

ABSTRACT

This project explores the application of machine learning techniques for classifying weather conditions—Sunny, Cloudy, or Rainy—based on historical meteorological data from Basel. The dataset includes variables such as mean temperature, humidity, cloud cover, atmospheric pressure, and precipitation. To enhance prediction accuracy, new features like temperature classification and the product of humidity and cloud cover were engineered. Imbalanced class distributions were addressed through upsampling techniques to ensure robust model performance across all weather types. Multiple classifiers were implemented and evaluated, including Random Forest, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and XGBoost. Among these, ensemble models such as Random Forest and XGBoost demonstrated superior performance in predicting weather conditions. This study highlights the effectiveness of machine learning in weather classification and sets the foundation for more advanced forecasting models using real-time or more granular data.