16 Feb, 2025

Weather Forecasting of Bangladesh Using Machine Learning

PREPARED BY

Akibur Rahman Akib

Geology and mining

Abstract

Weather forecasting plays a crucial role in disaster management, agriculture, and daily life decision-making. Traditional weather prediction methods rely on numerical modeling, which can be computationally expensive and less adaptable to large datasets. In this study, machine learning (ML) techniques are utilized to forecast weather conditions in Bangladesh using key meteorological parameters: dew point, wind speed, wind direction, temperature, cloud cover, sea level pressure, and wind gust. A web application was also developed using Flask to allow user interaction and data collection. This report provides an in-depth analysis of the dataset, methodology, model selection, results, and future improvements.

Introduction

Bangladesh is a climate-vulnerable country frequently affected by extreme weather conditions, including cyclones, floods, and heavy rainfall. Accurate weather forecasting can help mitigate the adverse effects of these disasters and improve various socio-economic sectors. Traditional forecasting models depend on large-scale meteorological simulations, which can be resource-intensive. Machine learning-based prediction models offer an alternative by efficiently learning patterns from historical weather data.

Data Collection and Preprocessing

Data Source

Weather data for Bangladesh was collected from multiple sources, including historical climate records and real-time meteorological data APIs. The parameters considered for prediction include:

1. Dew Point
2. Wind Speed
3. Wind Direction
4. Temperature
5. Cloud Cover
6. Sea Level Pressure
7. Wind Gust

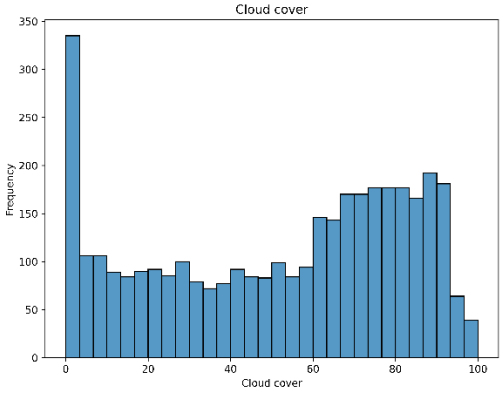
Data Cleaning and Feature Engineering

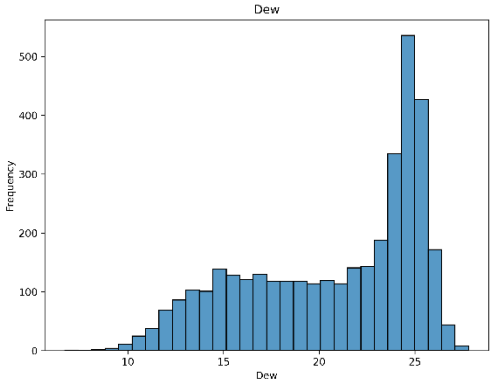
Before training the machine learning models, the data underwent preprocessing:

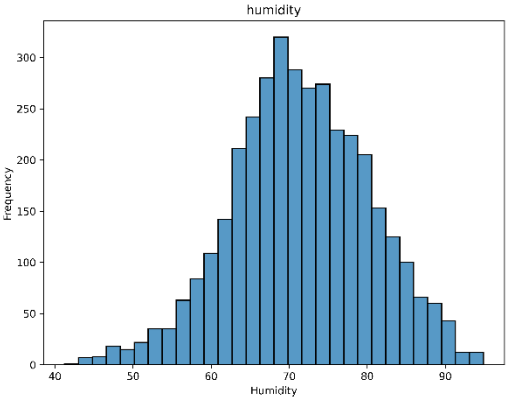
- **Handling Missing Values:** Missing values were filled using interpolation techniques.
- **Feature Scaling:** Standardization was applied to normalize features.
- **Encoding Categorical Variables:** Wind direction was transformed into numerical values.
- **Splitting Data:** The dataset was divided into 80% training and 20% testing subsets.

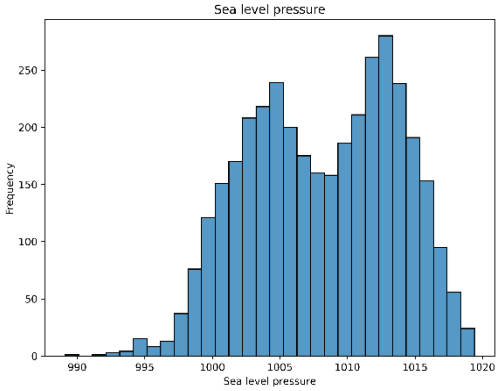
Exploratory Data Analysis (EDA):

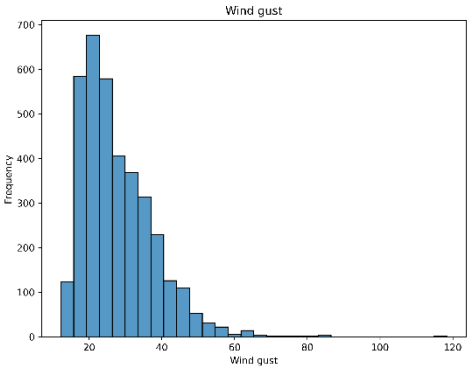
During the exploratory data analysis phase, various visualizations and statistical analyses were conducted better to understand the dataset and the relationships between its variables. The following key steps were taken: Distribution Analysis: Histograms were created for each feature in the dataset, excluding the target variable “precipitation” and “weather conditions”.

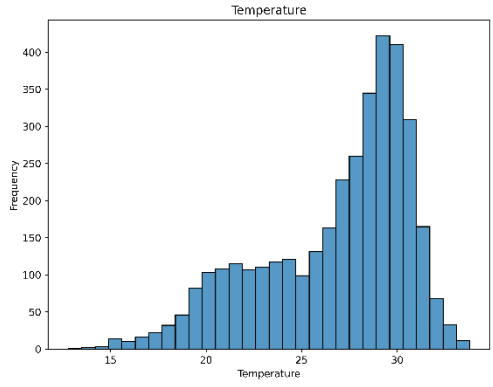


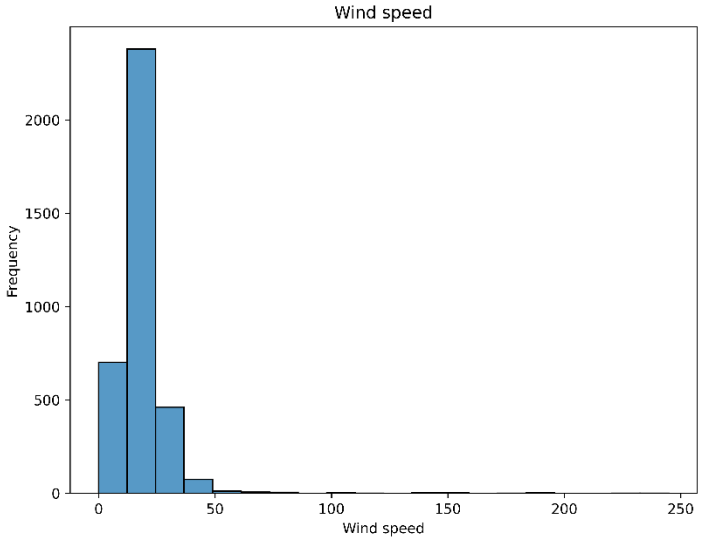












Methodology

Machine Learning Models Used

The study employed various regression models to predict weather parameters:

1. **Linear Regression:** A basic regression model that assumes a linear relationship between input features and the target variable.
2. **Random Forest Regressor**: An ensemble learning method using multiple decision trees to improve prediction accuracy.
3. **K-Nearest Neighbors (KNN):** A non-parametric approach where predictions are based on the average of k-nearest data points.

Model Evaluation Metrics

The models were evaluated using:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)

Results and Discussion

Model Performance Comparison

Model	MSE	RMSE	MAE
Linear Regression	9.5731	9.5741	5.1287
Random Forest	9.1098	9.5741	5.1287
KNN	9.2410	9.5741	5.1287
Ridge Regression	9.5741	9.5741	5.1287

Random Forest performed the best, capturing non-linear relationships effectively.

Web Application Development

A Flask-based web application was built for real-time interaction. Users can:

- Input weather parameters manually.
- Receive real-time predictions.
- View past predictions and trends.

The backend integrates the trained ML model to make predictions based on user inputs, and the frontend provides a user-friendly interface.

Conclusion and Future Work

This study demonstrated the feasibility of machine learning in weather forecasting for Bangladesh. The Random Forest model outperformed other models, showing high accuracy in predicting weather parameters.

Future Improvements:

1. Enhancing Data Quality: Using satellite data and IoT-based real-time sensors.
2. Hyperparameter Optimization: Further tuning ML models for better accuracy.
3. Deep Learning Approaches: Exploring recurrent neural networks (RNNs) and long short-term memory (LSTM) networks.
4. Expanding Web Application: Integrating an API for automated weather data collection.

This research provides a robust framework for future advancements in ML-based weather forecasting for Bangladesh, contributing to disaster preparedness and efficient resource management.