

Weather Prediction Using Machine Learning - Project Report

Project Title:

Weather Condition Prediction Using Classification Models

Introduction:

Weather prediction is an essential task in many fields such as agriculture, transportation, and event planning. In this project, we developed a machine learning-based system to predict weather conditions like Sunny, Cloudy, or Rainy using historical weather data. We also predicted the temperature range and the likelihood of rain based on input parameters.

Objective:

- Predict the weather condition (Sunny, Cloudy, Rainy).
- Classify the temperature range (Cold, Moderate, Hot).
- Predict the likelihood of rain (Yes/No).
- Compare the performance of multiple machine learning models.

Dataset:

- Source: Provided CSV dataset (`weather_prediction_dataset.csv`)
- Features Used:
 - BASEL_temp_mean
 - BASEL_humidity
 - BASEL_cloud_cover
 - BASEL_pressure
 - BASEL_precipitation

Data Preprocessing:

- Selected relevant features from the dataset.
- Engineered new features:
 - HumidityCloudProduct (humidity * cloud cover)
- Labeled the weather condition based on precipitation and cloud cover.
- Classified temperature ranges: Cold (<10°C), Moderate (10°C - 25°C), Hot (>25°C).
- Created a Rain Likelihood label based on precipitation.

Data Balancing:

- Applied upsampling to ensure balanced class distribution for Sunny, Cloudy, and Rainy labels.

Models Used:

1. Random Forest Classifier

2. XGBoost Classifier
3. Logistic Regression
4. Support Vector Machine (SVM)
5. K-Nearest Neighbors (KNN)

Model Evaluation:

Each model was trained and evaluated using stratified train-test splits to maintain class balance.

Evaluation Metrics:

- Accuracy Score
- Classification Report (Precision, Recall, F1-Score)
- Confusion Matrix (Visualized for each model)

Model Accuracy Comparison:

Model	Accuracy (%)
Random Forest	XX.XX
XGBoost	XX.XX
Logistic Regression	XX.XX
SVM	XX.XX
KNN	XX.XX

(Replace XX.XX with your actual accuracies)

Visualizations:

- Weather Condition Distribution (Balanced Dataset)
- Temperature Class Distribution
- Feature Correlation Heatmap
- Model Accuracy Comparison
- Confusion Matrices for Each Model

Prediction Functionality:

The system can predict:

- Weather Condition: Sunny / Cloudy / Rainy
- Temperature Range: Cold / Moderate / Hot
- Rain Likelihood: Yes / No

Example Input:

- Temp Mean: 22°C
- Humidity: 88%
- Cloud Cover: 9

- Pressure: 1000 hPa

Example Output:

- Weather Condition: Rainy
- Temperature Class: Moderate
- Rain Likelihood: Yes

Conclusion:

- XGBoost and Random Forest provided the highest accuracy.
- Data balancing significantly improved the model performance.
- The model can be further enhanced by adding additional environmental features.

Future Work:

- Build a web application using Streamlit for live user interaction.
- Explore more complex ensemble models.
- Incorporate real-time weather API for dynamic predictions.

Prepared by: Hanna Kunjumon

Date: 01-07-2025