<RAINFALL PREDICTION USING MACHINE LEARNING-PYTHON>

Submitted for

Artificial Intelligence Machine Learning CSET301

Submitted by:

(<E23CSEU1142>) ABHIJEET SINGH

Submitted to

MR. SHWETANG DUBEY SIR (Add your Lab Faculty Name)

JAN-MAY 2025

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

INDEX

Abstract

This project aims to predict rainfall based on various meteorological features using machine learning algorithms. Leveraging classification models like Logistic Regression, XG Boost, and SVM, the study explores the effectiveness of different techniques in forecasting rainfall using historical data. Key aspects of preprocessing such as handling missing values, balancing imbalanced classes, and feature scaling are emphasized to enhance
model performance.

Introduction

Rainfall prediction is a critical task in meteorology and has direct implications for agriculture, disaster management, and water resource planning. Accurate forecasts help mitigate the effects of floods and droughts. This study utilizes machine learning approaches to model rainfall prediction from a real-world dataset, there by contributing to improved forecasting systems.

Related Works

Recent works in weather forecasting have employed machine learning techniques including Support Vector Machines (SVM), Decision Trees, and Neural Networks for classification tasks. Various studies emphasize the role of feature selection and class balancing in enhancing model accuracy. The integration of ensemble models like XG Boost has also been shown to outperform traditional methods in similar classification problems.

Problem Statement

The primary objective is to predict the binary outcome of whether rainfall occurs or not, based on numerical weather attributes. The dataset is imbalanced, with a skew toward non-rainy instances, making it a challenge to achieve balanced model accuracy across both classes.

Contribution

• Implemented thorough data preprocessing including handling missing values and renaming columns.
• Conducted Exploratory Data Analysis (EDA) with visualizations such as histograms, boxplots, and correlation heatmaps.
• Handled class imbalance using Random Over Sampler from the `imblearn` library.
• Applied feature scaling to normalize inputs.
• Trained and evaluated three classification models: Logistic Regression, XG Boost, and Support Vector Classifier.
• Compared model performance using ROC-AUC scores for both training and validation datasets.

Conclusion

The project successfully demonstrates the use of multiple classification algorithms for predicting rainfall. XG Boost and SVM showed relatively high validation performance compared to Logistic Regression. Balancing the data significantly improved model predictions, highlighting the importance of preprocessing in classification problems .Future work could include exploring time-

series models or deep learning architectures for further performance gains.

Future work

- Integrate more real-world features (e.g., humidity, wind speed).
- Use deep learning models like LSTM or GRU.
- Real-time predictions using live weather APIs.

References

1. Scikit-learn: Machine Learning in Python – Pedregosa et al. (2011)

2. XG Boost: A Scalable Tree Boosting System – Chen & Guestrin (2016)

3. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets – Lemaitre et al. (2017)

4. Seaborn and Matplotlib documentation for visualization tools.

GITHUB LINK

https://github.com/Abhi20037500/AIML-project