VIT

Vellore Institute of Technology

Oecmed io be University urvder section 3 ni' üGC Act 1956)

Rain Prediction Using Machine Learning (SDG – Goal 13)

B ShabareshwarReddy | P Siva Nandhan Reddy | Akash A| Dr. Asnath Victy Phamila Y | SCOPE

Motivation/ Introduction

Weather prediction, especially rainfall forecasting, is crucial for sectors like agriculture, disaster management, and urban planning. Traditional methods often lack accuracy due to the complexity of atmospheric processes. Machine Learning offers a data-driven approach that can model these complex relationships efficiently. This project explores how supervised ML algorithms, particularly Random Forest, can be leveraged to predict rainfall based on various meteorological parameters.

SCOPE of the Project

This project focuses on building a binary classification model to predict the likelihood of rainfall on a given day. Using historical weather data, the system aims to learn patterns and relationships among features like temperature, humidity, dew point, wind conditions, and sunshine duration. The ultimate goal is to provide an accurate, automated prediction system that supports decision-making in weather-sensitive fields..

Methodology

The dataset consists of 366 daily weather records and includes features such as temperature, dew point, pressure, humidity, wind speed, and direction. After preprocessing the data (including handling missing values and converting categorical rainfall labels into binary), Exploratory Data Analysis (EDA) was conducted using histograms, boxplots, and a correlation heatmap. Highly correlated variables were dropped to reduce redundancy. The class imbalance in the dataset (more rainy days than non-rainy ones) was addressed by downsampling the majority class. A Random Forest model was trained and optimized using GridSearchCV, followed by cross-validation to evaluate its generalization performance.

To enhance the **Methodology** section of your project poster for the rainfall prediction project, we can add specific details about the **GridSearchCV** process used in your Jupyter Notebook (Rainfall_Prediction_using_Machine_Learning.ipynb) and include a relevant formula to provide mathematical context. This will make the section more informative and technically robust, aligning with academic poster expectations. Below, I'll provide the updated content for the Methodology section, including additional information about GridSearchCV and a formula related to the Random Forest Classifier's hyperparameter tuning. I'll also suggest how to integrate this into your existing poster content and Word file preparation.

Understanding GridSearchCV in Your Project

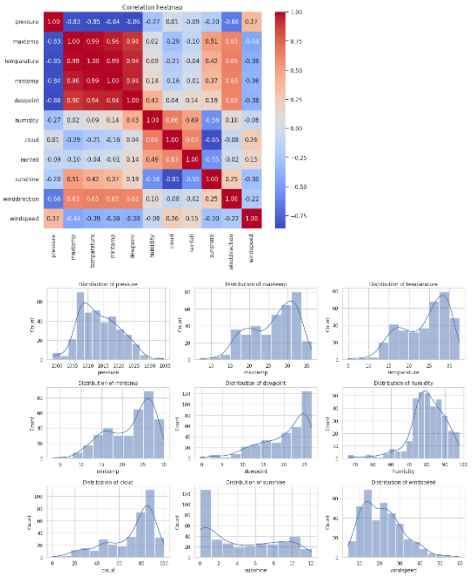
In your notebook, you used GridSearchCV to optimize the hyperparameters of the Random Forest Classifier. GridSearchCV systematically tests combinations of hyperparameters (e.g., n_estimators, max_depth, min_samples_split) to find the best-performing model based on cross-validation scores. The best parameters found were:

- n_estimators=50
- min_samples_split=10
- max_depth=None
- max_features='sqrt'
- min_samples_leaf=1

$$Gini(t) = 1 - \sum (p\_i^2)$$

Results

The best Random Forest model achieved a **test accuracy of approximately 74.5%**, with a precision of 77% for "no rainfall" and 72% for "rainfall" classes. Cross-validation yielded a mean score of **81.9%**, indicating good model stability. The model successfully predicted unseen data with reliable performance. Visualization techniques like seaborn plots were used extensively to understand feature distributions and interdependencies, while a confusion matrix and classification report provided detailed performance insights.



Conclusion/ Summary

This project demonstrates the feasibility and effectiveness of using machine learning models, particularly Random Forest, for rainfall prediction. The model showed robust performance and can be improved further with more data and additional features like geographical location and seasonality. Future work may include testing with real-time data and deploying the model as a web or mobile application for public use.

Contact Details

bheemi.shabareshwar2023c@vitstudent.ac.in

peta.siva2023@vitstudent.ac.in

akash.annam2023@vitstudent.ac.in

Acknowledgments/ References

Rainfall Dataset (sdg-13).

Tools used: Python, Pandas, Seaborn, Scikit-learn, MLxtend in Google Colab.