International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 2 Feb. 2025, pp: 130-133 www.ijaem.net ISSN: 2395-5252

Rainfall and Climate prediction Using Machine Learning

Suraj B S
Department of Computer Science and Engineering PESCE, Mandya

C Chethana
Department of Computer Science and Engineering PESCE, Mandya

Chetan kumar V
Department of Computer Science and Engineering PESCE, Mandya

Sindhu B S
Department of Computer Engineering and Engineering PESCE, Mandya

Date of Submission: 01-02-2025 Date of Acceptance: 10-02-2025

Abstract— Rainfall Prediction is a tough undertaking due to erratic patterns of rainfall and climate fluctuations all throughout the world. Rainfall forecasts helps to prevent floods and even assists in agriculture for producing crops. Rainfall prediction is aided by machine learning, which uses previously recorded weather data to identify unknown trends. This paper predicts rainfall using a machine learning classification technique. Of the total data, 80% is the training data and 20% is the tested data. Analysis is done on the number of actual and expected forecasts. The accuracy of the employed classification model is found to be 85%. The accuracy of numerous different machine learning techniques is also compared.

Keywords— machine learning; rainfall prediction; random forest classifier

I. INTRODUCTION

One of the essential steps in the weather forecasting process is predicting precipitation because erratic precipitation can result in overflow or dry spells, which helps prevent floods and droughts and safeguard infrastructure and human life. As a result of climate change, floods and other extreme events are predicted to become more frequent and have catastrophic consequences [1]. Moreover, India is an agricultural land, and prediction of precipitation is very essential for growing crops. It advises people to take critical precautions beforehand and protect their crops from rainfall. Weather forecasts can be divided as Short- term forecasts, medium-term forecasts, and Long- term forecasts. Short-term forecasts are related to forecasts for a few hours. Forecasts for the medium term span four to ten days, while those for the long term span . Long-term forecasts span more than ten days, while medium-term forecasts span four to ten days. When it comes to flood forecasting, short- and medium-term forecasts are useful [2].

Due to the unpredictable nature of rainfall, rainfall forecasting is a difficult task. One factor contributing to this uncertainty is climate change, which both threatens communities and supports their growth [3]. Therefore, it is essential to make accurate rainfall forecast predictions. Intelligent prediction models have been found to produce better results than traditional models that rely on statistics.. But the results of these intelligent models can still be improved on basis of the accuracy. In this paper, machine learning approach is used for rainfall prediction. Since, there are many algorithms for machine learning the paper focus on the classification approach that distinguishes the whole data provided, according to the training given by the model and helps in getting a better forecast. In this paper, comparison for better accuracy with other machine learning approaches that are based on supervised learning is also done and it can be concluded that the approach for classification used outperformed all other classifiers depicting the successful implementation of the technique for precipitation prediction.

Machine learning can be used to abstract unknown patterns from previous data to predict the future trend [4]. Weather forecasting based on the historical data is a tough but rewarding endeavor that entails various challenges that must be overcome to reach optimal results[5].Weather- related data is compiled to study various attributes such as

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 2 Feb. 2025, pp: 130-133 www.ijaem.net ISSN: 2395-5252

temperature, pressure, humidity and wind speed. Machine learning techniques incline to forecast upcoming weather conditions using unknown patterns and relationships between elements of historical weather data[6]. Classification is a supervised learning approach in machine learning which can be performed on both structured and unstructured data. It categorizes a given set of data into classes. These classes are known as targets which are used to predict a given class having particular features. It is used to classify new observations on the basis of training data. The unlabeled data is passed to the machine learning model and the model is then trained to classify the labels. The test data checks for the labels and processing is done by the machine learning model. The output of classifier is a probability value between 0 and 1 or Yes and NO.

II. RELATED WORK

Patil [7] demonstrates use of Multiple Linear Regression and Neural networks to predict rainfall. Maqaleh[8] presents two forecasting methods one is grounded on statistical models and the other is established by means of artificial neural network. Box-Jenkins model is used as first method which is employed to calculate time series. A modified artificial neural network model is the second method which uses adaptive slope and momentum parameter to upgrade the weights in back propagation neural network. Rahman[9] uses machine learning techniques such as Decision Tree, Naïve Bayes, K-Nearest Neighbors, and Support Vector Machines their prediction accuracy are bonded by fuzzy logic. Dutta[10] compares machine learning and artificial neural network. The machine learning uses LASSO regression which is more accurate than the artificial neural network approach. Basha [11] signified the Deep Learning Approach by means of the Multilayer Perceptron and Auto - encoder Neural Network for expecting the rainfall and concluded that artificial neural network performed better. Senthil[12] examined algorithms Naive Bayes, K- Nearest Neighbour algorithm, Decision Tree, ANFIS, ARIMA, SLIQ, Neural Network and fuzzy logic. The comparison shows that decision trees and k mean clustering are finest data mining technique for predicting rainfall. Gupta[13] analyzed that neural networks deliver improved effects than any of the other algorithms with 10 neurons in hidden layer. K. R.[14] practices clustering and classification data mining techniques for rainfall prediction. Mishra[15] implements Feed Forward Neural Network (FFNN) for forecasting the rainfall. Nigam [16] showed that Simple Recurrent Neural Network achieves improved result on rainfall prediction although LSTM is better for temperature prediction. Kadam [5] conveys that it is likely to efficiently predict Rainfall by implementing machine learning techniques. The choice algorithm of hinge on the nature of the data confined, attributes measured and the magnitude of the data. Ridwan[17] proposed methodology to predict rainfall in Malaysia by using machine learning algorithms and taking data from 10 stations covering the study area and resolved that prediction using autocorrelation function is better as compared to prediction using projected error. Hussein [18] determines multiple time scales, features, pre-processing techniques, and machine learning algorithms but concluded artificial neural network as best approach for long term prediction. Mehdizaleh[19] implemented hybrid support vector machine model with whale optimization algorithm to study water resource management. The input for support vector machine is preprocessed using random forest, principal component analysis and Pearson's correlation. But the hybrid model performed well as compared to simple support vector machine. Mohammadi[20] formulated forecasting model for streamflow combining adaptive neuro –fuzzy inference system and shuffled frog leaping algorithm. The model was implemented on data taken from two different rivers and it was concluded that the model gave better results than simple adaptive neuro –fuzzy inference system. Olivia[21] investigated the role of meteorological factors in Malaysia on different time horizons. The multilayer perceptron with artificial neural network was implemented giving better outcome in some horizons and the adaptive neuro-fuzzy inference system gave enhanced results at some other horizons depending on the input factors provided in different horizons.

III. PROPOSED WORK

This research uses the Australian weather dataset [22] to estimate rainfall. This dataset was chosen since it takes into account a wide range of factors, including pressure, humidity, wind direction, rainfall, minimum and maximum temperatures, and more. Each of these factors contributes to a more accurate weather forecast.. First of all, the dataset is imported to google collab and then data is preprocessed for feature selection. It helps in removing the invalid and missing data. Normalization is part of data preprocessing that aids in fixing the range of independent variables for better feature approximation. After data preprocessing model training is done, for which data is split into training set and test set. Fig. 2 shows the rainfall prediction model used which consist of 80% train set and 20% test set of the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 2 Feb. 2025, pp: 130-133 www.ijaem.net ISSN: 2395-5252

whole dataset. The train set is one on basis of which model training is done and the test set is one which helps in making predictions. The Random Forest classifier is used for the classification, and it produces YES and NO results for the rain forecast, which is then compared to the actual rain conditions in the research area. The overall model accuracy provided by the suggested method is 85%.

Rainfall Prediction Model

A. Random Forest Classifier

Random Forest classifier uses the decision trees as building blocks. It chooses a random sample from the given dataset and constructs a decision tree for each sample. The predictions made for each sample by the decision tree is result of the minimum correlation between each other. Random forest classifier gives better accuracy results when the dataset is large. Moreover, the training time taken is less in Random Forest Classifier as compared to other algorithms. The code used to implement Random Forest Classifier in Python is shown below. While implementing the Random Forest classifier in the proposed model the value of n estimators is taken to be 150 because greater number of trees brings about better accuracy and random state is set to be 0.

IV. RESULTS AND DISCUSSIONS

The implementation of the Rainfall prediction model assists in finding the variations for actual forecast and prediction forecast the outcome achieved from the model. The count of actual and prediction rainfall is done considering four cases. Case1 considers both actual and prediction values as false so that is the case when there is no rainfall witnessed at all. Case 2 reflects actual rainfall as false and prediction forecast as true. It depicts that in reality there was no rainfall observed whereas the weather forecast showed signs of rain. Case 3 depicts actual forecast as true and prediction as false. This represents that it rained those days when there was no estimate of rain in prediction. Case 4 takes both actual and prediction result as true. So here the observation matches completely with the actual weather conditions for rainfall.

Table 1 elaborates the chances for actual and prediction forecast depicting all four cases. It can be realized that 74% chances match having no rain tomorrow for both actual and prediction. Only 3% chances exist when there is no actual rain though the prediction for rain was confirmed. Out of all, 11% chances occurred when actually it was rainy but there was no prediction of rain tomorrow. However, 12 % chances were when actual and prediction for rain was true for next day as rainy day. So, overall 86% chances of the total chances exhibited correct results for actual versus forecast outcomes for rain tomorrow. Thus, the model imposed has given its best results considering all major parameters that are accountable for the rainfall and aids in making predictions using which people can be made aware of the weather deviations and measures can be taken to handle them.

Chances	Actual Rain Tomorrow	Predicted Rain Tomorrow
74%	NO	NO
3%	NO	YES
11%	YES	NO
12%	YES	YES

A. Comparison of results

On the basis of the machine learning approach, comparison is made between different classifiers that are based on supervised learning such as Naive Bayes, support vector machine, K nearest neighbor and Random forest in terms of accuracy as this is one parameter which needs to be attained for determining the improvement in results and it can be concluded from Fig. 4 that Random Forest Classifier has the highest accuracy of 85% followed by K nearest neighbor (KNN) with 80.7% accuracy and then Naive Bayes with 78% accuracy. The lowest accuracy is 51% for support vector machine (SVM).

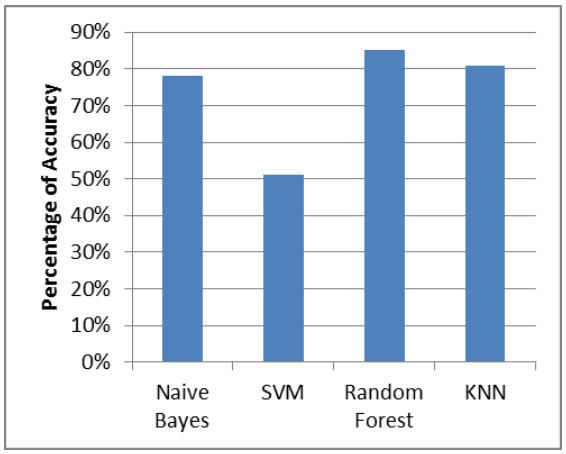

Fig. 4 Comparison of Accuracy

V. CONCLUSION

Predicting rainfall is crucial for food production planning, flood prevention, and efficient water resource management. The machine learning classification method used in this work helps to improve rainfall prediction forecasts because most

rainfall datasets have irregular patterns. Since the Random Forest classifier can handle missing values and a large dataset with less training time than other algorithms, it is used here.. At 85%, it has produced the most accurate result. When compared to alternative approaches, the suggested framework produced the best results, demonstrating improvement in the outcomes. The technique will be justified in future work using a wider variety of datasets.

REFERENCES

[1]. S. Aftab, M. Ahmad, N. Hameed, M. S. Bashir, I. Ali, and Z. Nawaz, "Rainfall prediction using data mining techniques: A systematic literature review," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 5, pp. 143–150, 2018, doi: 10.14569/IJACSA.2018.090518.

[2]. A. Y. Barrera-animas et al., "Machine Learning with Applications Rainfall prediction : A comparative analysis of modern machine learning algorithms for time-series forecasting," Mach. Learn. with Appl., vol. 7, no. October 2021, p. 100204, 2022, doi: 10.1016/j.mlwa.2021.100204.

[3]. M. A. Nayak and S. Ghosh, "Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier," Theor. Appl. Climatol., vol. 114, no. 3–4, pp. 583–603, 2013, doi: 10.1007/s00704-013-0867-3.

[4]. D. Salgaonkar and B. Chaudhari, "Survey on Rainfall Prediction System," no. 09, pp. 9–11, 2021.

[5]. M. Hasan et al., "ScienceDirect Search and rescue operation in flooded areas : A survey on emerging sensor networking-enabled IoT-oriented technologies and applications," Cogn. Syst. Res., vol. 67, pp. 104–123, 2021, doi: 10.1016/j.cogsys.2020.12.008.

[6]. K. Dutta and G. P*, "Rainfall Prediction using Machine Learning and Neural Network," Int. J. Recent Technol. Eng., vol. 9, no. 1, pp. 1954–1961, 2020, doi: 10.35940/ijrte.a2747.059120.

[7]. C. Z. Basha, N. Bhavana, P. Bhavya, and V. Sowmya, "Rainfall Prediction using Machine Learning Deep Learning Techniques," Proc. Int. Conf. Electron. Sustain. Commun. Syst. ICESC 2020, no. Icesc, pp. 92–97, 2020, doi: 10.1109/ICESC48915.2020.9155896.

[8]. K. R. B and G. A, "Rainfall Prediction Using Data Mining Techniques - A Survey," vol. 83, no. 8, pp. 23–30, 2013, doi: 10.5121/csit.2013.3903.

[9]. K. R. B and G. A, "Rainfall Prediction Using Data Mining Techniques - A Survey," vol. 83, no. 8, pp. 23–30, 2013, doi: 10.5121/csit.2013.3903.

[10]. N. Mishra, H. K. Soni, S. Sharma, and A. K. Upadhyay, "A comprehensive survey of data mining techniques on time series data for rainfall prediction," J. ICT Res. Appl., vol. 11, no. 2, pp. 167–183, 2017, doi: 10.5614/itbj.ict.res.appl.2017.11.2.4.

[11]. J. Ranjani, V. K. G. Kalaiselvi, A. Sheela, D. Deepika Sree, and G. Janaki, "Crop Yield Prediction Using Machine Learning Algorithm," Proc. 2021 4th Int. Conf. Comput. Commun. Technol. ICCCT 2021, pp. 611–616, 2021, doi: 10.1109/ICCCT53315.2021.9711853.

[12]. W. M. Ridwan, M. Sapitang, A. Aziz, K. Faizal, A. Najah, and A. El-shafie, "Rainfall forecasting model using machine learning methods : Case study," Ain Shams Eng. J., vol. 12, no. 2, pp. 1651–1663, 2021, doi: 10.1016/j.asej.2020.09.011.