Weather Forecasting Report

1. Introduction

1.1 Overview of Weather Forecasting

Weather forecasting plays a critical role in various sectors, including agriculture, aviation, disaster management, and urban planning. Accurate predictions enable better decision-making and preparedness for extreme weather conditions. This report details a comprehensive weather forecasting study using time series analysis and anomaly detection techniques.

1.2 Objectives

- Perform exploratory data analysis (EDA) to uncover trends, patterns, and seasonality.
- Develop and evaluate a forecasting model using Facebook Prophet.
- Detect anomalies using STL + ESD to identify data inconsistencies or extreme weather events.
- Assess model performance and suggest improvements for better accuracy.

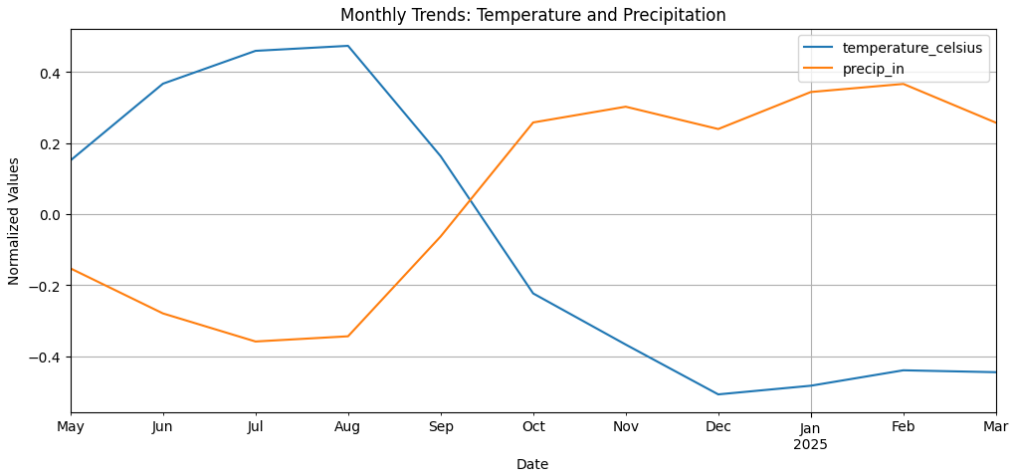
2. Data Preprocessing

2.1 Dataset Description

The dataset contains weather records with 41 attributes, including temperature, precipitation, wind speed, pressure, humidity, visibility, and air quality metrics. The primary timestamp column is `last_updated`, which represents time intervals of 15 minutes.

2.2 Data Cleaning

- **Original Shape:** (60,218, 41)
- **Shape after Removing Outliers:** (25,425, 41)
- **Handling Missing Values:**
    - Dropped columns with >40% missing values.

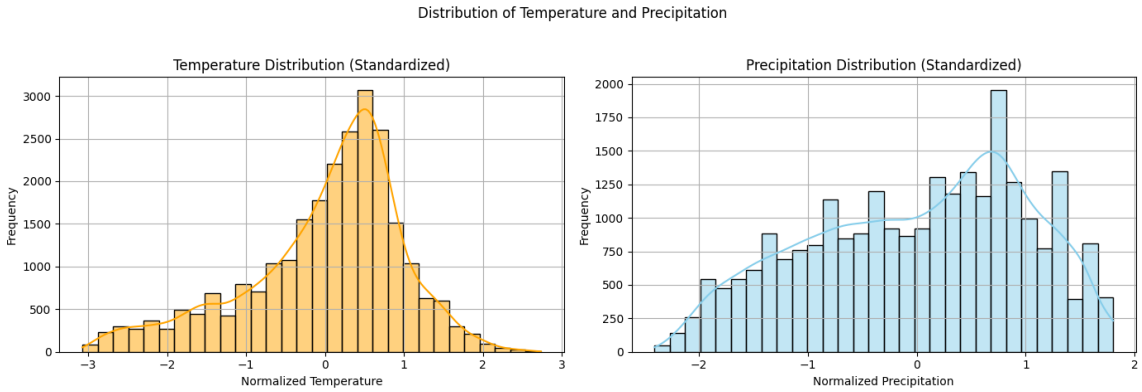


- Filled numeric values with medians and categorical values with modes.

2.3 Feature Engineering

- **Normalization:** StandardScaler applied to numeric columns to ensure consistency.
- **New Variables Created:** Derived features such as daily temperature change rate and rolling averages to capture trends.

3. Exploratory Data Analysis (EDA)



3.1 Seasonality and Trends

- Monthly average temperature analysis showed significant seasonal variations.
- Wind speed and precipitation levels exhibited cyclical trends, possibly influenced by monsoon seasons.

3.2 Geographical Visualization

- Temperature and precipitation variations were mapped across different regions.
- Hotspots of extreme weather conditions were identified using heatmaps.

4. Forecasting Model: Facebook Prophet

4.1 Why Prophet?

Facebook Prophet is an additive time series model effective in handling missing values, seasonality, and trend shifts. It provides intuitive parameter tuning and works well with irregular data.

4.2 Model Training

- **Training Data:** Used `last_updated` as the timestamp.
- **Target Variables:** `precip_in` and `temperature_celsius`.
- **Hyperparameter Tuning:**
  - Adjusted seasonal parameters for better fitting.
  - Incorporated holiday effects to improve forecasting accuracy.

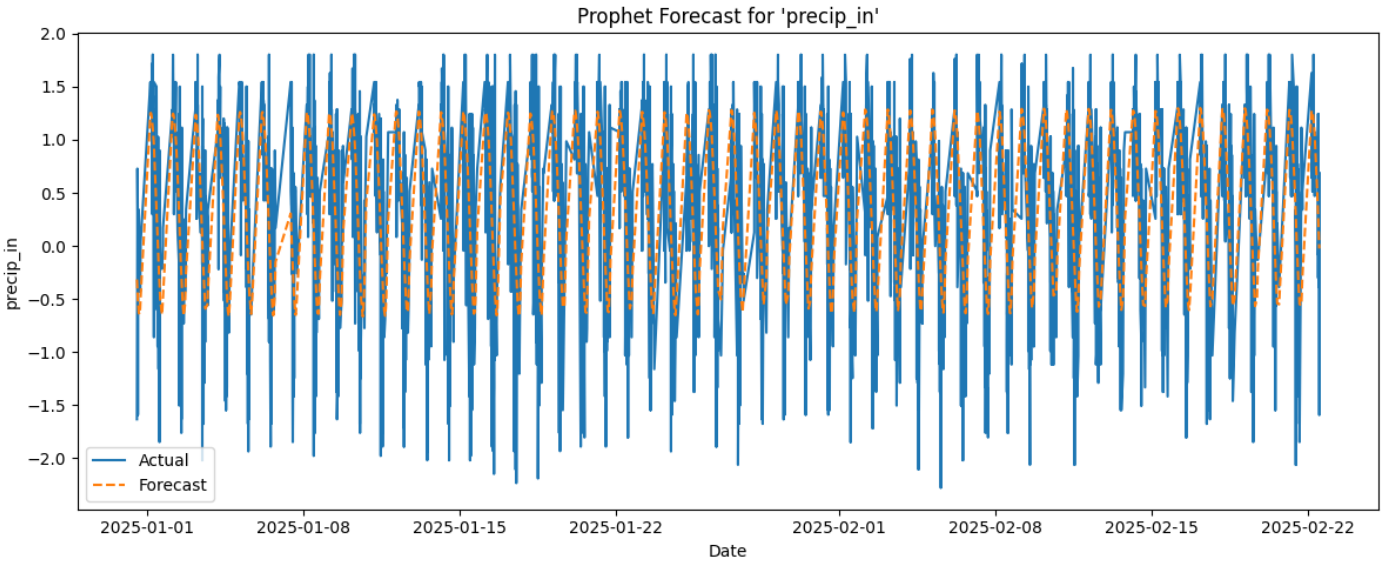
5. Model Evaluation

5.1 Performance Metrics

🔍 Evaluation for 'precip_in':

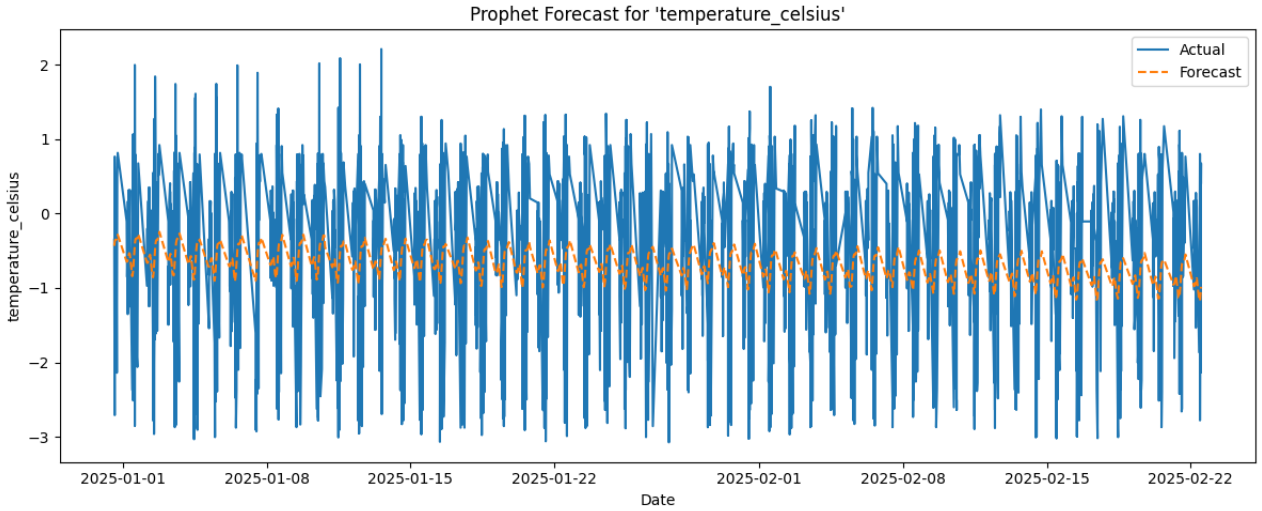
- MAE: 0.5466

- RMSE: 0.6936
- MAPE: 340.69%
- R^2: 0.3761



🔍 Evaluation for 'temperature_celsius':

- MAE: 0.9612
- RMSE: 1.1167
- MAPE: 397.81%
- R^2: -0.0112



5.2 Interpretation

- **Temperature Forecasting:** Poor R^2 value (-0.0112) suggests weak predictive performance, indicating high volatility in temperature changes.
- **Precipitation Forecasting:** Moderate performance, though the high MAPE indicates forecasting instability.

6. Anomaly Detection (STL + ESD)

6.1 Methodology

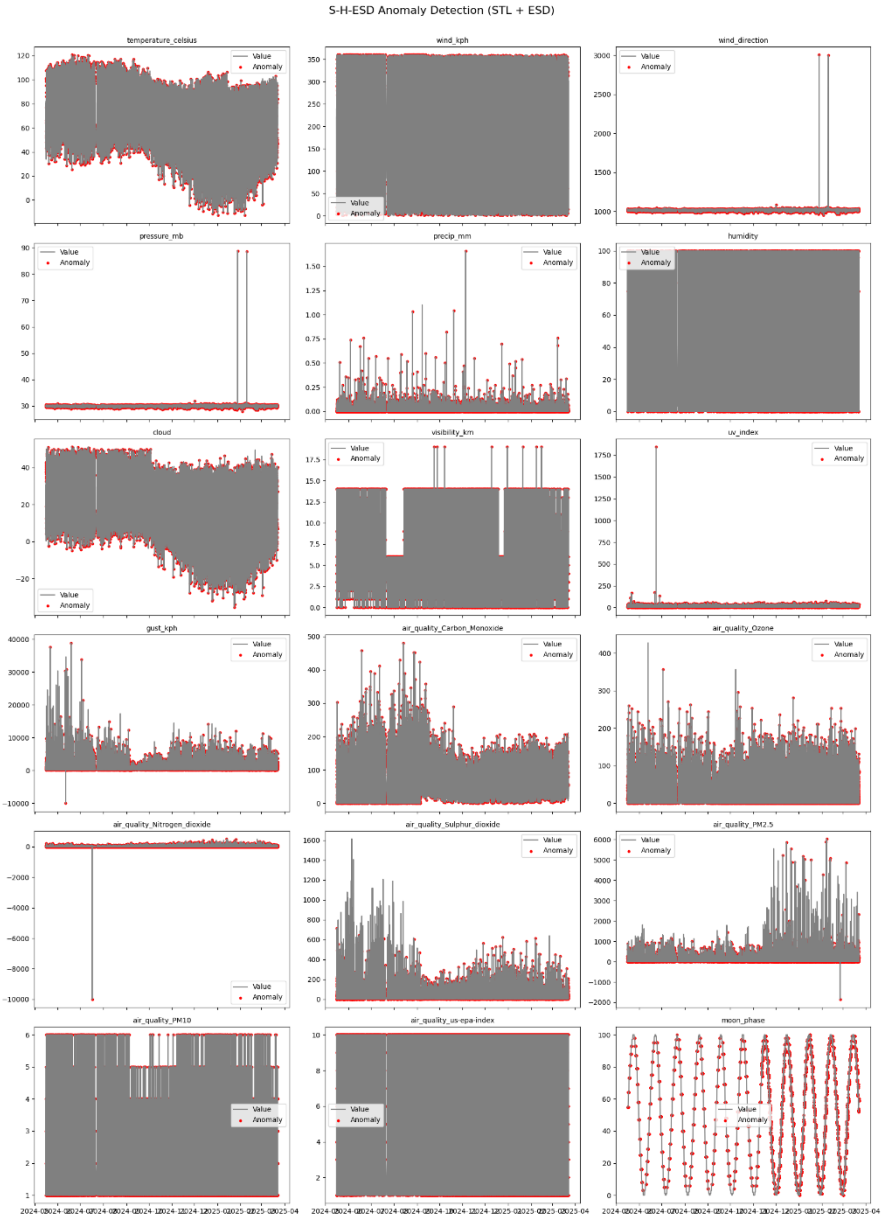
- **STL Decomposition:** Breaks time series into trend, seasonal, and residual components.
- **ESD Test:** Detects extreme values by identifying statistical outliers in the residuals.

6.2 Findings

- High spikes observed in `wind_kph`, `pressure_mb`, and `air_quality_PM10`.
- Irregular patterns in `visibility_km` and `cloud` suggest possible sensor errors or extreme weather conditions.

6.3 Visualization

- Red dots indicate detected anomalies across multiple weather variables.
- Certain anomalies may be due to faulty sensor readings rather than real environmental changes.



  - Detected anomalies in key weather variables such as temperature, wind speed, pressure, precipitation, humidity, and air quality indices.
  - High spikes observed in `wind_kph`, `pressure_mb`, and `air_quality_PM10`.

- Irregular patterns in `visibility_km` and `cloud` indicate possible sensor errors or extreme weather conditions.

7. Insights & Recommendations

7.1 Key Findings

- **Temperature variations are highly volatile, affecting forecast accuracy.**
- **Precipitation models need further tuning due to high error rates.**
- **Anomalies in wind speed, pressure, and air quality metrics indicate potential sensor errors or extreme weather conditions.**

7.2 Potential Improvements

- **Use ensemble models** (ARIMA, SARIMA, XGBoost) to improve forecasting performance.
- **Include multivariate factors** (humidity, pressure) to enhance predictive accuracy.
- **Enhance data quality** by addressing sensor anomalies before model training.

8. Next Steps

8.1 Enhancing Forecast Accuracy

- Compare Prophet with additional models.
- Tune hyperparameters to optimize Prophet's performance.
- Integrate additional features such as atmospheric pressure trends.

8.2 Final Deliverable

- **Comprehensive Report:** Includes all analyses, evaluations, and visualizations.
- **Github Repository:** Contains code, datasets, and documentation.