Li, Wei Li, Jianping Li, and Chang Yin. 2018. "Comparision of Four Machine Learning Techniques for the Prediction of Prostate Cancer Survivability." *2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. https://doi.org/10.1109/iccwamtip.2018.8632577.

Whangchai, Niwooti, Daovieng Yaibouathong, Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. "Effect of Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn Culture." *Chemosphere* 301 (August): 134638.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review." *Chemosphere* 299 (July): 134390.

Yen, Meng-Hua, Ding-Wei Liu, Yi-Chia Hsin, Chu-En Lin, and Chii-Chang Chen. 2019. "Application of the Deep Learning for the Prediction of Rainfall in Southern Taiwan." *Scientific Reports* 9 (1): 1–9.

Yuan, Lifeng, and Kenneth J. Forshay. 2021. "Enhanced Streamflow Prediction with SWAT Using Support Vector Regression for Spatial Calibration: A Case Study in the Illinois River Watershed, U.S." *PloS One* 16 (4): e0248489.

Zhang, Pengcheng, Yangyang Jia, J. Gao, Wei Song, and H. Leung. 2020. "Short-Term Rainfall Forecasting Using Multi-Layer Perceptron." http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8468083.

TABLES AND FIGURES

Table 1. Comparison of Accuracy and Precision achieved during the evaluation of Logistic Regression and XGBoost models for classification with different iterations.

Algorithm	Accuracy	Precision
LR	79.0%	77.4%
LR	79.3%	77.9%
LR	79.4%	78.1%
LR	79.5%	78.4%
LR	79.5%	78.3%

LR	79.5%	78.5%
LR	79.5%	78.4%
LR	79.4%	78.1%
LR	79.4%	78.1%
LR	79.2%	76.8%
XBG	94.7%	91.6%
XBG	94.8%	91.7%
XBG	94.9%	94.8%
XBG	94.9%	95.1%
XBG	94.9%	95.1%
XBG	95.0%	95.2%
XBG	95.0%	95.3 %
XBG	94.9%	95.4%
XBG	94.9%	95.4%
XBG	94.9%	94.1%

Table 2. Experimental analysis in Google Colab for Accuracy, Precision, Recall, F1 Score and support for LR and XGB. XGB provides better Accuracy and Precision than LR.

Model	Accuracy	Precision	Recall	F1 Score	Support
LR	79.37	78.00	74.14	76.13	18789
XGB	94.89	94.37	97.12	94.38	18789

Table 3. Statistical Analysis of Mean, Standard deviation and Standard Error of Accuracy and Precision of LR and XGB algorithms. There is a Statistically significant difference in Accuracy and Precision values between the algorithms. XGB had the highest Accuracy (94.89%) and Precision (94.37%) compared with LR Accuracy (79.37%) and Precision (78.00%). But the standard error mean is higher for the precision in XGB in comparison with LR.

GROUP		N	Mean	Std. Deviation	Std.Error Mean
Accuracy	LR	10	79.3700	0.16364	0.05175
	XGB	10	94.8900	0.08756	0.02769
Precision	LR	10	78.0000	0.52705	0.16667
	XGB	10	94.3700	1.48328	0.46905

Table 4. Comparison of the significance level for LR and RF algorithms with value $p < 0.05$. Both LR and RF have a significance level less than 0.05 in terms of accuracy with a 95% confidence interval.

	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
Accuracy	3.020	.019	-264.44 -264.44	18 13.76	.000 .000	-15.52 -15.52	0.058 0.058	-15.643 -15.646	-15.396 -15.393
Precision	6.687	.019	-32.88 -32.88	18 11.23	.000 .000	-16.37 -16.37	0.497 0.497	-17.415 -17.462	-15.324 -15.277

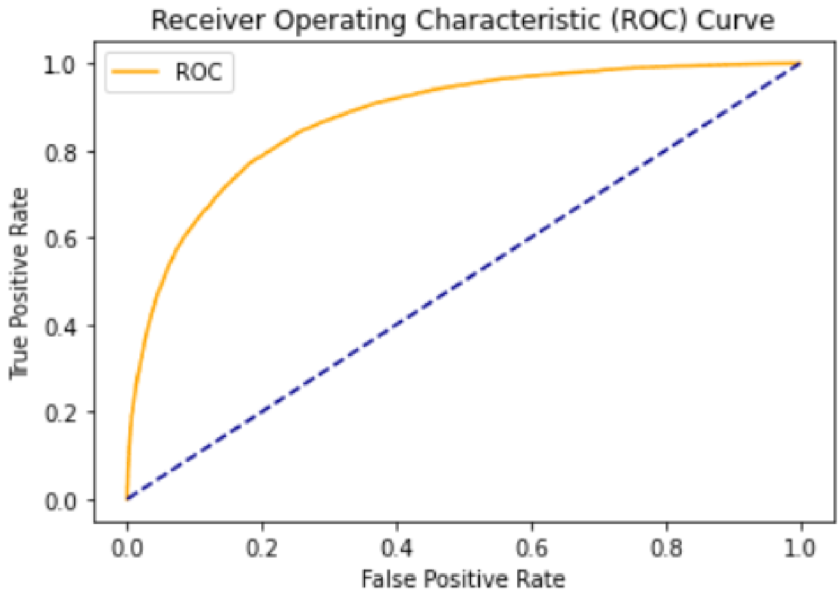

Fig. 1. Receiving Operating characteristic (ROC) Curve for LR

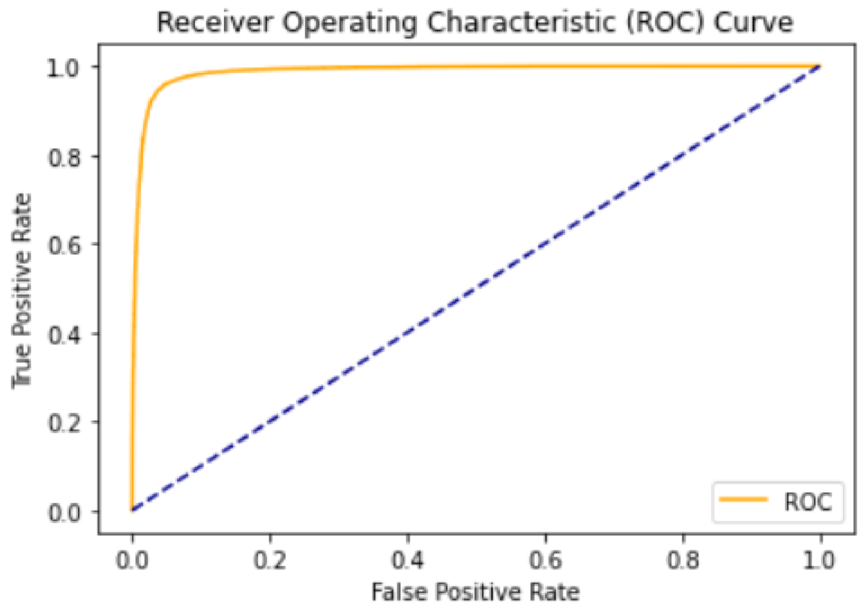

Fig. 2. Receiving Operating characteristic (ROC) Curve for XGB

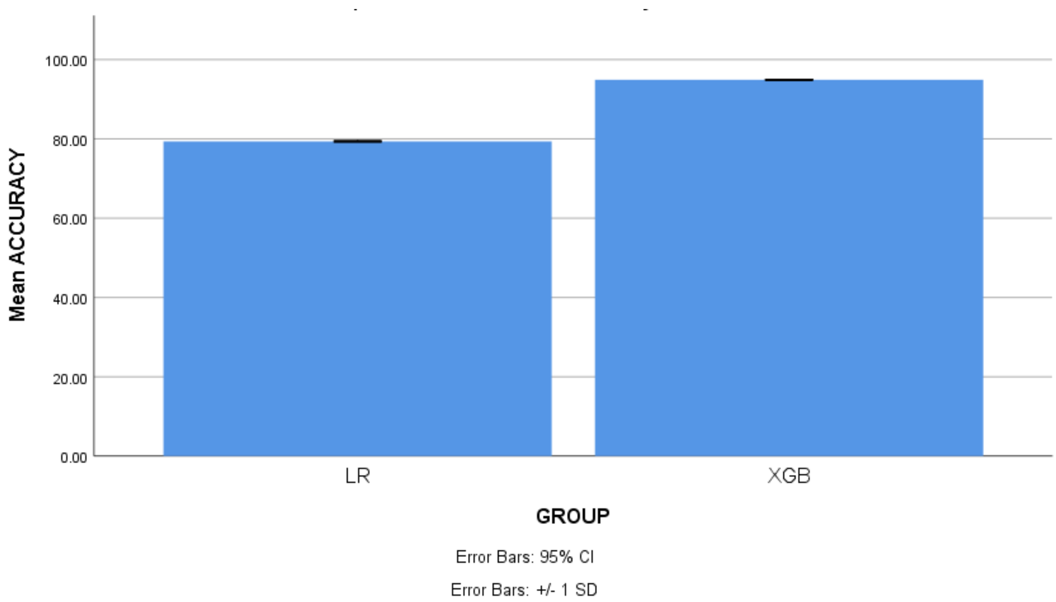

Fig. 3. Bar Chart representing the comparison of mean accuracy of Rainfall prediction using LR and XGB algorithms. XGB produces better accuracy and more consistent results. X-axis: LR vs XGB. Y-axis: Mean Accuracy ± 1 SD.

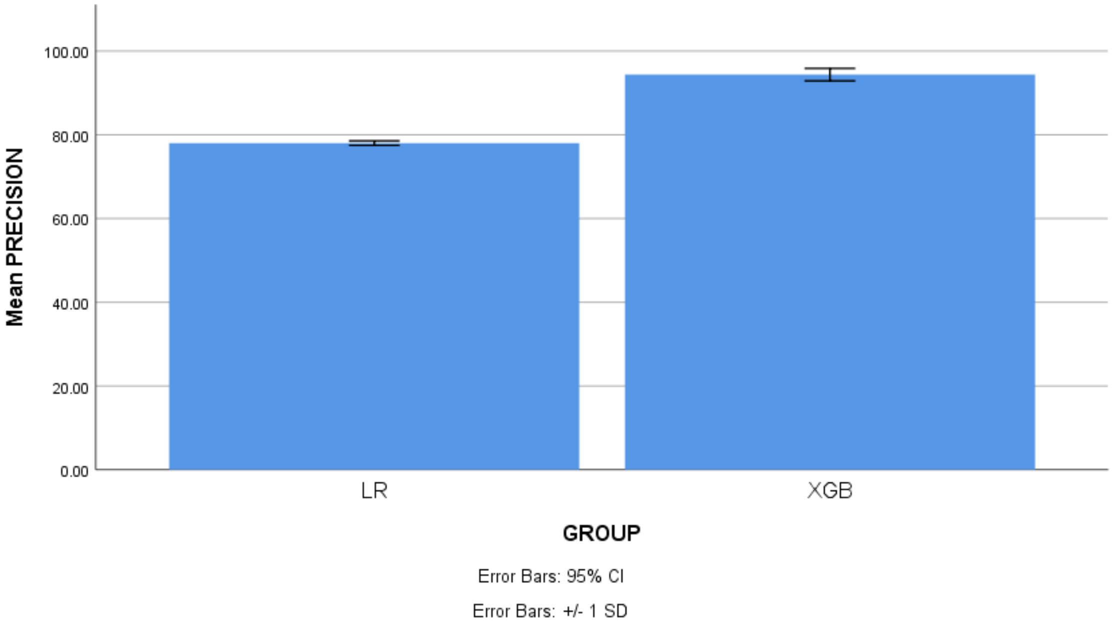

Fig. 4. Bar Chart representing the comparison of mean precision of Rainfall prediction using LR and XGB algorithms. XGB appears to produce better precision and more consistent results. X-axis: LR vs XGB. Y-axis: Mean Precision ± 1 SD.