Abstract:

Rainfall prediction is a challenging task due to the complex nature of atmospheric factors involved. Traditional methods often fail to provide accurate forecasts, prompting the need for alternative approaches. This article presents a machine learning-based solution for rainfall prediction using Python. By leveraging libraries such as Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, XGBoost, and Imblearn, alongside a dataset containing atmospheric data, a predictive model is constructed.

The process begins with importing necessary libraries and loading the dataset into a Pandas DataFrame. Data cleaning techniques are applied to handle null values and inconsistencies, ensuring the dataset's readiness for analysis. Exploratory Data Analysis (EDA) techniques are then employed to gain insights into the data's distribution, correlation, and potential patterns. Visualizations aid in understanding the relationships between different atmospheric factors and rainfall occurrence.

Following EDA, the dataset is prepared for model training by separating features and target variables. Due to class imbalance, the training data is balanced using the Random Over-Sampling technique. Feature scaling is applied to normalize the data for stable and efficient training. Three classification algorithms, namely Logistic Regression, XGBoost, and Support Vector Classifier (SVC), are trained on the prepared dataset.

Model performance is evaluated using both training and validation data, with metrics such as ROC AUC score providing insights into each model's accuracy. The results indicate that Logistic Regression and SVC exhibit satisfactory performance, with relatively low gaps between training and validation accuracies. Confusion matrices and classification reports further validate the models' effectiveness in predicting rainfall occurrence.

In conclusion, the presented machine learning approach offers a promising solution for rainfall prediction, demonstrating the potential of leveraging atmospheric data and advanced algorithms to enhance forecast accuracy. Further research and refinement of models could lead to improved predictions, contributing to better preparedness for weather-related events