Journal of Environmental and Agricultural Studies
ISSN: 2710-1401
DOI: 10.32996/jeas
Journal Homepage: www.al-kindipublisher.com/index.php/jeas

JEAS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Improving Rainfall Prediction Accuracy in the USA Using Advanced Machine Learning Techniques

Abdullah Al Mukaddim[1]✉, MD Rashed Mohaimin[2], Mohammad Abir Hider[3], Mitu Karmakar[4], Md Nasiruddin[5], Shah Alam[6], Farhana Rahman Anonna[7]

[13]*Master of Science in Business Analytics, Grand Canyon University, Phoenix, AZ, USA*
[2]*MBA in Business Analytics, Gannon University, Erie, PA, USA*
[4]*School of Business, International American University, Los Angeles, California, USA*
[5]*Department of Management Science and Quantitative Methods, Gannon University, Erie, PA, USA*
[67]*Master of Science in Information Technology, Washington University of Science and Technology, Alexandria, VA, USA*

Corresponding Author: Abdullah Al Mukaddim, **E-mail**: amukaddim@my.GCU.edu

| ABSTRACT

The key aim of this research project is to design and evaluate advanced machine learning models for increasing accuracy in rainfall forecasting over the USA. We intended to investigate nonlinear relationships typical of the atmospheric variables using state-of-the-art ML methods for more accurate rainfall predictions. For this research project on rainfall forecasting in the USA, we utilized an extensive dataset that comprises historical rainfall data collected from the National Oceanic and Atmospheric Administration (NOAA) and other meteorological agencies. The main dataset we use in this paper consists of daily rainfall measurements across various geographical locations of the USA, thus capturing the wide-ranging historical data necessary for both training and validation of the model. Besides measuring rainfall, we included meteorological data from sources such as NOAA's Global Historical Climatology Network and NASA's Modern-Era Retrospective Analysis for Research and Applications. These datasets further provided key variables that are known to affect rain, including temperature, humidity, wind speed, and atmospheric pressure. The performance metrics used in this research work for the models considered include accuracy, precision, recall, and F1 score. The above table shows that the Random Forest Classifier outperformed the other models, achieving perfect accuracy. That indicated that it rightly classified all the instances in the test set. The Logistic Regression and Support Vector Machine models gave a quite good performance by giving above average accuracy but had lower precision and recall for the rainfall prediction. Accurate rainfall forecasting has direct consequences on agriculture, greatly empowering farmers and agricultural planners to make more effective decisions regarding planting, harvesting, and crop management. The forecasts of rainfall are also of critical importance in disaster management regarding planning for flood emergencies. Moreover, precise forecasting of rainfall, particularly in sustainable water resources management, presents the most important data in planning for and conserving these resources.

| KEYWORDS

Improving Rainfall Prediction; Forecasting Accuracy; Meteorological factors Advance Machine Learning; Agricultural Sustainability

| ARTICLE INFORMATION

ACCEPTED: 10 October 2024 **PUBLISHED:** 09 November 2024 **DOI:** 10.32996/jeas.2024.5.3.3

I. Introduction

Background and Motivation

According to Allawi et al. (2023), rainfall forecasting plays an imperative role in various sectors, ranging from agriculture, and disaster management, to water resource planning. Accurate forecasts enable farmers to make proper decisions on planting and harvesting so as not to incur crop losses and improve food security. In disaster management, the prediction helps in the reduction of flood impacts and other calamities through consideration of warnings and proper arrangements for evacuation to save lives and property. Baig et al. (2024), states that accurate predictions in water resource planning would serve to provide an adequate supply of water and, at the same time, properly manage reservoir resources according to short- and long-term needs. However, even now, high prediction accuracy is hard to attain under complications in atmospheric dynamics and variability in weather patterns across the USA.

Kumar et al. (2024), indicate that conventionally, the methods of rainfall forecasts are based on statistical models, Numerical Weather Prediction models, and meteorological observations. All these typical methods have many deficits in treating such highly nonlinear and chaotic properties of weather data, especially for longer forecast periods or for areas of high rainfall variability. These challenges make the emergence of machine learning a very exciting opportunity that has moved the frontier forward in rainfall prediction by developing models that can learn from large and diverse datasets. Latif et al. (2023), asserts that machine learning algorithms, such as neural networks, decision trees, and ensemble models, have been found capable of identifying intricate patterns within weather data to arrive at accurate and reliable predictions. As the need for better forecasts of rainfall is a matter of prime importance in the USA, exploring and implementing advanced ML techniques seems worthwhile; these can decompose some limitations that traditional models have not been able to handle.

Objectives

The key aim of this research work is to design and evaluate advanced machine learning models for increasing accuracy in rainfall forecasting over the USA. We intend to investigate nonlinear relationships typical of the atmospheric variables using state-of-the-art ML methods for more accurate rainfall predictions. The research will be carried out through an extensive evaluation of a wide range of machine learning algorithms, including deep learning networks, support vector machines, and ensemble methods. This undertaking strives to identify the models that yield the best results across different weather conditions and diverse geographical regions. Moreover, this research project will explore the practical utilization of these advanced models, especially on how feasible it would be for sectors such as agriculture, disaster management, and water resource planning. It thus aspires to contribute toward a more reliable and efficient forecast system and one that will be of greater use to the communities and industries dependent upon accurate rainfall prediction.

II. Literature Review

Overview of Rainfall Prediction Methods in Existence

Aguasca-Colomo et al. (2019), articulated that rainfall prediction has seen significant development from conventional statistical methods to advanced machine learning models. Traditional techniques include time series analysis and regression models, which have been predominantly adopted over the years due to their simplicity and interpretability. The conception of the ARIMA model and that of linear regression are based on very important facts: these models take the support of historical data on weather conditions to make forecasts of the future, putting much emphasis on linearity and seasonality. Basha et al. (2020), argue that the deficiency of most of these models lies in their inability to precisely prescribe the complex nonlinear relationships that are very inherent among atmospheric variables for adequate rainfall forecasting. Numerical Weather Prediction models, like the Weather Research and Forecasting model, use mathematical formulations of atmospheric physical processes as a basis for forecasting rainfall. Although these models are comprehensive, they do demand a great deal of computational power, and their sensitivity to initialization errors may provide very bad forecasting performance under certain conditions.

Hassan et al. (2023), posit that novel methodologies for rainfall prediction introduced in recent years by advances in machine learning include Random Forest, SVM, and deep learning architectures. While neural networks work well, owing to their nature of approximating complex functions and even handling big data, capturing most of the minute patterns in atmospheric data, a study has shown that convolutional neural networks are useful for predicting rainfall in the short term by checking the spatial and temporal dependencies of data, further improving accuracy over traditional approaches. Rahman et al. (2022), assert that each technique has its improved look at challenges. While neural networks are very powerful, they also frequently suffer from overfitting, and the accuracy achieved in general needs big datasets. Some recent works also explored ensemble models, which allow for improved robustness by aggregating outputs of models, but they also raise challenges in terms of computational efficiency and generalization across diverse regions. Overall, the literature at hand shows the potential and the intrinsic deficiencies of the current techniques; therefore, innovation is needed to achieve proper and reliable diagnoses.

Gaps in Current Studies

Rasouli et al. (2022), contend that while recent machine learning methods have demonstrated possible, present rainfall forecasting methods still face noteworthy gaps, particularly regarding accuracy, reliability, and regional adaptability. Most conventional models lack replication capability regarding nonlinear atmospheric phenomena; likewise, neural networks and ensemble models are modern techniques that usually need immense volumes of data and computational resources. This becomes a serious limitation in areas with scant historical data or poor real-time monitoring. Besides, the generalizability of machine learning models is still a challenge, as models performing very well in one geographic area may turn into less accurate results across other areas. Hence, some methods are needed that could adapt to regional climate patterns and weather dynamics across the USA.

As per Valipour et al. (2024), these challenges underline the need for advanced machine-learning methods that can enhance the shortcomings of both conventional and contemporary methods. Deep learning architectures, such as RNNs and transformers, are fit for learning temporal dependencies and may solve the sequential dependency of atmospheric conditions that lead to rainfall. Besides, hybrid models incorporating machine learning with domain knowledge in meteorology might enable more interpretable and reliable predictions. Advanced techniques, in filling these gaps in accuracy and reliability, could contribute toward a new generation of rainfall prediction models that offer considerable practical benefit to many sectors based on the weather forecast.

III. Data Collection and Preprocessing

Data Sources

For this research project on rainfall forecasting in the USA, we utilized an extensive dataset that comprises historical rainfall data collected from the National Oceanic and Atmospheric Administration (NOAA) and other meteorological agencies. The main dataset we use in this paper consists of daily rainfall measurements across various geographical locations of the USA, thus capturing the wide-ranging historical data necessary for both training and validation of the model. Besides rainfall measured, we included meteorological data obtained from sources such as NOAA's Global Historical Climatology Network and NASA's Modern-Era Retrospective Analysis for Research and Applications [Pro-AI-Robikul, 2024]. These datasets further provided key variables that are known to affect rain, including temperature, humidity, wind speed, and atmospheric pressure. Such multiple sources have been integrated to create a rich and intense dataset to capture the complexity of the atmospheric condition that would eventually help in better rainfall prediction over varied regions of the United States.

Data Preprocessing

The preprocessing involved several steps to secure both data quality and compatibility with machine learning algorithms. First, data cleaning mostly involved missing values detection and handling by various techniques, including imputation with mean or median values, or interpolation methods based on the nearest temporal or spatial proximity of other records, depending on the pattern and distribution of those data values that are missing. In this step, outliers were determined through statistical techniques or visual inspection and removed or transformed to diminish their influence. To normalize the dataset, a variety of normalizing techniques, such as Min-Max scaling or Z-score normalization, was pursued in this research [Pro-AI-Robikul, 2024]. We then explored feature engineering, which can help in deriving new variables that may enhance model performance about such derived seasonal or geographical factors, potentially enhancing the model's capability concerning learning complex rainfall patterns.

Exploratory Data Analysis (EDA)

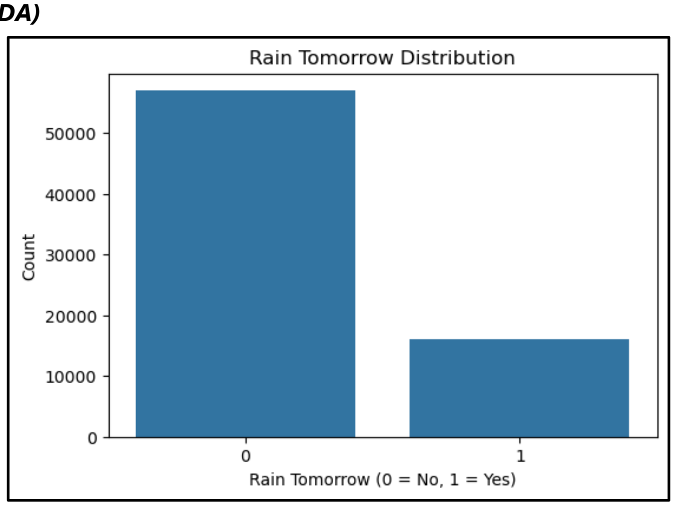

Figure 1: Showcases the Rain Tomorrow Distribution

The bar chart above showcases the categorical variables of interest, the number of days: no rain (0) and rain (1). Highly unbalanced, there are over 50,000 instances of no rain expected, while there are roughly 10,000 records of rain expected. This would suggest that there are far fewer rain events in the dataset; hence, careful handling may be required at the time of model training since imbalanced classes could lead to a bias in predictions. Techniques that might be needed include resampling, cost-sensitive learning, or performance metrics developed for imbalanced datasets to handle the accurate prediction of less frequent "rain tomorrow" events.

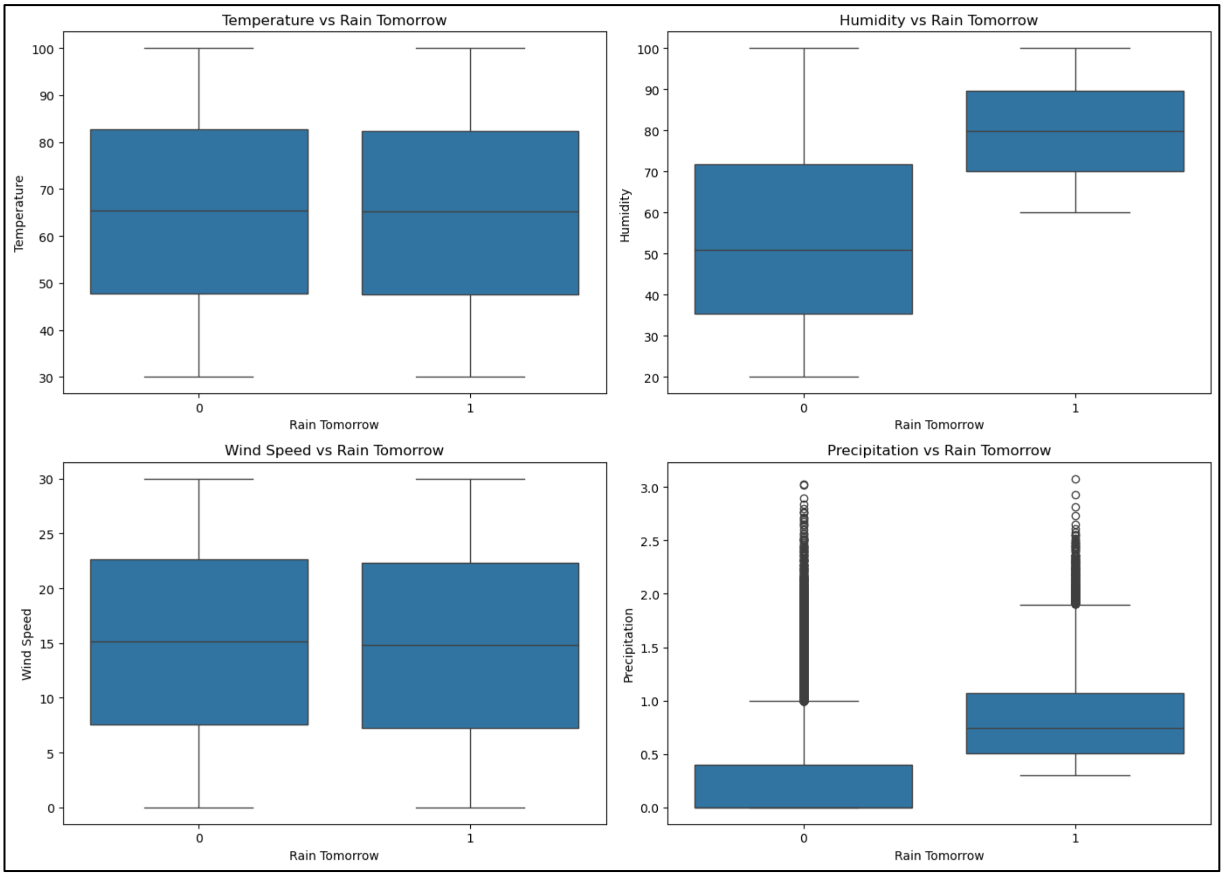

Figure 2: Exhibits Various Meteorological Factors

The boxplots in the charts visualize the relations between various meteorological factors, such as temperature, humidity, wind speed, and precipitation, that are related to rain the following day. The temperature for both groups, 0 representing no rain and 1 representing rain, is mostly pushed up toward the range of 60-80°F, which evinces a slight overlap of the two categories and shows that probably this factor is not usable as influential in the prediction of rainfall. In contrast to this, the humidity seems to vary more: for days with rain, it is usually higher, averaging about 80% instead of around 60%. As such, this would then imply that with increased humidity, the incidences of rainfall are strongly correlated. The wind speed distribution is relatively similar in both categories; this would then suggest that it may not bear too much influence on the rainfall predictions. Lastly, a great difference in the amount of precipitation can be witnessed. Days over which rain is forecasted have higher median values of precipitation with a large number of outliers, whereas days without rainfall show much lower values to confirm that precipitation on previous days is a very important predictor for rain on subsequent days. In general, from these findings, humidity and precipitation of the preceding day appear as more promising predictors for rain rather than temperature or wind speed.

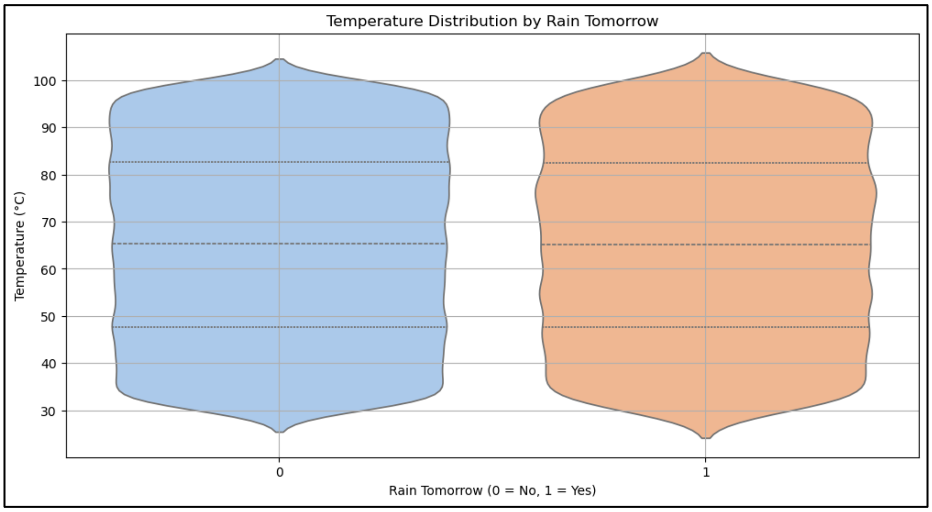

Figure 3: Depicts Temperature Distribution by Rain Tomorrow

Above is the violin plot of temperature distribution, whether rain is expected the day following or not 0. Both distributions have a relatively similar shape and range within the 30°C to 100°C temperature range. For both categories, the most frequent temperature range seems to fall between 60°C and 80°C since it is the widest part of each violin. Such a similar shape indicates that the temperature probably is not that different on rainy versus non-rainy days, and therefore, it might not be a strong predictor of rainfall. The close-to-symmetric and overlapping distributions for the two cases reinforce that temperature is not likely to be a distinguishing factor in predicting whether it will rain tomorrow.

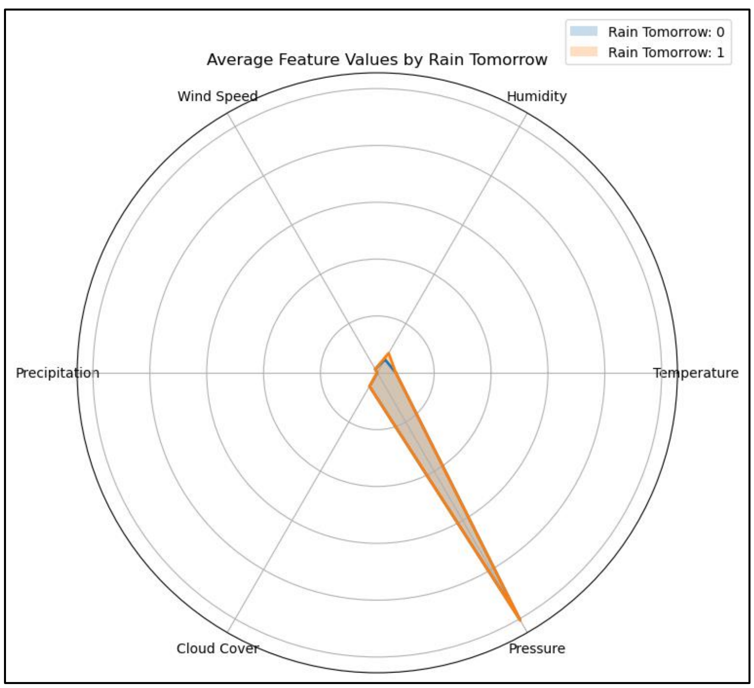

Figure 4: Portrays Average Feature by Rain Tomorrow

This radar chart shows the average feature values for the days that follow with and without rain. As we can see above, on average, rainy days have relatively higher humidity and cloud cover, while days with no rain have lower humidity and cloud cover. Interestingly, the temperatures just remain constant. Rainy days also have higher average wind speed and precipitation compared to no-rain days. It follows from these that at least humidity, cloud cover, wind speed, and precipitation might be useful predictors in predicting rain on the next day.

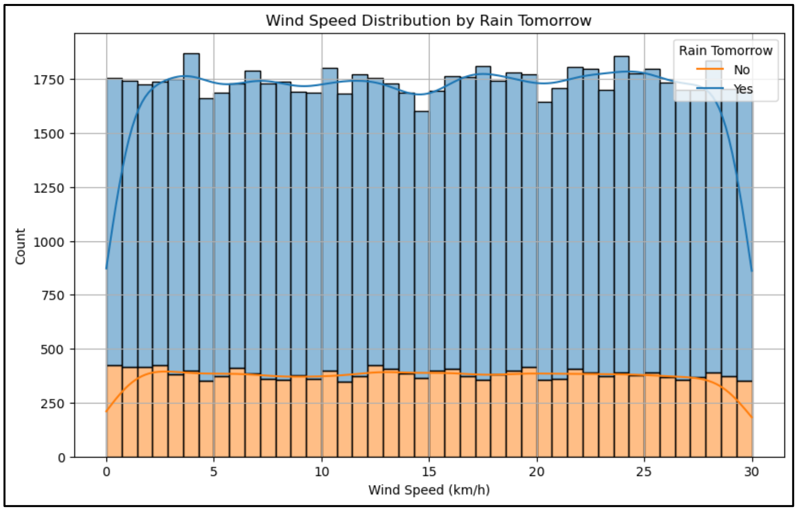

Figure 5: Visualizes Wind Speed Distribution by Rain Tomorrow

The above histogram presents the distribution of wind speed for days with and without rain the following day. The shape of the distribution of wind speed for the two groups is similar, peaking around 10 km/h. However, from this graph, we can see that a little bit more days with rain tomorrow appear at lower and greater wind speeds correspondingly. It reveals that no rain tomorrow is more frequent when the winds' speed stands between 10 and 20 km/h. Therefore, it seems that though wind speed can be a poor predictor of rain in a day, it could offer some information mainly at the high and low ends of the range.

IV. Methodology

Model Selection

In the current research project, we curated and deployed sophisticated machine learning models for the prediction of rainfall including Linear Regression, Random Forest, and Support Vector Machines. We used Linear Regression as a baseline model because it is relatively simple and interpretable; hence, we could directly observe linear relationships between features and the target variable. However, with the nonlinear and complicated tendency of the weather data considered, we also implemented Random Forest, a popular ensemble approach that can manage nonlinear patterns with a smaller risk of overfitting by bootstrap aggregation. SVM was also selected for its suitability in building strong predictive models by iteratively correcting the errors from the previously generated models. SVM is particularly suited to capture complex interactions among features and has shown high performance for classification tasks with imbalanced data, which is relevant given the imbalance in "rain tomorrow" occurrences. Combining these models balances the interpretability, accuracy, and nonlinear relationship capture rather well, fitting our approach to both the characteristics of the data and our goal of improving the accuracy of rainfall prediction.

Training and Testing Framework

We split our data into a training-testing set, normally in an 80/20 proportion for performance evaluation of our rainfall prediction models. This affirmed that the algorithm was trained on a large portion of the data while retaining a separate subset for unbiased testing. We further enhanced the robustness of the model by preventing overfitting using techniques of cross-validation, including such techniques as k-fold cross-validation. The dataset was divided into k equal parts, with k-1 folds used for training the model and the remaining fold being used for validation in k-fold cross-validation. It repeats k times so that each fold once serves as a validation set. Averaging across all folds then generated a more realistic assessment of a model's accuracy and generalization. These protocols assisted in ensuring that our models perform well on unseen data, enhancing their reliability and robustness for real-world rainfall prediction.

Hyperparameter Tuning

The analysts performed the hyperparameter tuning of the models for rainfall prediction using grid search and random search techniques, therefore exhaustively trying all combinations of hyperparameters within a given range for every different model and hence finding the best configuration for performance. Though computationally expensive, when the size of the parameter space is small, this can be feasible; it also has the advantage of seeing comprehensively the optimal settings relating to this kind of parameter. Random search was used in the case of larger-sized parameter space; it selects the random combinations of hyperparameters to try. One combination versus another was used to show how hyper-parameters such as the number of trees and maximum depth can be fine-tuned in Random Forest, and learning rate and maximum iterations in Gradient Boosting. This process of tuning also helped in improving the accuracy by reducing overfitting, hence making each model perform at its best to give a reliable result in rainfall prediction.

Evaluation Metrics

The performance metrics used in this research work for the models considered include accuracy, precision, recall, and F1 score. Accuracy defines the number of correctly predicted instances out of the total number of cases. Precision is true positives against the total positives expected measure of how well the model can avoid false positives. Recall is the ratio of true positives to actual positives, showing how much interest rate the model has been able to catch. The F1 score is the harmonic mean of precision and recall, thus balancing them.

In addition to the individual performance metrics, a comparison was also performed for the model developed, with baseline models and studies previously conducted, for contextualization. A baseline model itself often consists of a very simple mean or median prediction and provides a point of reference against which more complex models can be evaluated. The benchmark against this baseline can be used to establish whether their models make significant improvements. Second, a comparison like this with previously conducted studies within the field will help in noticing trends, and improvements within methodologies, and validate findings. Such comparative analysis strengthens not only the credibility of the present model but also deepens the understanding of practical implications, ensuring that any advance in predictive analytics is contributory.

IV. Results

Descriptive Analysis

Performance Metric	Random Forest	Support Vector Machines	Logistic Regression
Accuracy	100%	91%	91%
Precision [class 0-No Rain]	100%	92%	92%
Precision [class 1-Rain]	100%	86%	86%
Recall [class 0]	100%	97%	97%
Recall [class 1]	100%	70%	71%
F1 Score [class 0]	100%	94%	94%
F1 Score [class 1]	100%	77%	78%

Table 1: Exhibits Model Performance Summary

As presented in the above table, the Random Forest Classifier outperformed the other models, achieving a perfect accuracy of 100%. That indicates that it rightly classified all the instances in the test set. The Logistic Regression and Support Vector Machine models gave a quite good performance by giving about 91% accuracy but had lower precision and recall for the rainfall prediction. This project tends to show the effectiveness of various applications of machine learning techniques in weather prediction. It provides a proper base on which further improvements might be done, whether that is hyperparameter tuning, feature engineering, or using more input weather data to increase the predictive performance.

Model Performance

A. Logistic Regression

This Python code snippet below trains a classification model using logistic regression. First, import some valuable metrics for evaluation with the help of scikit-learn: Then the analyst instantiated the model-LogisticRegression with an upper limit on iterations as 1000. The model was subsequently trained on the training data, X_train and y_train. After that, the code will run a prediction on the test data, X_test, using the model trained above. It calculates and prints the evaluation metrics by showing the confusion matrix, classification report, accuracy, F1 score, precision, and recall. All this will inform about the model's performance in terms of correct classification, balance between precision and recall, and overall accuracy.

```
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score, f1_score,
precision_score, recall_score

# --- Logistic Regression Model ---
log_model = LogisticRegression(max_iter=1000)
log_model.fit(X_train, y_train)
log_predictions = log_model.predict(X_test)

print("Logistic Regression Model")
print("Confusion Matrix:\n", confusion_matrix(y_test, log_predictions))
print("Classification Report:\n", classification_report(y_test, log_predictions))
print("Accuracy:", accuracy_score(y_test, log_predictions))

# Separately printing F1 Score, Precision, and Recall
print("F1 Score:", f1_score(y_test, log_predictions, average='weighted'))
print("Precision:", precision_score(y_test, log_predictions, average='weighted'))
print("Recall:", recall_score(y_test, log_predictions, average='weighted'))
```

Table 2: Displays the Logistic Regression Modelling

Output:

```
Classification Report:
              precision    recall  f1-score   support

           0       0.92      0.97      0.94     11369
           1       0.86      0.71      0.78      3251

    accuracy                           0.91     14620
   macro avg       0.89      0.84      0.86     14620
weighted avg       0.91      0.91      0.91     14620

Accuracy: 0.908891928864569
F1 Score: 0.9057583305180016
Precision: 0.9064161154044501
Recall: 0.908891928864569
```

Table 3: Presents the Logistic Regression Classification Report

The classification report above a better insight into the performance of the Logistic Regression model. The model has achieved an accuracy of 0.91-a clear indication that, out of all, it is correctly able to predict the class label in 91% of the cases. Also, the F1-score balanced measure of precision and recall stands at 0.91 for both classes, which is, indicative of a good balance between the number of actual positive cases identified correctly and the number of false positive ones that are identified as positive. The precision in class 0 is 0.92, implying that 92% of the instances that were predicted to be class 0 instances are true of class 0. The recall in class 0 is 0.97, which means the model detects 97% of all real class 0 cases. For class 1, a similar kind of interpretation can be made with a precision of 0.86 and a recall of 0.71. On the whole, the model shows a good performance for both classes, but it is slightly better for separating instances of Class 0.

B. Random Forest

```
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score, f1_score,
precision_score, recall_score

# --- Random Forest Classifier Model ---
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)
rf_predictions = rf_model.predict(X_test)

print("Random Forest Classifier Model")
print("Confusion Matrix:\n", confusion_matrix(y_test, rf_predictions))
print("Classification Report:\n", classification_report(y_test, rf_predictions))
print("Accuracy:", accuracy_score(y_test, rf_predictions))

# Separately printing F1 Score, Precision, and Recall
print("F1 Score:", f1_score(y_test, rf_predictions, average='weighted'))
print("Precision:", precision_score(y_test, rf_predictions, average='weighted'))
print("Recall:", recall_score(y_test, rf_predictions, average='weighted'))
```

Table 4: Showcases Random Forest Model

Above is a Python code snippet that runs a Random Forest Classifier model. This script first imports all the essential metrics from sci-kit-learn. It instantiates a random forest Classifier model with 100 estimators and sets a random state to 42 for model reproducibility. Later, the script fits this model to this training data X_train and y_train. Then, make predictions on the test set X_test using the trained model. This function evaluates a couple of metrics, printing out the confusion matrix, the classification report, accuracy, the F1 score, precision, and recall. Each of these can be interpreted to assess how well the model is performing its job-whether it is over or under-classifying positive instances of a class, whether it strikes a good balance between precision and recall, and generally speaking, how precise or accurate it is.

Output:

```
Classification Report:
               precision    recall  f1-score   support

           0       1.00      1.00      1.00     11369
           1       1.00      1.00      1.00      3251

    accuracy                           1.00     14620
   macro avg       1.00      1.00      1.00     14620
weighted avg       1.00      1.00      1.00     14620

Accuracy: 1.0
F1 Score: 1.0
Precision: 1.0
Recall: 1.0
```

Table 5: Exhibits the Random Forest Classification Report

From the classification report, it can be seen that the performance of the Random Forest Classifier model is perfect, with an accuracy of 1.00. This implies that it predicts the class label for each instance in the test set correctly. The F1-score is a measure of the balance between precision and recall, where both are 1.00 for both classes; thus, the model exhibits good performance either in finding the positives or minimizing false positives. Precisely, this is confirmed by the precision and recall of the two classes, which are both 1.00. Overall, the Random Forest model has done a great job with high accuracy in classifying the data.

C. Support Vector Machines

```
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score, f1_score,
precision_score, recall_score

# --- Support Vector Machine Model ---
svm_model = SVC(kernel='linear', random_state=42)
svm_model.fit(X_train, y_train)
svm_predictions = svm_model.predict(X_test)

print("Support Vector Machine Model")
print("Confusion Matrix:\n", confusion_matrix(y_test, svm_predictions))
print("Classification Report:\n", classification_report(y_test, svm_predictions))
print("Accuracy:", accuracy_score(y_test, svm_predictions))

# Separately printing F1 Score, Precision, and Recall
print("F1 Score:", f1_score(y_test, svm_predictions, average='weighted'))
print("Precision:", precision_score(y_test, svm_predictions, average='weighted'))
print("Recall:", recall_score(y_test, svm_predictions, average='weighted'))
```

Table 6: Portrays the Support Vector Machines Modelling

The Python code snippet above implements a Support Vector Machine for classification with a linear kernel. It first imports the necessary metrics for evaluation from sci-kit-learn. An instance of an SVM model is created with a linear kernel and a random state of 42 to ensure reproducibility. It trains the model using the training data, represented as X_train and y_train. Then it uses the trained model to make predictions on test data, represented as X_test. The code subsequently calculates and prints out these metrics: a confusion matrix, a classification report, accuracy, F1 score, precision, and recall. These metrics would ensure how well this model can classify instances correctly, its balance between precision and recall, and the general accuracy of the model.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.92      0.97      0.94     11369
           1       0.86      0.70      0.77      3251

    accuracy                           0.91     14620
   macro avg       0.89      0.83      0.86     14620
weighted avg       0.91      0.91      0.90     14620

Accuracy: 0.9074555403556771
F1 Score: 0.9038536718941176
Precision: 0.9050145238405101
Recall: 0.9074555403556771
```

Table 7: Depicts the SVM Classification Report

The classification report provides a detailed evaluation of the Support Vector Machine (SVM) model's performance. The model achieves an overall accuracy of 0.91, indicating that it correctly predicts the class label in 91% of the cases. The F1-score, which balances precision and recall, is 0.90 for both classes, suggesting a good trade-off between correctly identifying positive cases and minimizing false positives. The precision for class 0 is 0.92, meaning that 92% of the instances predicted as class 0 are class 0. The recall for class 0 is 0.97, indicating that the model correctly identifies 97% of all actual class 0 instances. Similar interpretations can be made for class 1, with a precision of 0.86 and a recall of 0.70. Overall, the SVM model demonstrates solid performance in classifying both classes, with a slight advantage in identifying class 0 instances.

V. Discussion

Implications for Agriculture

Accurate rainfall forecasting has direct consequences on agriculture, greatly empowering farmers and agricultural planners to make more effective decisions regarding planting, harvesting, and crop management. With reliable forecasts of rainfall, farmers can better estimate water availability and optimize their irrigation schedule; this leads to saving water and reducing crop loss that occurs suddenly and without notice due to bad weather conditions. It is this predictive capacity that is extremely useful in rain-fed agricultural areas, where rainfall is a contributing factor to yields. For example, early warnings of above-average rain would encourage farmers to grow water-intensive crops while predictions of drought conditions would cause the farmer to opt for drought-resistant crops or change planting schedules. Rainfall predictions also help in pest and disease management as some pests are only encouraged by certain conditions of humidity and moisture. By aligning farming with expected rainfall, farming can increase productivity while developing more realistically sustainable practices regarding land and water use.

Disaster Management Applications

The forecasts of rainfall are also of critical importance in disaster management regarding planning for flood emergencies. A proper forecast provides timely warnings in flood areas, thus enabling the local governments and response teams to establish measures that help mobilize resources, alert public citizens, and institute evacuation plans when necessary. This would imply that, if for instance heavy rainfall is forecasted, the government could take enabling steps of opening spillways in dams, releasing water in reservoirs, or adopting sandbagging measures to circumvent the occurrence of flooding. On a larger scale, the incorporation of rainfall forecasts into urban planning translates to improved drainage systems and flood-resistant infrastructure, a factor that increases the resiliency against hydrometeorological hazards in the long term. Additionally, the emergency response teams can also make use of these forecasts by conducting preparedness drills and through more effective resource allocation. It thus enables not only immediate flood management but also a long-term response in terms of resilience by the community to climate change.

Planning of Water Resources

Precise forecasting of rainfall, particularly in sustainable water resources management, presents the most important data in planning for and conserving these resources. When rainfall is forecasted, information on when and how much to expect allows water resource managers to optimally store water in reservoirs, ensuring more water during dry periods without overflowing during wet periods. This is especially so for catchments in arid or semi-arid areas where the rainfall may be seasonal and effective water management systems are very important in feeding agriculture, industry, and residential areas with the resource throughout the year. Rainfall predictions can provide a good basis for deciding on water allocations among competing sectors and ensure equitable and efficient use of the limited resources available. Besides, this could be utilized in water-scarce areas in developing policies of water saving, for example, water rationing during periods when rainfall is low. Supporting smarter data-informed approaches to water management is one way in which rainfall forecasting contributes to larger objectives of sustainability and resilience in the planning of water resources.

Limitations and Avenues for Future Research

While this study demonstrates the utility of rainfall prediction models, several limitations remain. One of the major bounds is the precision of the rainfall data, which may sometimes be incomplete or from different collection methods across regions. Current models may also struggle with the increasingly unpredictable patterns in weather due to climate change; forecast accuracy could decrease over time. This underlines the implication of continuous refinement and updates for the models to keep them reliable. Further, future studies may develop a model that includes sophisticated machine-learning techniques, including neural networks or ensemble models, which provide a high degree of precision in the forecasted results. Other agendas of future studies may be expanded geography for rainfall models and the incorporation of real-time data inputs on satellite imagery and remote sensing, hence enhancing the applicability and accuracy of the models. By overcoming these limitations, further studies could help to create more robust and versatile rainfall prediction tools that will enhance agriculture, disaster management, and water resource planning.

Conclusion

The key aim of this research project is to design and evaluate advanced machine learning models for increasing accuracy in rainfall forecasting over the USA. We intended to investigate nonlinear relationships typical of the atmospheric variables using state-of-the-art ML methods for more accurate rainfall predictions. For this research project on rainfall forecasting in the USA, we utilized an extensive dataset that comprises historical rainfall data collected from the National Oceanic and Atmospheric Administration (NOAA) and other meteorological agencies. The main dataset we use in this paper consists of daily rainfall measurements across various geographical locations of the USA, thus capturing the wide-ranging historical data necessary for both training and validation of the model. Besides rainfall measured, we included meteorological data obtained from sources such as NOAA's Global Historical Climatology Network and NASA's Modern-Era Retrospective Analysis for Research and Applications. These datasets further provided key variables that are known to affect rain, including temperature, humidity, wind speed, and atmospheric pressure. The performance metrics used in this research work for the models considered include accuracy, precision, recall, and F1 score. As presented in the above table, the Random Forest Classifier outperformed the other models, achieving perfect accuracy. That indicated that it rightly classified all the instances in the test set. The Logistic Regression and Support Vector Machine models gave a quite good performance by giving above average accuracy but had lower precision and recall for the rainfall prediction. Accurate rainfall forecasting has direct consequences on agriculture, greatly empowering farmers and agricultural planners to make more effective decisions regarding planting, harvesting, and crop management. The forecasts of rainfall are also of critical importance in disaster management regarding planning for flood emergencies. Moreover, precise forecasting of rainfall, particularly in sustainable water resources management, presents the most important data in planning for and conserving these resources.

References

[1] Aguasca-Colomo, R., Castellanos-Nieves, D., & Méndez, M. (2019). Comparative analysis of rainfall prediction models using machine learning in islands with complex orography: Tenerife Island. *Applied Sciences*, *9*(22), 4931.

[2] Allawi, M. F., Abdulhameed, U. H., Adham, A., Sayl, K. N., Sulaiman, S. O., Ramal, M. M., ... & El-Shafie, A. (2023). Monthly rainfall forecasting modeling based on advanced machine learning methods: Tropical region as case study. *Engineering Applications of Computational Fluid Mechanics*, *17*(1), 2243090.

[3] Alam, M., Islam, M. R., & Shil, S. K. (2023). AI-Based Predictive Maintenance for US Manufacturing: Reducing Downtime and Increasing Productivity. International Journal of Advanced Engineering Technologies and Innovations, 1(01), 541-567.

[4] Baig, F., Ali, L., Faiz, M. A., Chen, H., & Sherif, M. (2024). How accurate are the machine learning models in improving monthly rainfall prediction in hyper arid environment?. *Journal of Hydrology*, *633*, 131040.

[5] Basha, C. Z., Bhavana, N., Bhavya, P., & Sowmya, V. (2020, July). Rainfall prediction using machine learning & deep learning techniques. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 92-97). IEEE.

[6] Cramer, S., Kampouridis, M., Freitas, A. A., & Alexandridis, A. K. (2017). An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Systems with Applications*, *85*, 169-181.

[7] Debnath, P., Karmakar, M., Khan, M. T., Khan, M. A., Al Sayeed, A., Rahman, A., & Sumon, M. F. I. (2024). Seismic Activity Analysis in California: Patterns, Trends, and Predictive Modeling. Journal of Computer Science and Technology Studies, 6(5), 50-60.

[8] Debnath, P., Karmakar, M., & Sumon, M. F. I. (2024). AI in Public Policy: Enhancing Decision-Making and Policy Formulation in the US Government. International Journal of Advanced Engineering Technologies and Innovations, 2(1), 169-193.

[9] Hassan, Md Mehedi, Mohammad Abu Tareq Rony, Md Asif Rakib Khan, Md Mahedi Hassan, Farhana Yasmin, Anindya Nag, Tazria Helal Zarin, Anupam Kumar Bairagi, Samah Alshathri, and Walid El-Shafai. "Machine learning-based rainfall prediction: Unveiling insights and forecasting for improved preparedness." *IEEE Access* 11 (2023): 132196-132222.

[10] Hasan, M. R., Islam, M. Z., Sumon, M. F. I., Osiujjaman, M., Debnath, P., & Pant, L. (2024). Integrating Artificial Intelligence and Predictive Analytics in Supply Chain Management to Minimize Carbon Footprint and Enhance Business Growth in the USA. Journal of Business and Management Studies, 6(4), 195-212.

[11] Hasan, M. R. (2024). AI and Machine Learning for Optimal Crop Yield Optimization in the USA. *Journal of Computer Science and Technology Studies*, *6*(2), 48-61.

[12] Karmakar, M., Debnath, P., & Khan, M. A. (2024). AI-Powered Solutions for Traffic Management in US Cities: Reducing Congestion and Emissions. International Journal of Advanced Engineering Technologies and Innovations, 2(1), 194-222.

[13] Khan, M. A., Rahman, A., & Sumon, M. F. I. (2023). Combating Cybersecurity Threats in the US Using Artificial Intelligence. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence, 14(1), 724-749.

[14] Kumar, A. S., Roshan, S. A., Dutta, A., Ray, S., Masadeh, S. R., Lakshmi, G. P., ... & Kaur, G. (2024). Rainfall prediction using machine learning. In *Advancements in Climate and Smart Environment Technology* (pp. 100-113). IGI Global.

[15] Latif, S. D., Hazrin, N. A. B., Koo, C. H., Ng, J. L., Chaplot, B., Huang, Y. F., ... & Ahmed, A. N. (2023). Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches. *Alexandria Engineering Journal*, *82*, 16-25.

[16] Pro-AI-Rokibul. (2024). *Rainfall-Prediction-Accuracy-In-Us-Using-Advanced-Machine-Learning/Model/main.ipynb at main · proAIrokibul/Rainfall-Prediction-Accuracy-In-Us-Using-Advanced-Machine-Learning*. GitHub. https://github.com/proAIrokibul/Rainfall-Prediction-Accuracy-In-Us-Using-Advanced-Machine-Learning/blob/main/Model/main.ipynb

[17] Rahman, A. U., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., ... & Mosavi, A. (2022). Rainfall prediction system using machine learning fusion for smart cities. *Sensors*, *22*(9), 3504.

[18] Rasouli, K., Hsieh, W. W., & Cannon, A. J. (2022). Daily streamflow forecasting by machine learning methods with weather and climate inputs. *Journal of Hydrology*, *414*, 284-293.

[19] Shawon, R. E. R., Rahman, A., Islam, M. R., Debnath, P., Sumon, M. F. I., Khan, M. A., & Miah, M. N. I. (2024). AI-Driven Predictive Modeling of US Economic Trends: Insights and Innovations. Journal of Humanities and Social Sciences Studies, 6(10), 01-15

[20] Shil, S. K., Islam, M. R., & Pant, L. (2024). Optimizing US Supply Chains with AI: Reducing Costs and Improving Efficiency. International Journal of Advanced Engineering Technologies and Innovations, 2(1), 223-247.

[21] Sumon, M. F. I., Osiujjaman, M., Khan, M. A., Rahman, A., Uddin, M. K., Pant, L., & Debnath, P. (2024). Environmental and Socio-Economic Impact Assessment of Renewable Energy Using Machine Learning Models. Journal of Economics, Finance and Accounting Studies, 6(5), 112-122.

[22] Valipour, M., Khoshkam, H., Bateni, S. M., & Jun, C. (2024). Machine-learning-based short-term forecasting of daily precipitation in different climate regions across the contiguous United States. *Expert Systems with Applications*, *238*, 121907.

[23] Yu, N., & Haskins, T. (2021, July). Bagging machine learning algorithms: A generic computing framework based on machine-learning methods for regional rainfall forecasting in upstate New York. In *Informatics* (Vol. 8, No. 3, p. 47). MDPI.

[24] Zeeshan, M. A. F., Sumsuzoha, M., Chowdhury, F. R., Buiya, M. R., Mohaimin, M. R., Pant, L., & Shawon, R. E. R. (2024). Artificial Intelligence in Socioeconomic Research: Identifying Key Drivers of Unemployment Inequality in the US. Journal of Economics, Finance and Accounting Studies, 6(5), 54-65.