Machine Learning–Driven Temperature Forecasting Using NOAA Weekly Weather Data

Presented by: Ch. Sowmya Sree
Roll No: AP23110010053
Guided by: Dr. Anusha Nalajala

Introduction

- Weather forecasting is essential for agriculture, aviation, transport, and energy.

- Temperature plays a major role in understanding climate conditions.

- Dataset used: **NOAA CORGIS Weather Dataset (2016)**.

- Contains weekly data: temperature, rainfall, wind, and locations.

- Goal: Build a **machine learning model** to predict average weekly temperature.

Problem Statement

- Traditional forecasting methods can be slow and struggle with non-linear weather patterns.
- Large datasets require faster and more accurate prediction methods.

Need an ML model that can:

- Handle missing/inconsistent data
- Work with mixed features
- Predict temperature reliably

Objective: Build a **data-driven model** to forecast weekly temperature.

Dataset Overview

- **Total Rows:** 16,743

- **Total Features (Columns):** 14

- **Target Variable:** data_temperature_avg_temp

- **Number of Unique Cities/Places:** 122 Weather Forecast Offices (WFOs)

- **Main Features:**
  data_precipitation, data_temperature_avg_temp, data_temperature_max_temp, data_temperature_min_temp, data_wind_direction, data_wind_speed, station_city, station_state, station_location, date_full, date_month.

Methodology

Data Preprocessing:
 Cleaned data, removed duplicates, handled missing values, renamed columns for consistency, extracted date features (year, month, week)

Feature Engineering:
 Selected numeric features, dropped irrelevant columns, kept only useful variables for prediction

Model Selection:
 Used Random Forest Regressor and XGBoost Regressor for temperature forecasting

Methodology

Train–Test Split:
 Split data into 80% training and 20% testing

Model Training:
 Trained both ML models using selected features to predict average weekly temperature

Evaluation Metrics:
 Used R², MAE, RMSE to compare accuracy and performance

Visualization:
 Created Feature Importance, Actual vs Predicted, and Residual plots

Model Comparison (Random Forest vs XGBoost)
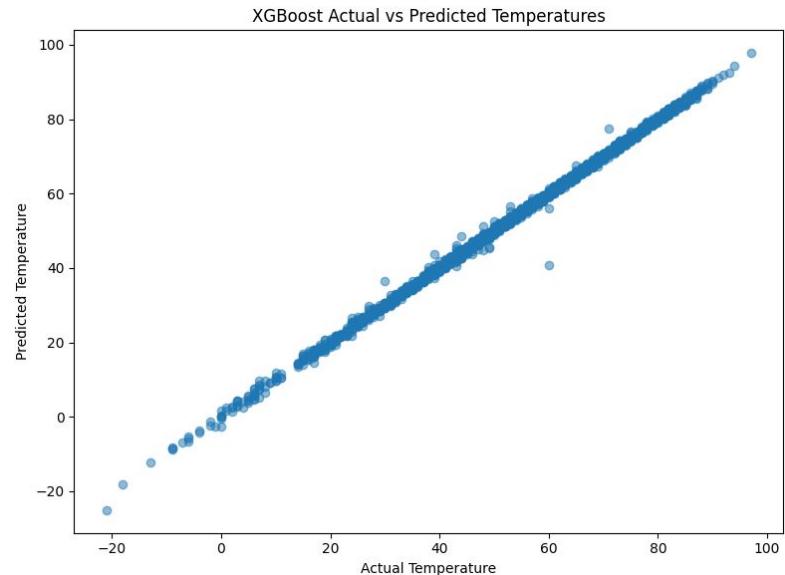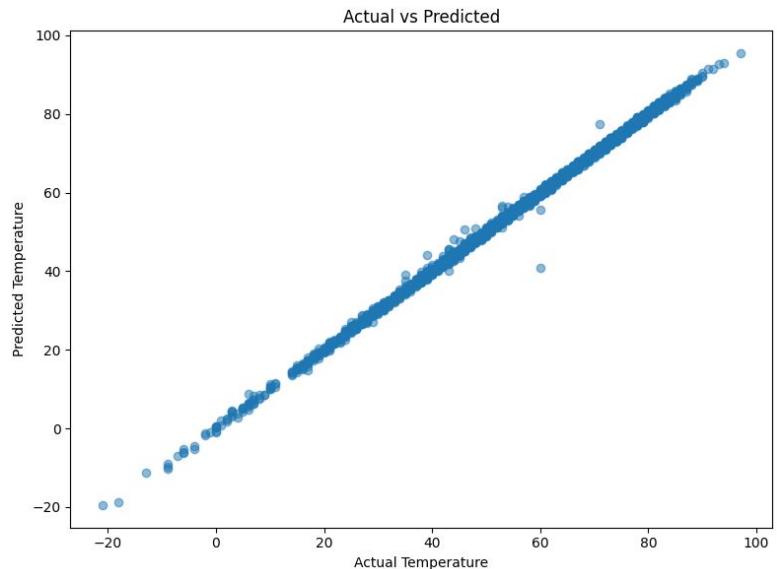
Metric	Random Forest	XGBoost
R²	0.9989	0.9988
MAE	0.327	0.405
RMSE	0.609	0.663

Model Comparison (Random Forest vs XGBoost)

Best Model: Random Forest

- **Handles Noise Better:** Random Forest reduces overfitting by averaging many decision trees.
- **Dataset Size is Small:** RF works very well on smaller/medium datasets like NOAA 2016.
- **Temperature Features are Simple:** RF captures relations between Min Temp, Max Temp, Wind easily and also Random Forest is better because it gives higher accuracy, lower error, and more stable predictions for this dataset.

Scatter Plot Actual vs Predicted

Conclusion

- Machine learning successfully predicted weekly temperatures using the NOAA CORGIS dataset.
- Random Forest proved to be the most accurate and reliable model.
- Model achieved very high accuracy with minimal errors.
- ML-based forecasting is fast, scalable, and more efficient than traditional methods.
- This project shows how ML can support environmental and climate prediction tasks.

Thank You