Project - Time-Series Analysis and Forecasting of AQI in Delhi

Group Members:

Vishal Kumar - 22PCS5132

Ravi Kumar - 22PCS5151

Vineet Kumar - 22PCS5131

Course – B.Sc. Physical Science with Computer Science

Subject - Data Mining for Knowledge Discovery

Guided By: Sanjeet sir

Abstract

This project investigates the air quality trends in Delhi by analysing the concentrations of PM2.5 and PM10—two critical pollutants affecting human health. Using Facebook's Prophet model, historical data was processed to identify seasonal and long-term trends and predict future values. Results demonstrated the model's capability to forecast pollution levels, providing valuable insights for urban planning and public health initiatives.

---

1. Introduction

Air pollution has emerged as a significant environmental and health issue in urban areas, particularly in Delhi, where AQI levels frequently surpass hazardous thresholds. Among various pollutants, PM2.5 and PM10 are critical as they pose severe respiratory and cardiovascular risks.

This project aims to analyse historical PM2.5 and PM10 data using time-series techniques to forecast future pollution levels. The study leverages Facebook's Prophet Model for trend forecasting, offering an intuitive approach to model seasonality and irregular data patterns.

---

2. Literature Review

Research in AQI prediction often employs statistical models like ARIMA and machine learning techniques. While these methods are effective, they require rigorous parameter tuning and struggle with irregularities in data. Prophet, developed by Facebook, offers a robust alternative, excelling in cases with strong seasonality and missing data.

---

3. Methodology

3.1 Dataset and Data Source

The dataset was sourced from the **Central Pollution Control Board (CPCB)**'s online platform, specifically through their **Advanced Search** feature. The data includes hourly averages of PM2.5 and PM10 levels in Delhi, spanning [1 Jan 2021 – 20 Nov 2024].

3.2 Data Preprocessing

Steps taken for data preparation:

1. **Feature Selection**: Selected only the relevant attributes such as Date, PM2.5 and PM10.
2. **Data Cleaning:** Removed missing or inconsistent entries.
3. **Exploratory Analysis:** Visualized trends and seasonal patterns to understand data characteristics.

3.3 Model: Prophet Model developed by Facebook (FB Prophet Model)

The Prophet model was selected for its:

- **Ease of Use:** Minimal parameter tuning required.
- **Seasonality Handling:** Built-in mechanisms for capturing yearly and weekly trends.
- **Robustness:** Effective even with missing data or outliers.

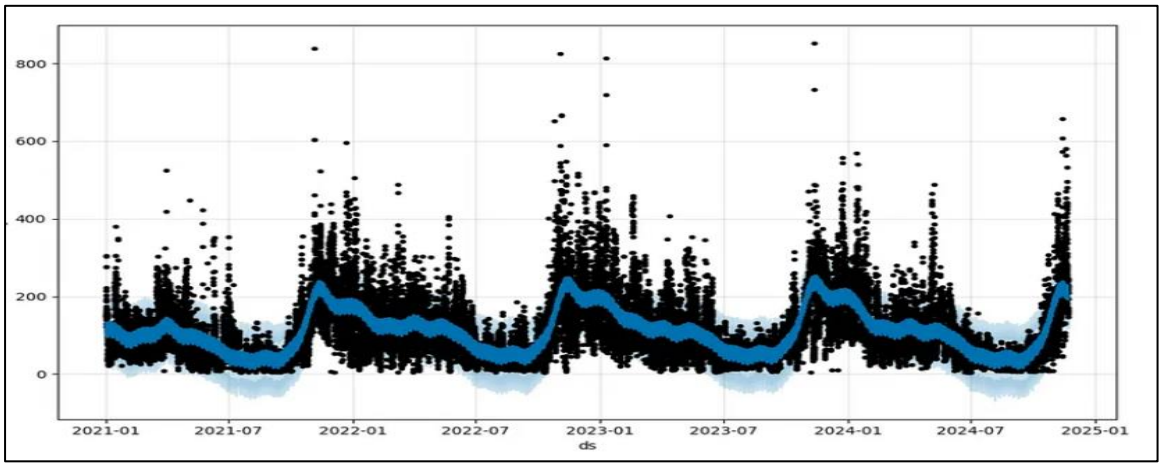
3.4 Implementation

1. **Model Training:** Separate models were trained for PM2.5 and PM10.
2. **Trend Analysis:** Long-term trends were extracted to identify overall pollution patterns.
3. **Forecasting:** Forecasted values for the next day were generated.

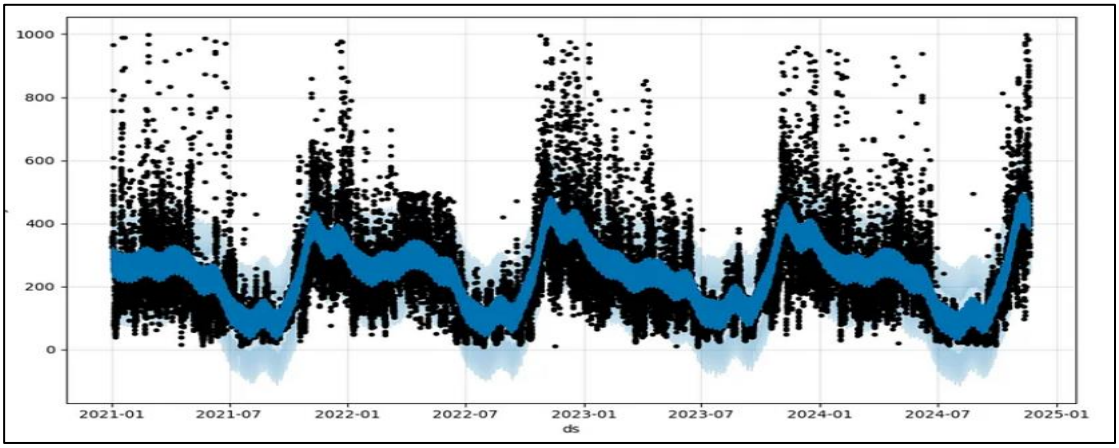
---

4. Results

Trend Analysing of PM2.5



Trend Analysing of PM10



4.1 Exploratory Data Analysis

- PM2.5 and PM10 levels exhibited clear seasonal spikes, particularly during winter months and festive periods (e.g., Diwali).
- A steady increase in pollution levels was observed over the years, correlating with urbanization and vehicular emissions.

4.2 Forecasting values of AQI (21 Nov 2024)

PM2.5

	ds	yhat	yhat_lower	yhat_upper
32058	2024-11-21 00:00:00	223.864839	159.972692	299.432349
32059	2024-11-21 01:00:00	219.140368	153.317157	293.176714
32060	2024-11-21 02:00:00	214.809537	143.277243	279.968458
32061	2024-11-21 03:00:00	210.991376	138.740991	285.876456
32062	2024-11-21 04:00:00	206.841370	134.619364	277.302504
32063	2024-11-21 05:00:00	202.447330	138.245382	275.509044
32064	2024-11-21 06:00:00	199.556678	130.168921	272.779043
32065	2024-11-21 07:00:00	200.204422	123.344059	270.705839
32066	2024-11-21 08:00:00	204.516407	136.531047	276.517361
32067	2024-11-21 09:00:00	209.929669	141.459116	284.689944
32068	2024-11-21 10:00:00	212.798691	138.619121	285.277158
32069	2024-11-21 11:00:00	211.071713	137.416143	277.189449
32070	2024-11-21 12:00:00	205.671075	130.677155	277.951940
32071	2024-11-21 13:00:00	199.430212	128.216012	269.660932
32072	2024-11-21 14:00:00	194.726123	115.301232	266.118373
32073	2024-11-21 15:00:00	192.061797	120.889445	260.473533
32074	2024-11-21 16:00:00	190.744909	118.516477	265.286534
32075	2024-11-21 17:00:00	190.638967	117.192769	261.451890
32076	2024-11-21 18:00:00	192.924835	116.463208	258.090072

PM10

	ds	yhat	yhat_lower	yhat_upper
32058	2024-11-21 00:00:00	447.677710	311.822151	581.814735
32059	2024-11-21 01:00:00	434.565645	301.480698	560.273805
32060	2024-11-21 02:00:00	422.773275	283.387860	568.403383
32061	2024-11-21 03:00:00	411.311134	279.041050	543.420141
32062	2024-11-21 04:00:00	398.841650	260.968004	531.492353
32063	2024-11-21 05:00:00	387.422786	260.297126	519.605426
32064	2024-11-21 06:00:00	382.114629	245.943708	519.240913
32065	2024-11-21 07:00:00	386.556198	254.427513	520.085202
32066	2024-11-21 08:00:00	398.840859	273.668562	536.931690
32067	2024-11-21 09:00:00	411.929227	279.993436	548.251268
32068	2024-11-21 10:00:00	418.526605	291.141859	553.759319
32069	2024-11-21 11:00:00	416.093753	271.878730	548.430603
32070	2024-11-21 12:00:00	407.638148	277.261750	539.401174
32071	2024-11-21 13:00:00	398.087056	255.883995	524.750549
32072	2024-11-21 14:00:00	390.183219	258.997706	533.107624
32073	2024-11-21 15:00:00	383.880733	246.089073	523.524706
32074	2024-11-21 16:00:00	379.255924	233.872926	509.894337
32075	2024-11-21 17:00:00	379.126941	237.321659	512.559906
32076	2024-11-21 18:00:00	387.855731	262.595027	527.329543

4.3 Forecasting Results

- Forecasted PM2.5 levels indicate a peak during [Nov - Jan].
- Forecasted PM10 levels indicate a peak during [Nov - Feb].
- The model captured both long-term trends and seasonal variations effectively.

---

5. Discussion

The results confirm the utility of the Prophet model in AQI forecasting, especially for pollutants like PM2.5 and PM10. The insights gained include:

- A significant upward trend in pollution levels during winter.
- Seasonal variations align with weather patterns and anthropogenic activities.

Challenges:

- Handling abrupt pollution spikes due to unforeseen events.
- Limited data granularity for real-time forecasting.

---

6. Conclusion

This study highlights the effectiveness of the Prophet model in analysing and forecasting air quality in Delhi. The model successfully captured trends and seasonal patterns, providing actionable insights for policymakers. These forecasts can aid in implementing targeted measures to mitigate air pollution, especially during high-risk periods.

---

7. References

1. Central Pollution Control Board (CPCB), *Advanced Search Platform*, Dataset Link.
2. https://facebook.github.io/prophet/docs/quick_start.html#python-api
3. Time Series Analysis using Facebook Prophet - GeeksforGeeks.
