Weather Trend Forcasting

PM Accelerator mission : “Welcome to PM Accelerator , a beacon of guidance for aspiring and experienced PMs alike . We’ve designed this platform to offer training, education, and job opportunities for product Managers , creating room for constant improvement and shaping the next generation of PM’s. Whether you are a newbie guaranteed to find a new opportunity with the help of PM Accelerator. "

Data Cleaning and Preparation

- **Duplicates & Missing values:** Checked for duplicate timestamp and redundant entries and removed them to ensure data integrity . Any missing data point were addressed by interpolation or by omitting those record.
- **Outlier Detection :** Outlier were identified using statistical methods. Extreme value in features like wind speed or pressure were flagged . Depending on context , such outliers were either retained or replaced/ removed to avoid skewing models.

Exploratory Data Analysis (EDA)

- **Statistical Summary :** Computed descriptive statistics for key variables (temperature, humidity, wind speed etc) to understand central tendencies and spread. This helps variability and detect potential data issues.
- **Distribution and Trend Visualization:** Plotted histograms and boxplot for each feature to visualize their distributions and spot anomalies. Time-series line charts of temperature revealed clear seasonal patterns - for example , winter months had lower temperature than summer . These seasonal trends are critical for selecting appropriate forecasting models.
- **Feature Relationships:** Used correlation metrics and heatmap to explore inter-variable inter-relationships. For instance , a strong positive correlation between temperature and dew point was observed. Such insights suggest which feature might serve as predictors in multivariate models.

Forecasting Models

- **Classical Methods:** Applied ARIMA/ SARIMA and exponential Smoothing models to capture linear trends and seasonality. ARIMA is effective for short- term stationary series while handling periodic patterns.
- **Machine Learning Models:**Tested Regression based methods on lagged features and exogenous inputs. These can model nonlinear relationships if carefully tuned. TrainInData's guide notes that time-series can be turned into regression problems by engineering lag and time features. We also considered Facebook Prophet for it's ease with multiple seasonality and holiday effects.
- **Model Comparison:** All models were implemented and their forecasts visualized. For transparency , baseline forecasts (like seasonal naive or linear drift) were also computed to benchmark advanced models. Model hyperparameters were tuned

Model Evaluation

- **Metrics :** Forecast accuracy was quantified using metrics : MAE, RMSE and MAPE , which are common in time series studies. These metrics give absolute , squared , and percentage error perspective respectively.

- **Forecast Plots :** Plotted actual vs predicted values over time to visually assess performance (forecasts should track real trends). Also examined multi- step forecast error growth (error typically increases for no longer horizons)

Model Performance	
Models	Performance
Logistic	accuracy->69.96
Decision Tree	accuracy->85.43
Random Forest	accuracy->90.20
ARIMA	MAE-> 4.23
Prophet	MAE->28.92

Analysis & Insights

- **Exogenous Features:** Incorporated other weather variables (humidity, wind speed, pressure) and calendar features (day-of-week, month) as predictors. Including these exogenous inputs can improve forecasts, since they carry additional information**.** For example, knowing humidity helped fine-tune temperature predictions due to their high correlation.

- **Insights from Patterns:** The analysis reaffirmed that temperature and humidity are strongly seasonally linked: winter brought cooler, more humid conditions, while summer was hotter/drier. Identifying this guides the choice of seasonal cycles in models. Also, strong correlations (e.g. temperature–dew point) suggest multivariate models or engineered features can boost accuracy. Outliers in wind or pressure may indicate extreme weather events; recognizing them in EDA allows special handling (e.g., anomaly modeling)

CONCLUSION

This project demonstrates a complete end-to-end machine learning workflow:

1. Data Cleaning
2. EDA
3. Feature Engineering
4. Model Training
5. Model Comparison
6. Insight Generation

The structured approach aligns with PM Accelerator's mission of applying analytical thinking and real-world problem-solving techniques.

The final results highlight the importance of model selection, proper preprocessing, and thoughtful data analysis in achieving reliable predictive performance.