

IJS GS



ISSN: 2488-9229
FEDERAL UNIVERSITY
GUSAU-NIGERIA

INTERNATIONAL JOURNAL OF SCIENCE FOR GLOBAL SUSTAINABILITY

Rainfall Prediction Models for Katsina State, Nigeria: Machine Learning Approach

Umar Iliyasu, G.N Obunadike and Eli Adama Jiya

Department of Computer Science, Faculty of Computing and Artificial Intelligence,
Federal University Dutsin-Ma, Katsina State, Nigeria.

Corresponding Author's E-mail: umarilivasut@gmail.com

Received on: April, 2023

Revised and Accepted on: May, 2023

Published on: July, 2023

ABSTRACT

Weather patterns and rainfall are essential pieces of information that drive the agricultural sector. For a peasant farmer in Katsina, knowledge of pattern of rainfall is a vital determinant of which crops to plant and when to commence planting. Considering its implications for agriculture, water resource management, and disaster preparedness, this paper developed rainfall prediction models specifically tailored to the unique patterns in Katsina State, Nigeria. The work used Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) with 10 years historical rainfall data obtained from the Nigerian Meteorological Agency Katsina where data were collected from Nigerian Meteorological Agency Katsina State, then preprocessed and subjected to the models, while the evaluation matrices that were used are Precision, Recall, Accuracy, F1-Score, R-Square, Mean Square Error (MSE) and Root Mean Square Error (RMSE). The results of this research indicate that the ANN model outperforms the MLR model with R-Square of ANN equal 0.532 and that of MLR equal to 0.099 and also the precision, recall and f1-score for ANN are 0.666, 0.666 and 0.661 respectively while for MLR they are 0.580, 0.583 and 0.564 respectively. These findings suggest that the ANN model is better at capturing the linear relationship between input variables and rainfall.

Keywords: Artificial Neural Network (ANN), Machine Learning, Multiple Linear Regression (MLR), Rainfall Prediction

1.0 INTRODUCTION

Rainfall prediction is a critical aspect of weather forecasting with significant implications for various sectors, including agriculture, water resource management, and disaster preparedness. Accurate rainfall forecasts can help farmers optimize irrigation schedules, plan crop cultivation, and mitigate the risks associated with droughts or heavy rainfall. Traditionally, rainfall predictions have relied on statistical methods and meteorological models. However, the rapid advancements intelligent Machine Learning Algorithms (MLA) and increasing databases of rainfall dataset have opened up new possibilities for improving the accuracy of rainfall predictions (Ojo and Ogunjo, 2021).

This study focuses on predicting rainfall in Katsina State, located in northern Nigeria. Katsina State has a predominantly agricultural economy, with a

substantial reliance on rainfall for crop production and livestock rearing. Therefore, accurate rainfall forecasting is essential for effective agricultural planning, resource allocation, and water management in the region. By leveraging machine learning algorithms and historical rainfall data, this paper developed a reliable model for predicting unique rainfall patterns of Katsina State. The model has the potential to provide timely and accurate rainfall predictions, enabling farmers, water resource managers, and other stakeholders to make informed decisions. By optimizing agricultural practices, managing water resources efficiently, and enhancing disaster preparedness, the proposed model can contribute to the sustainable development and resilience of Katsina State.

The significance of this study lies in its potential to empower farmers and agricultural stakeholders in Katsina State with valuable information for decision-making. Accurate rainfall predictions can

help farmers optimize irrigation schedules, plan planting and harvesting activities, and adopt appropriate crop management practices. This, in turn, can lead to increased agricultural productivity, reduced losses, and improved food security. Additionally, water resource managers can utilize these predictions to allocate water resources effectively, ensuring sustainable water supply and management in the region. By incorporating machine learning techniques into rainfall prediction, this research also contributes to the broader scientific community's understanding of the application of artificial intelligence in weather forecasting and climate studies (Grace and Suganya, 2020),

The methodology for this research involves several key steps. Initially, historical rainfall data specific to Katsina State will be collected from reliable meteorological sources named Nigerian Meteorological Agency Katsina State, considering monthly mean maximum, minimum temperature and rainfall data for the period of ten (10) years i.e. between 2006 and 2015.

The data will then undergo preprocessing, including data cleaning, outlier removal, and normalization, to ensure its quality and consistency. Feature engineering techniques will be employed to extract relevant meteorological features, such as mean maximum, minimum temperature and geographical information that may influence rainfall patterns in Katsina State.

Two machine learning algorithms will be reviewed and compared to determine which model is best suited for rainfall prediction. The first model will use ANN to predict the rainfall pattern while the second model will use MLR. The chosen model will be trained using historical rainfall data and Mean Squared Error (MSE) will be used to assess their performance as well as Root Mean Squared Error (RMSE).

2.0 RELATED WORK

A research to predict flood occurrences in Nigeria was conducted by Ighile et al., (2022) using logistic regression (LR) and ANN. The result of the research showed that ANN outperformed the LR

model, achieving a prediction rate of 76.4% , compared to 62.5% with LR.

In another research by According to sources like Patil et al., (2021), weather and precipitation are intricate, nonlinear processes that need sophisticated algorithms to obtain a reliable prediction. For predicting rainfall, a number of machine learning models are utilized, including as MLR, ANN, K-means, and Naive Bayes. To locate and forecast rainfall, these models extract, train, and test data sets

The paper also explores the use of Decision Trees for recommending crops based on rainfall prediction. The results show that rainfall can be reasonably predicted, enabling crop recommendations.

Using 33 historical rainfall datasets, Lawal et al., (2021) used three machine learning algorithms—decision tree, LR, and support vector machine (SVM)—to forecast flood incidents in Kebbi State. The study revealed that logistic regression demonstrated the highest performance accuracy among the algorithms tested. However, picking the right attributes play a crucial role in any ML model development. This was highlighted by the researchers, indicating the importance of selecting relevant variables and ensuring their accessibility for accurate predictions.

According to a study conducted by Grace and Suganya, (2020), rainfall prediction is considered one of the most crucial techniques for predicting meteorological conditions in any country. In the case of India, the work predicted rainfall with MLR using weather-related attributes as input data, leading to more accurate rainfall estimates. Compared with other models, the accuracy of the outperforms other existing techniques in the literature.

In a different study by Estévez et al., (2020), dynamic pattern of rainfall for semiarid locations for forecasted using geographic and position-related data as inputs into ANN. The accuracy of the research had between 0.76 to 0.90 as correlation coefficients, indicating a good level of accuracy in

rainfall prediction. It is important to note that, as of now, that the researchers could not find researches in Nigeria that used ML on environmental and location-related attributes to predict rainfall.

Dutta and Gouthaman, (2020) research on models for predicting rainfall based on ANN in India. The researchers compare two rainfall forecast methods to determine the most accurate one. Existing techniques for rainfall prediction struggle with the complexity of available data, particularly when nonlinear trends are present. Machine learning and neural network approaches are explored, and the algorithm delivering the best results is employed for prediction. The research highlights the importance of rainfall estimation for effective rainwater harvesting and emphasizes the role of accurate weather forecasting in reaching a wider audience.

Garg and Pandey, (2019) addressed the increasing concern of detecting and forecasting rainfall in countries lacking advanced technology. The paper compares the findings obtained from different techniques and employs Support Vector Regression (SVR), SVM, and K-Nearest Neighbors (KNN) to estimate future rainfall. Accurate rainfall prediction can have significant implications in areas such as agriculture, health, and water supply.

In the context of estimating precipitation, Shah et al., (2018) emphasized the importance of climatic insights for various industries such as agriculture and research. The paper discusses factual procedures and machine learning approaches used for forecasting and estimating precipitation. The goal is to provide relevant climate and atmospheric information to stakeholders, enabling them to understand changes in precipitation, temperature, and humidity.

Parmar et al., (2017) highlighted the significance of predicting heavy rainfall for meteorological departments as it directly impacts the economy and human well-being. Rainfall forecasting accuracy is crucial for countries like India, heavily reliant on agriculture. Traditional statistical techniques often fail due to the dynamic nature of the atmosphere.

Artificial Neural Networks, with their ability to handle nonlinear rainfall data, offer superior performance. The paper provides a review of different methodologies and algorithms used by researchers for rainfall prediction, aiming to make the study's techniques accessible to non-experts.

3.0 METHODOLOGY

The methodology used by this research work is given as: Data collection, Data pre-processing, Splitting of data into training and testing, machine learning algorithm selection which are Artificial Neural Network and Multiple Linear Regression, Turing the algorithms and evaluate the performance of the algorithms using precision, recall, accuracy, F1-score, Root mean square error, Mean square error and R-square while the tools used are Python Programming, Numpy and Pandas for data manipulation, Matplotlib for data visualisation. These are discussed in detailed.

3.1 Dataset Description

The dataset used in this study contains rainfall measurements for Katsina State from the year 2006 to 2015, obtained from the Nigerian Meteorological Agency Katsina State. The data includes meteorological information collected daily for a period of 10 years in the Katsina State. The target variable to be predicted using machine learning algorithms is to know the days that rain will fall for sure." The dataset also includes features such as Year, Month, Minimum temperature, Maximum temperature, highest rainy days and rainfall, where the input features are Year, Month, Minimum temperature, Maximum temperature and highest rainy days, while the target variable is Rainfall. For the years, it was a numerical value, Month were entered as continuous data (Non-numerical), Minimum and Maximum were entered as float (Numerical) and highest rainy days as integer.(Numerical). After it was entered, the data were split into Training and testing dataset and 80% were used for training, while the remaining 20% for testing.

3.2 Data Pre-processing Missing data

To handle missing data, an analysis was conducted to identify variables with missing values. Instead of discarding a significant amount of data, the variables without data were analyzed by separating them according to cities to minimize data loss. And the missing data were replaced with N/A.

3.2.1 Data normalization

To ensure that all variables take values between 0 and 1, the data was normalized using min-max scaling. This step is important to avoid variables with large magnitudes from dominating the machine learning algorithms.

3.2.2 Detection of outliers

Outliers within the dataset were identified and removed. By eliminating outliers, the model's accuracy is improved, as outliers can adversely affect the performance of the model.

3.3 Feature Selection

In this research work, the dataset includes features such as Year, Month, Minimum temperature, Maximum temperature, highest rainy days, and rainfall. The feature selection technique that was used is Correlation analysis which was used to determine which features have the strongest relationship with the target variable (Rainfall). And with this overfitting was reduced.

3.4 Formulation of prediction model

3.4.1 Multiple Linear Regression

MLR is a statistical method used to determine if a dependent variable and any number of independent variables are related. In this case, multiple linear regression is applied to predict rainfall using independent variables such as Minimum temperature and Maximum temperature. The model is trained using the collected dataset, and predictions are made based on the fitted model.

The multiple linear regression is as follows:

$$y = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + \dots$$

where y is the predicted rainfall value, a_1, a_2, a_3, a_4 are the predictor coefficients and x_1, x_2, x_3, x_4 are predictor variables.

3.4.2 Artificial Neural Networks (ANN)

ANN is ML algorithms that originated as a model of human neurons. In this study, ANN is trained using various features such as Minimum temperature and Maximum temperature to analyze and predict rainfall based on these features. The neural network architecture used can range from simple networks with a single input neuron, a hidden neuron, and an output neuron to more complex structures. The network learns to identify patterns and make predictions based on the training dataset.

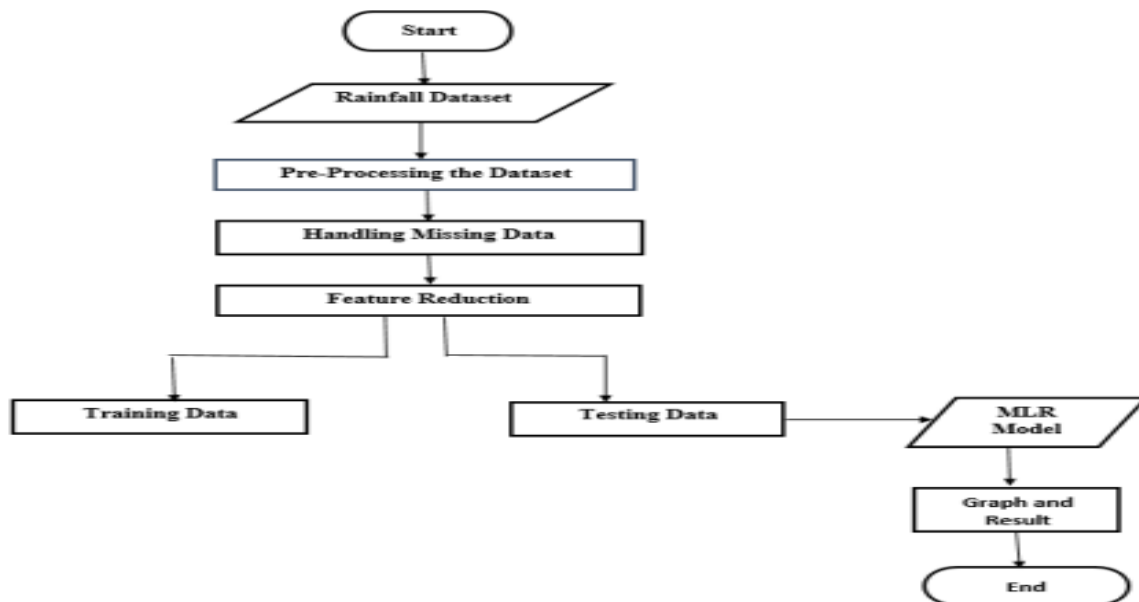


Figure 1: Rainfall Prediction Model

The methodology involves importing the rainfall dataset, performing data preprocessing, handling missing data, reducing features using Principal Component Analysis (PCA), splitting the data into training and testing sets, applying models, calculating evaluation metrics such as MSE and RMSE, plotting scatter plots to compare predicted and testing data, and presenting the results.

4.0 RESULTS AND DISCUSSION

This part presents the findings of the result which involve precision, recall, accuracy, F1-Score, R-Square, Mean Square Error and Root Mean Square Error on MLR model for predicting rainfall and ANN-based rainfall prediction methods. The goal of the research is to identify an algorithm that provides accurate rainfall predictions. From Table 1, the table compares the performance of two models, MLR and ANN, in predicting rainfall. The evaluation metrics used to assess the models are MSE and RMSE.

Table1: Model Comparative table

MODEL	MLR	ANN
MSE	7699.66	3996.38
RMSE	87.74	63.21
R-SQUARE	0.099	0.532

The MSE is use to calculates the average squared difference between the predicted rainfall values and the actual rainfall values. A lower MSE indicates better accuracy, as it means the predicted values are closer to the actual values. In this case, the MLR model has an MSE of 7699.66, while the ANN model has a lower MSE of 3996.38 and more so, the R-Square for ANN is 0.532. This suggests that the ANN model performs better in terms of accuracy, as it has a significantly lower MSE.

MSE measures the square average difference between the projected values of the original unit (rainfall), and the actual values. is represented by this statistic. A lower RMSE denotes higher accuracy, similar to MSE. The MLR model has an RMSE of 87.74, while the ANN model has a higher RMSE of 63.21 and the R-square is 0.099. Again,

this shows that the ANN model outperforms the MLR model in terms of accuracy, as it has a lower RMSE.

From the MSE and RMSE values, it can be concluded that the ANN model provides more accurate predictions of rainfall compared to the MLR model. The ANN model has a lower MSE and RMSE, indicating that it modelled the linear relationship between the input variables and the past rainfall data better than MLR. On the other hand, the MLR model seems to have higher errors and deviations in its predictions, resulting in a higher MSE and RMSE.

Moreso, using Precision, Recall, Accuracy and F1-Square, the result gotten is presented in the Table 2 for each of the algorithms.

Table 2: Evaluation matrix table

Algorithm	Precision	Accuracy	Recall	F1-Square
MLR	0.580	0.583	0.583	0.564
ANN	0.666	0.666	0.666	0.661

Based on the provided results, it appears that the MLR (Multiple Linear Regression) and ANN

(Artificial Neural Network) models achieved different performance metrics for the rainfall

prediction task. For Multiple Linear Regression (MLR), the precision of 0.580 indicates that around 58% of the predicted rainy days by the MLR model were actually true positives. The accuracy of 0.583 suggests that the MLR model classified approximately 58.3% of the instances correctly. The recall of 0.583 means that the MLR model correctly identified around 58.3% of the actual rainy days while the F1-Score of 0.564 is the harmonic mean of precision and recall, providing an overall measure of model performance. This indicates that the MLR model achieved a balanced trade-off between precision and recall.

For Artificial Neural Network (ANN), The precision of 0.666 suggests that approximately 66.6% of the predicted rainy days by the ANN model were true positives. The accuracy of 0.666 indicates that the ANN model classified around 66.6% of the instances correctly. The recall of 0.666 means that the ANN model correctly identified approximately 66.6% of the actual rainy days and F1-Score of 0.661 represents an overall measure of model performance, considering both precision and recall. This indicates that the ANN model achieved a balanced trade-off between precision and recall, similar to the MLR model.

5.0 CONCLUSION

In conclusion, this research highlights the significance of accurate rainfall prediction for various sectors, particularly in Katsina State, Nigeria, where agriculture heavily relies on rainfall. The study demonstrates the potential of machine learning algorithms to enhance rainfall forecasting by leveraging historical data. Specifically, the MLR and ANN models were both assessed for their accuracy in predicting rainfall. Comparing the MLR and ANN models, the ANN model appears to have slightly outperformed the MLR model in terms of precision, accuracy, recall, and F1-Score. The results of this research indicate that the ANN model outperforms the MLR model with R-Square of ANN equal 0.532 and that of MLR equal to 0.099 and also the precision, recall and f1-score for ANN are 0.666, 0.666 and 0.661 respectively while for MLR they are 0.580, 0.583 and 0.564 respectively. For further work, researches can be done for other locations because

this scope of this is Katsina State and moreso, other Machine learning algorithms can be looked at.

REFERENCE

- Dutta, K. and Gouthaman, P. (2020). Rainfall prediction using machine learning and neural network. *International Journal of Recent Technology and Engineering (IJRTE)*, 9(1), 1954-1961. DOI:10.35940/ijrte.A2747.059120
- Estévez, J.A. Bellido-Jiménez, X. Liu, A.P. and García-Marín (2020), Monthly precipitation forecasts using wavelet neural networks models in a semiarid environment, *Water* 12 (7) (2020) 1909. <https://doi.org/10.3390/w12071909>
- Garg, A., and Pandey, H. (2019). Rainfall prediction using machine learning. *International Journal of Innovative Science and Research Technology*, 4(5), 56-58. DOI:10.13140/RG.2.2.26691.04648
- Grace, R. K., and Suganya, B. (2020). Machine learning based rainfall prediction. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 227- 229). IEEE. DOI:10.1109/ICACCS48705.2020.9074233
- Ighile, E.H., Shirakawa, H. and Tanikawa, H. (2022), A study on the application of GIS and machine learning to predict flood areas in Nigeria, *Sustainability* 14 (9) (2022) 5039. <https://doi.org/10.3390/su14095039>
- Lawal, Z. K., Yassin, H. and Zakari, R.Y. (2021), Flood prediction using machine learning models: a case study of Kebbi state Nigeria, in *IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, IEEE, 2021, pp. 1–6.
- Ojo A. S. and Ogunjo, S. T (2022) Machine learning models for prediction of rainfall over Nigeria, *Scientific African* 16, Published by Elsevier B.V. on behalf of African Institute of Mathematical Science. <https://doi.org/10.1016/j.sciaf.2022.e01246>
- Parmar, A., Mistree, K., and Sompura, M. (2017). Machine learning techniques for rainfall prediction: A review. In *International Conference on Innovations in information Embedded and Communication Systems (Vol. 3)*. ISO 9001:2008 Certified Journal

Patil, D., Badarpura, S., Jain, A., Student, U. G., and Gupta G. (2021), A. Rainfall Prediction using Linear approach & Neural Networks and Crop Recommendation based on Decision Tree. DOI:10.17577/IJERTV9IS040314

Shah, U., Garg, S., Sisodiya, N., Dube, N., and Sharma, S. (2018). Rainfall prediction: Accuracy

enhancement using machine learning and forecasting techniques. In 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 776-782). IEEE. DOI:10.1109/PDGC.2018.8745763