Environmental Industry Letters
https://journals.tultech.eu/index.php/eil
ISSN: 2806-2965
Volume 2, Issue 2
DOI: https://doi.org/10.15157/eil.2024.2.2.44-56
Received: 05.06.2024; Revised: 29.08.2024 ; Accepted: 13.09.2024

Research Article

A Comparative Analysis of Rainfall-Prediction Using Optimized Machine Learning Algorithms

S Sindhura[1*], P Dedeepya[2], N Sampreet Chowdary[2], Katragadda Megha Shyam[3], Balamurali Krishna Thati[4]

[1] Department of Computer Science and Engineering, NRI Institute of Technology, Agiripalli, India
[2] Department of Computer Science and Engineering, PVP Siddhartha Institute of Technology, Vijayawada, India.
[3] Department of CSE, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Bengaluru, India
[4] Department of CSE, Dhanekula Institute of Engineering & Technology, Ganguru, Vijayawada., India
***ssindhurapraveen@gmail.com**

Abstract

The difficult challenge of predicting rainfall is brought on by the daily observations of erratic rainfall patterns and climatic fluctuations. Predicting when the rain will fall can help avoid floods and even aid in crop growth in agriculture. Timely and precise predictions can prevent loss of life and assets. The ability to forecast the amount of rainfall requires an understanding of weather-related elements such as pressure, humidity, wind speed, latitude, longitude, and precipitable water with varying x and y-axis parameters. The research in this study involves using fundamental machine learning techniques to create weather forecasting models that use the day's meteorological data to predict whether or not it will rain tomorrow. By utilizing previously identified trends from historical meteorological data, machine learning helps forecast rainfall. We are using a classification model in our supervised data model, and the techniques utilized to forecast the amount of rain include random forest, KNN, decision tree, and logistic regression. Using machine learning algorithms to examine past weather data and find patterns that can be applied to forecast future rainfall patterns is the suggested technique for rainfall prediction. A more accurate weather forecast is made possible by all of the aforementioned factors. We will handle the information that aids in eliminating erroneous and incomplete data. As a component of data preparation, normalization helps to improve feature approximation by adjusting the range of independent variables. Model training is carried out following data preparation, during which data is divided into training and test sets. The test set aids in prediction-making, while the training set serves as the foundation for model training.

Keywords: Machine learning, Rainfall prediction, Supervised learning, Weather forecasting

INTRODUCTION

The management of water levels and agriculture may be significantly impacted by the unpredictable nature of rainfall brought on by weather changes. Therefore, to lessen the possible negative effects of rainfall variability, accurate rainfall prediction models must be developed. A promising solution to this problem is to combine machine learning with environmental data analysis and the analysis of significant atmospheric features. Researchers have made great strides toward creating more accurate and dependable rainfall prediction models, which are crucial for several applications, such as water resource management, agricultural decision-making, and smart city planning. These models are made possible by utilizing historical weather data and sophisticated machine-learning algorithms.

Several algorithms can be used in the process of forecasting rainfall when predicting rainfall using machine learning. By uncovering latent patterns in past weather records, machine learning techniques can be utilized to forecast rainfall. Various machine learning models have been utilized for this purpose, including Gaussian naïve Bayes, RF, adaptive boosting, gradient boosting, MLP, linear regression, XGBoost, and further ones. To enhance the precision of rainfall forecasting, these algorithms are assessed and contrasted with real-time meteorological datasets and environmental data. Planning for agriculture and efficient water resource management have benefited greatly from the use of machine learning, which has demonstrated promising results in developing the accuracy of rainfall prediction systems. To train classification models, the method usually involves collecting previous weather information. Based on the input data, these models try to forecast if it will rain or not within a given timeframe. Neural networks, SVM, decision trees, RF, and other machine learning models can be used for this purpose. To determine whether future weather rainfall is there or not, they examine previous weather patterns and related variables.

K-Nearest Neighbors (KNN) is a simple technique, it may be resource-intensive, especially when dealing with large datasets, which limits its use for real-time forecasting. Its accuracy is made worse by its noise and irrelevant feature sensitivity. Although Random Forests are good at processing a wide range of data, they can overfit noisy data, which could affect the accuracy of their predictions. The interpretability of their collective nature is compromised, making it difficult to understand the critical decision-making process involved in rainfall forecasting. Although decision trees are understandable, their relevance to new rainfall patterns is limited due to their tendency towards overfitting. Practical rainfall forecasts are further complicated by the fact that Multilayer Perceptron (MLP) networks, which are excellent at capturing complex connections, are data-intensive and may have difficulties with feature scaling, particularly when faced with insufficient or uneven feature scaling.

The encouragement of large datasets and noise present challenges for vector machines (SVM), which reduces their accuracy in rainfall forecasting. Likewise, the dependence of logistic regression on linearity affects its ability to understand complex correlations among rainfall data, which may lead to less accurate forecasts. The reliability of Naive Bayes is reduced by its assumption of important independence of characteristics, which may not hold in the case of coupled rainfall features. Furthermore, while AdaBoost, Gradient Boosting, and XGBoost have strong prediction performance, they also require careful hyperparameter adjustment or are prone to overfitting noisy data. AdaBoost's overall performance is affected by its sensitivity to noise and outliers, which creates a major obstacle to its capacity to forecast rainfall patterns with accuracy.

The proposed methodology involves Gathering previous weather data, preprocessing it, using tree-based models to identify the most important features for rainfall prediction, and training a variety of classification models—comprising Adaboost, XGBoost, Gradient Boosting, Random Forest, Decision Trees, MLP, Support Vector Machine (SVM), Logistic Regression, and Naive Bayes. —are all part of the suggested methodology. The models will be assessed using perception recall, accuracy metrics, and F1 score. Tests have demonstrated that, when compared to other models, XGBoost has the greatest accuracy. This approach aims to optimize performance and produce accurate rainfall classification estimates by adjusting the model's hyperparameters. This approach may be used to create a reliable machine-learning model for rainfall prediction, considering the information from the sources previously stated. Decision-makers in the water management, agricultural, and other sectors that rely on precise rainfall forecasts for efficient planning and resource management will find great insights from this model.

LITERATURE REVIEW

The challenge is creating a machine learning model for rainfall prediction, an important environmental element influencing many industries like urban planning, agriculture, and water resource management. The objective is to develop a prediction model that, using past meteorological data, can anticipate patterns of rainfall. To reflect the variety in rainfall patterns, a sizable dataset covering several geographical locations and temporal scales should be used to train the model. Finding relevant patterns and correlations in the data that can be used to support precise predictions is the difficult part. The goal is to deliver accurate and timely rainfall forecasts so that interested parties may make well-informed decisions and lessen the possible effects of extreme weather occurrences linked to precipitation. The machine learning models will have a big impact on disaster planning and sustainable resource management in areas where rainfall patterns can change dramatically.

In the research [1], the author stated that the study on the incorporation of algorithms into climate finance decision support is presented with an emphasis on the effects of rainfall on Italian vineyards. The goal of the research is to forecast the intensity of rainfall every quarter using machine learning approaches. Three algorithms are compared in terms of performance: Random Forest and Neural Network. Preprocessing of the dataset for the experiment was done after it was obtained from the Piedmont area of Italy's agrometeorological office. The outcomes show that the Random Forest method is superior to the other two algorithms in terms of prediction accuracy and explainability, indicating that it can better anticipate rainfall by utilizing specific environmental factors that are important to the prediction. This work adds to the corpus of research on climate financing by illuminating how machine learning may help decision-makers manage climate risks in the food chain—specifically, in the Italian wine sector. In the paper [2], the author explains how to use a multiple linear model to estimate PRCP based on meteorological variables including temperature, wind speed, and dew point. The National Climatic Center provided the data for the study, and the model was developed using Python code that made use of the Pytorch package. The model's effectiveness was ascertained by comparing the average mean square error of the training and test sets of data. The results showed that the average mean square error improved by 85% during the testing phase when the same amount of data was used for both the training and test phases. However, it dropped to 59% when more data was gathered for the test phase than for the training phase.

The article [3] presented a rain prediction model based on machine learning techniques. The study aims to anticipate rain using modern techniques to increase the forecasts' accuracy and dependability. After obtaining 90% accuracy, 0.90 F1 score, and 0.90 area Under the curve, the Random Forest classifier was found to have the greatest performance, according to the data. Considering climate change and its effects on extreme weather and disasters, the study's finding highlights the importance of the rain prediction model and its ability to provide precise rain forecasting.

In paper [4], the author described how ensemble techniques, notably Random Forest and NARX neural network algorithms, were used to create a rainfall forecasting system. The suggested method seeks to increase district-level rainfall forecasting accuracy while requiring fewer calculations. The document highlights the importance of rainfall forecasting in several fields and endeavors, including urban planning, agriculture, hydroelectric power production, water resource management, disaster preparedness, and climate research. Along with data analysis, feature extraction, model training, real-time data integration, ensemble approaches, uncertainty estimates, model improvement, and early warning systems, it also covers the role of machine learning in rainfall prediction.

In the paper [5], the authors explored the application of machine learning algorithms for rainfall prediction in Ghana's several ecological zones. The study employed five classification algorithms (K-Nearest Neighbor, Decision Tree, Random Forest, Multilayer Perceptron, and Extreme Gradient Boosting) to develop models for rainfall prediction. The effectiveness of these algorithms was evaluated with training/testing data ratios and a variety of metrics. The results showed that random forest, extreme gradient boosting, and multilayer perceptron performed well across all zones, whereas K-Nearest Neighbor fared the lowest. The study concluded that machine learning techniques, particularly Random Forest, may be used to predict rainfall in Ghana's various ecological zones. The paper [6] looked at the analysis of rainfall data from various locations using random forest and support vector machine learning algorithms. The work increases the precision and accuracy of rainfall forecasts by using feature extraction. The methodology used in the study includes data collection, pre-processing, and the application of machine learning techniques such as random forest, logistic regression, decision trees, and neural networks. The results of the study demonstrate how accurate the rainfall predictions made by the algorithms were.

In the paper [7], the authors suggested a strategy for predicting rainfall based on past meteorological data by applying machine learning techniques. The system's goal is to deliver precise weather forecasts so that farmers may make well-informed choices regarding farming tasks like planting and irrigating. SoftMax logistic regression, data cleaning, and prediction system assessment are the main topics of the work. The 'underground' website provided the study with their dataset, which included historical weather information from eleven cities in Myanmar. To forecast rainfall categories based on variables including temperature, humidity, wind speed, and precipitation, the system uses the SoftMax logistic regression algorithm. The testing findings show that the system can use SoftMax to predict rainfall with an accuracy of 83%. According to the study's findings, farmers and other stakeholders may find the suggested rainfall forecast system helpful in making decisions about weather-dependent operations. The study offers a thorough analysis of the methodology, system architecture, and assessment metrics utilized to gauge the rainfall forecast system's effectiveness.

In the study [8], the authors showed how the system for predicting rainfall using machine learning techniques—more especially, the SoftMax logistic regression method—can be utilized based on historical meteorological data. The system's goal is to deliver precise weather forecasts so that farmers may make well-informed choices regarding farming tasks like planting and irrigating. In research [9], SoftMax logistic regression, data cleaning, and prediction system assessment are the main topics of the study. The 'underground' website provided the study with their dataset, which included historical weather information from eleven cities in Myanmar. To forecast rainfall categories based on variables including temperature, humidity, wind speed, and precipitation, the system uses the SoftMax logistic regression algorithm. Also, in the study [10], which is based on the SoftMax logistic regression algorithm, the experimental findings show that the system can forecast rainfall with an accuracy of 83%. The study offers a thorough analysis of the methodology, system architecture, and assessment metrics utilized to gauge the rainfall forecast system's effectiveness.

In the paper [11], based on ground observation data from the China Meteorological Administration (CMA), the rainfall prediction model has been investigated, and it forecasts rainfall using data mining techniques. This research compares Naïve Bayes, Support Vector Machine (SVM), and Back Propagation Neural Network (BPNN), three different approaches. The prediction model is recommended to be evaluated using two criteria: RR (rainfall data rate) and RO (overall rate). The research evaluates the model's performance using many observation station factors, including latitude, longitude, altitude, average temperature, and the chance of rainfall in the past. The experimental results show that BPNN performs better in average RO and RR as compared to NB and SVM. According to the study's findings, the suggested rainfall prediction model can be a helpful tool for forecasting rainfall in diverse areas and be used in a variety of contexts, such as disaster relief, agriculture, and water resource management [12]. In paper [13], the authors describe how research based on real-time rainfall prediction for the

Hyderabad region using a machine learning approach describes a brand-new hybrid machine learning method for forecasting the region's monsoon rainfall. The study employs a combination of regression and classification models, such as several regression models and the Random Forest, XGB, and K neighbors' classifiers, to increase prediction accuracy [14]. A weighted average of the Random Forest Classifier and MLP Regressor produced the best results. With the recommended method, a mean absolute error of 1.063540 and a practical accuracy of 0.92537 are obtained. To prove the superiority of the suggested weighted hybrid algorithm, the study also evaluates the performance of other hybrid models and individual algorithms using various performance indicators [15], [16]. The study's results provide encouraging outcomes in terms of accuracy and performance measures, indicating the approach's potential value in assisting agriculture sector stakeholders in making decisions based on precise rainfall forecasts for rainfall prediction [17], [18]. The authors evaluate the performance of these algorithms using different metrics such as accuracy, precision, recall, and F1 score. They conclude that the random forest algorithm performs the best for rainfall prediction [19].

METHODOLOGY

Data Description

The dataset consists of a total of 54245 instances and 23 characteristics. The Rain Today and Rain Tomorrow are the two main labels. The training and test data in the dataset were divided 80:20.80 percent of the training and set Test outcomes: 20%

Data Preprocessing

To preprocess the data, we used the read_csv() method to read a CSV file called "weatherAUS.csv" into a Pandas Data Frame. Next, by utilizing the head() function to show the data frame's first few rows, the code does some preliminary data exploration. To eliminate any leading or following white spaces, it additionally renames the data frame's columns using the rename() method. Using the unique() method, the code determines how many unique values are there in each Data Frame column. For additional analysis, this might assist in identifying categorical variables that could require encoding or transformation. Lastly, we use the describe() function to display the data frame's summary statistics. This offers information on the distribution and range of values in the data, which may be useful in locating abnormalities or outliers that require attention.

Algorithms based on Machine Learning and evolution

The output label and all the selected features are integrated into a new dataset after the feature selection procedure. To ensure that the model uses the train data for training and the test data to confirm its performance when predicting results, the new dataset must be split into testing and training segments at a 20:80 ratio. We are utilizing nine different machine learning approaches to compute the values given the above data.

1. Support Vector Machines: A Support Vector Machine (SVM) is a supervised machine

learning algorithm used for classification and regression analysis. SVMs aim to maximize the margin distance between data points of different types by determining the optimal line or decision boundary. SVMs can achieve greater accuracy than simpler kernel functions and are especially effective at addressing binary classification issues; nevertheless, the approach requires more data and computing time to train. After being trained, SVMs can identify which side of the decision boundary new, unknown data points lie on to categorize them.

2.

$$\text{v } K(x_1, x_2) = \Theta(x_1)^T . x_2 \qquad (1)$$

which is used to convert the data points (m1, m2) in each input space to a new feature space using the mapping function Θ(x), where the data is parted by a hyperplane. The linear kernel function is considered for the model and is given by:

$$K(x_1, x_2) = {x_1}^T . x_2 \qquad (2)$$

3. Multi-layer Perceptron Neural Networks (MLPNN): A Multilayer Perceptron (MLP) is a type of artificial neural network that consists of fully connected neurons organized in at least three layers: input, hidden, and output layers.MLPs are appropriate for both classification and regression problems because they can recognize complex patterns and correlations in data. The nonlinear activation functions, such as sigmoid or ReLU, are utilized to convert the incoming data into a format that is readable by the network. Multiple hidden layers in MLPs allow them to identify more complex patterns in the data and improve their performance across a range of applications.

$$f_i = \sum_{j=1}^{m} \omega_{ji} X_j + \omega_{o_i} \qquad (3)$$

4. K-Nearest Neighbors: The k-Nearest Neighbors algorithm is a non-parametric, supervised learning method used for classification and regression. It functions according to the idea that data points in the feature space that have similar attributes are near to one another. The output for categorization is decided by the k nearest neighbors casting most votes. The method uses proximity to classify or anticipate how to group a single data point. To do this, it needs to know the distance metrics and the value of k, which is the number of nearest neighbors to take into consideration. It needs careful consideration when choosing the distance metric and the amount of k, and it can be relatively costly, particularly for big datasets.
5. Random Forest: Random Forest is a machine learning algorithm used for both classification and regression tasks. To make predictions, it builds many decision trees during training and collects the results. This lowers the danger of overfitting by using a random selection of features and data for each tree. It produces accurate predictions by

combining the output of these several trees, which makes it useful for managing noisy data and big datasets with lots of variables. The combination of this approach's primary benefits—scalability and stability—has made it popular in a wide range of applications.

6. Decision Tree: Decision Trees are a fundamental machine-learning technique used for both classification and regression tasks. They arrange the data into a tree-like structure, with each leaf node holding the ultimate judgment or prediction, each internal node representing a characteristic or attribute, and each branch corresponding to a choice or result depending on that feature. To construct homogenous subsets concerning the objective variable, the dataset is recursively partitioned based on characteristics, to maximize information gain or minimize impurity at each stage. This framework offers a clear set of guidelines for decision-making and permits intuitive interpretation.
7. Naive Bayes: Naive Bayes is a probabilistic machine learning algorithm based on Bayes' theorem with the "naive" assumption of independence among features. By figuring out the conditional probability of each class given the input attributes, it may predict the possibility that a given sample will belong to a specific class. Because of its ease of computing and efficacy with small datasets, Naive Bayes maintains its efficiency despite its simplifying assumption of feature independence, particularly in text classification and other tasks involving high-dimensional data.

RESULTS AND DISCUSSION

Experimental setup and Dataset

To predict rainfall, it is necessary to build up a controlled environment in which different machine-learning models are tested and assessed using historical meteorological data. The procedure requires knowledge of variables such as temperature, humidity, pressure, and rainfall, all of which are present in the dataset. We build up the experimental setup using the Panda's package in Python, which enables data cleaning, preprocessing, and analysis. To eliminate missing values, discrepancies, and inconsistencies, the procedure starts with reading the dataset. By doing this, you can be confident that the data is consistent and clean for additional analysis. For training and validation, the data is then divided into features (X) and target variables (Y).

To determine which characteristics are most crucial for the prediction job, feature selection approaches like Recursive Feature Elimination (RFE) are used. Numerous machine learning models, including XGBoost, Random Forest, Extra Trees, Adaptive Boosting, Gradient Boosting, Multilayer Perceptron, Gaussian naïve Bayes, linear regression, and others, are evaluated using the preprocessed dataset.

The performance of these models is assessed using appropriate evaluation measures, such as mean squared error and R-squared score. The best-performing model is identified based on these evaluation metrics, and its performance is then improved by adjusting the model using hyperparameter optimization techniques.

There are 23 features and 145,460 instances in the dataset used in the "Prediction of

Rainfall Using Machine Learning" study. The dataset provides a thorough foundation for the machine learning-based rainfall prediction analysis, encompassing a wide range of meteorological factors over many years.

Evaluation parameters and formulations:

Assessing a predictive model's performance in the context of rainfall prediction is crucial to guaranteeing its accuracy in predicting meteorological conditions. This statistic measures the model's total accuracy in predicting the occurrence of rainfall. It gives a clue as to how effectively the model simulates both rainy and dry situations. However, data distribution imbalances may constrain the usefulness of accuracy as a single metric in the field of rainfall prediction, as non-rainy occurrences may significantly outnumber rainy ones. Recall and accuracy become essential at this point. In Figure 1, the precision of rainfall prediction measures how well the model forecasts the presence of rain, highlighting the need to reduce false positives to avoid needless alarms. Conversely, recall gauges how well the algorithm detects wet situations, reducing false negatives and guaranteeing that genuine rainfall occurrences are not overlooked.

$$Precision = \frac{TP}{FP+TP} \qquad (4)$$

$$Accuracy = \frac{TN+TP}{TP} + (TP+FP+FN+TN) \qquad (5)$$

$$Precision = \frac{TP}{FN+TP} \qquad (6)$$

$$Precision = \frac{2TP}{2TP+FN+FP} \qquad (7)$$

Comparison Methods:

We use algorithms to estimate whether it will rain tomorrow based on a variety of weather-related factors. First, the required libraries are imported: sci-kit-learn for machine learning, Label Encoder for encoding categorical variables for data processing. We then extract the rows from the dataset that have missing values. The label encoder is utilized to convert categorical variables, namely 'Rain Today' and 'Rain Tomorrow,' into numerical values.

After this the obtained accuracy for Decision Tree is 0.79, Random Forest is 0.84, KNN is 0.83, MLP Classifier is 0.84. SVM is 0.78.

Figure 2 shows a Heatmap with a Data Frame’s (df) correlation matrix for a subset of its continuous columns. Using Matplotlib, the plot's size is set in the first line.

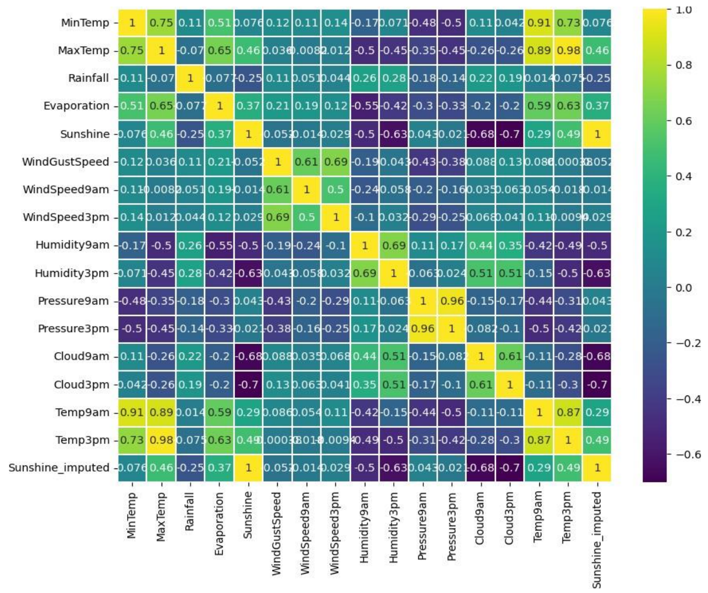

Figure 1. Heat Map

The count plots in Figure 2 show a comparison inside a Data Frame (DF) for two columns, "Cloud9am_imputed" and "Cloud3pm_imputed." The total figure dimensions are set to 10 widths by 4 height units. The following plotting commands are split amongst the two subplots once the code generates a subplot grid with one row and two columns. A Seaborn count plot is created for the "Cloud9am_imputed" column in the left subplot (1, 2, 1), showing the distribution of occurrences for each distinct value. Like this, a count plot for the "Cloud3pm_imputed" column may be found in the right subplot (1, 2, 2). Here we identified the count plot for the two columns; they are cloud3pm and cloud.

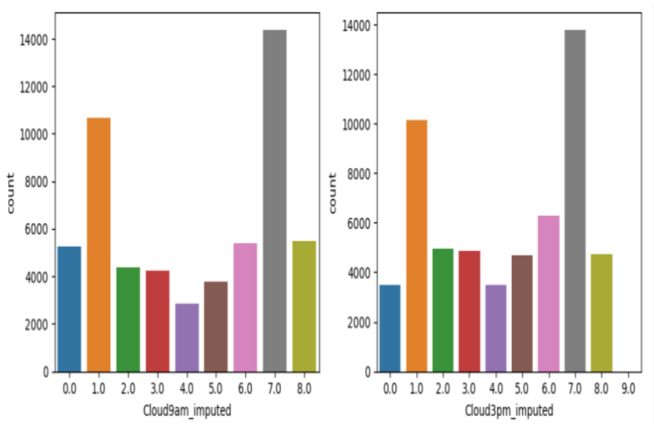

Figure 2. Count plot

For various algorithms, the Confusion Matrix is obtained in Figure 3 which represents the Random Forest classifier, KNN, MLPNN, and SVM classifiers.

Confusion matrixes for the various algorithms obtained:

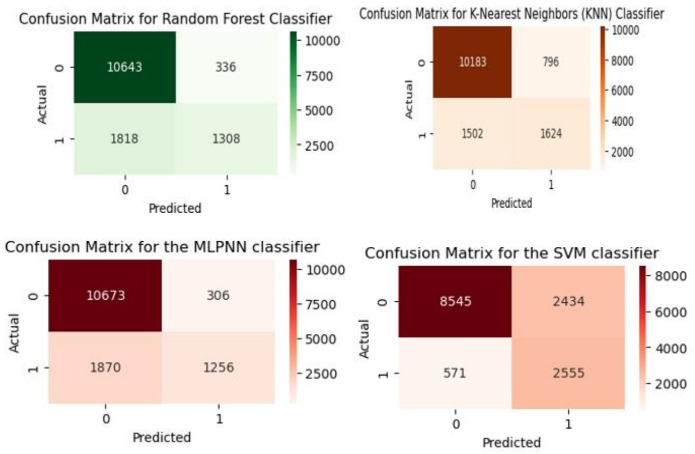

Figure 3. Confusion Matrix for the Models with Pso

CONCLUSION

The use of machine learning to predict rainfall has proven to be a useful technique, demonstrating the capacity to examine meteorological variables and past weather data to produce forecasts that are precise and timely. By utilizing sophisticated algorithms like random forest, support vector machines, and neural networks, it is now possible to identify intricate patterns in the data, which leads to predictions that are more accurate than those made using more conventional techniques. Ongoing development, cooperation with meteorological specialists, and the incorporation of real-time data will further enhance the efficiency of these models, benefiting industries such as agriculture, water resource management, and disaster preparedness. With ramifications for several applications, the effective use of machine learning in rainfall prediction represents a substantial breakthrough in weather forecasting.

CONFLICT OF INTERESTS

The authors confirm that there are no conflicts of interest in publishing this paper. All authors have independently conducted the research without any external influence.

REFERENCES

1. Bruni, M. E., Lazzaroli, V., Perboli, G., & Vandoni, C. (2023, June). Machine Learning to Forecast Rainfall Intensity. In 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 1762-1767). IEEE.

2. Ahmed, H. A. Y., & Mohamed, S. W. (2021, February). Rainfall prediction using multiple linear regressions model. In 2020 international conference on computer, control, electrical, and electronics engineering (ICCCEEE) (pp. 1-5). IEEE.

3. Muneer, M. S., Mustafa, S. M. N., Zehra, S. S., & Maqsood, H. (2023, January). Rain Predictive

Model using Machine learning Techniques. In 2023 International Multi-disciplinary Conference in Emerging Research Trends (IMCERT) (Vol. 1, pp. 1-6). IEEE.

4. Balamurugan, K., Rammya, B., Kumar, D., Kalanithy, K., & Mathan, J. N. (2023, September). Implementation of Effective Rainfall Forecast Model using Machine Learning. In 2023 4th International Conference on Smart Electronics and Communication (ICOSEC) (pp. 1655-1660). IEEE.
5. Praveen, S. Phani, Vamsi Saripudi, V. Harshalokh, T. Sohitha, S. Venkat Sai Karthik, and T. Venkata Pavana Surya Sreekar. "Diabetes Prediction with Ensemble Learning Techniques in Machine Learning." In 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), pp. 1082-1089. IEEE, 2023.
6. Praveen, S. Phani, Polavarapu Chaitanya, A. Mohan, Vahiduddin Shariff, Janjhyam Venkata Naga Ramesh, and Jayaprakash Sunkavalli. "Big Mart Sales using Hybrid Learning Framework with Data Analysis." In 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), pp. 471-477. IEEE, 2023.
7. Donepudi, Swapna, Sai Chaitanya Palagani, P. Sai Naga Pramod, Y. Rohith Kumar, S. Karthikeya, and S. Phani Praveen. "Brain Metastasis Tumor Detection using Image Segmentation and VGG16 Architecture." In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 150-155. IEEE, 2023.
8. Praveen, S. P., Bikku, T., Muthukumar, P., Sandeep, K., Sekhar, J. C., & Pratap, V. K. (2024). Enhanced Intrusion Detection Using Stacked FT-Transformer Architecture. Journal of Cybersecurity & Information Management, 13(2).
9. Praveen, S. Phani, Sai Srinivas Vellela, and R. Balamanigandan. "SmartIris ML: Harnessing Machine Learning for Enhanced Multi-Biometric Authentication." Journal of Next Generation Technology (ISSN: 2583-021X) 4, no. 1 (2024).
10. Anuradha, Chokka, D. Swapna, Balamuralikrishna Thati, V. Navya Sree, and S. Phani Praveen. "Diagnosing for liver disease prediction in patients using combined machine learning models." In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 889-896. IEEE, 2022.
11. Dedeepya, P., P. Sowmya, T. Deva Saketh, P. Sruthi, P. Abhijit, and S. Phani Praveen. "Detecting cyber bullying on twitter using support vector machine." In 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), pp. 817-822. IEEE, 2023.
12. Chittora, Pankaj, Tulika Chakrabarti, Papiya Debnath, Amit Gupta, Prasun Chakrabarti, S. Phani Praveen, Martin Margala, and Ahmed A. Elngar. "Experimental analysis of earthquake prediction using machine learning classifiers, curve fitting, and neural modeling." (2022).
13. Appiah-Badu, N. K. A., Missah, Y. M., Amekudzi, L. K., Ussiph, N., Frimpong, T., & Ahene, E. (2021). Rainfall prediction using machine learning algorithms for the various ecological zones of Ghana. IEEE Access, 10, 5069-5082.
14. Soe, M. M. (2023, February). Rainfall Prediction using Regression Model. In 2023 IEEE Conference on Computer Applications (ICCA) (pp. 113-117). IEEE.
15. Sivanantham, S., Kumar, P. H., Vardhan, S. N., Kumar, S. C., Kumar, T. A., & Pradeep, T. (2023, February). Rainfall Prediction using Machine Learning Techniques–A Comparative Approach. In 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS) (pp. 1472-1480). IEEE.
16. Krithik, P. S., Raghavan, M. A., & Vasanth, K. (2021, November). Real Time Rainfall Prediction for Hyderabad Region using Machine learning Approach. In 2021 Innovations in

Power and Advanced Computing Technologies (i-PACT) (pp. 1-6). IEEE.

17. Niu, J., & Zhang, W. (2015, August). Comparative analysis of statistical models in rainfall prediction. In 2015 IEEE International Conference on Information and Automation (pp. 2187-2190). IEEE.
18. Naveen, A., SS, N. P., & Rajesh, R. (2023, August). Bi-LSTM Model based Real-time Rainfall Prediction. In 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) (pp. 1078-1081). IEEE.
19. P. Dedeepya, D. Karishma, S. G. Manuri, T. Raghuvaran, V. Shariff and S. Sindhura, "Enhancing Cyber Bullying Detection Using Convolutional Neural Network," 2023 4th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2023, pp. 1260- 1267, doi: 10.1109/ICOSEC58147.2023.10276007.