AIR QUALITY
MONITORING SYSTEM

Mentor : Dr. Mainak Thakur

B25MT01

1. About the Project
2. Dataset Description and Analysis
3. Modeling
4. Health Recommedations
5. Challenges and Outcomes
6. Timeline

Why Study Air Pollution?

Air pollution has become a major global concern, affecting individuals with pre-existing conditions such as asthma, bronchitis, or cardiovascular diseases also millions of people.

Air quality monitoring systems are essential for creating healthier cities, reducing disease burden, and combating climate change. Real-time data and forecasting empower individuals, industries, and policymakers to take informed actions toward cleaner air.

DISASTERS CAUSED

According to the State of Global Air 2024 report, air pollution caused 8.1 million deaths in 2021, making it the second most important risk for mortality, especially among children under five. Every year, 6.7 million people lose their lives prematurely due to the combined impact of poor indoor and outdoor air.

Treat air pollution as public health emergency

Air pollution is linked to one in nine deaths around the world. In India, over a billion people live in areas where dangerous particulate matters exceed national quality standards. We must treat it as a public health emergency

Air pollution killed 1.5 million annually from 2009-2019

A study said that the entire 1.4 billion population of India live in areas having PM2.5 levels higher than the WHO-recommended yearly average of 5 micrograms per cubic metre.

A whopping 1.5 million deaths were reported in India annually from 2009 to 2019 due to long-term exposure to PM2.5 pollution every year, said the latest Lancet study.

Source : NEW INDIAN EXPRESS

Literature Review

Research Paper	Model	Performance Result
Multivariate Forecasting of Multiple Pollutants with Representative Deep Learning Architectures, S. Bianco(1), IEEE RTSI 2024 (8th Forum on Research and Technologies for Society and Industry Innovation), DOI: 10.1109/RTSI61910.2024.10761197.	BI-LSTM	Performed better than other models
Multivariate Spatial Prediction of Air Pollutant Concentrations with INLA Wenlong Gong(1), Published in Environmental Research Communications, October 2021 | DOI: 10.1088/2515-7620/ac2f92	BMSM with INLA	Multivariate model out performed univariate models
Mohammadi, F.(1). Prediction of atmospheric PM2.5 level by machine learning techniques in Isfahan, Iran. Scientific Reports, 14, 2109.	ANN	Deep learning method out performed other ML models
Z. Wang(1), Impact of Air Pollution on Daily Respiratory Hospitalization in Lanzhou, China, 2010 4th International Conference on Bioinformatics and Biomedical Engineering, Chengdu, China, 2010, pp. 1-4, DOI: 10.1109/ICBBE.2010.5514806.	GAM	GAM outperformed simpler models in assessing air pollution's impact
Bera, B., Bhattacharjee(1). PM2.5 concentration prediction during COVID-19 lockdown over Kolkata metropolitan city, India using MLR and ANN models. Environmental Challenges, 4, 100155.	MLR & ANN	ANN out performed MLR. Other pollutants along with meteorological factors were used.

Where can we improve?

- GNNs can model complex spatial relationships between multiple locations
- By capturing interactions between variables across space and time, GNNs improve forecasting accuracy.
- Can estimate pollutants like PM2.5, NO_2, SO_2, etc., at unmeasured locations based on nearby sensor data.

PROBLEM STATMENT

To implement accurate temporal forecasting and spatial gap filling models with health recommendations.

Temporal forecasting is predicting the values of variable over time.

- Here we wish to get the predictions of pollutants for the next Meteorological Stations.

We get the temporal forecasting values for particular regions. In spatial Gap Filling, we use these predictions and the previous patterns to estimate the pollutants at each location.

Health recommendations help assess the risks associated with various pollutants and provide guidance on minimizing exposure.

02.
Data Description and Analysis

Ground Station Data Description

Ground station data of the Chennai state containing 6 locations data for the years 2023 - 2024.

Main Attributes include

- Latitude, Longitude
- Solar Radiation
- Relative Humidity (%)
- Wind Speed (m/s)
- Wind Direction (deg)

Contains NULL Values

Data Source

- CPCB WEBSITE
- AQICN
- Format : CSV

Data Cleansing (Ground Station Data)

Outlier Detection & Removal

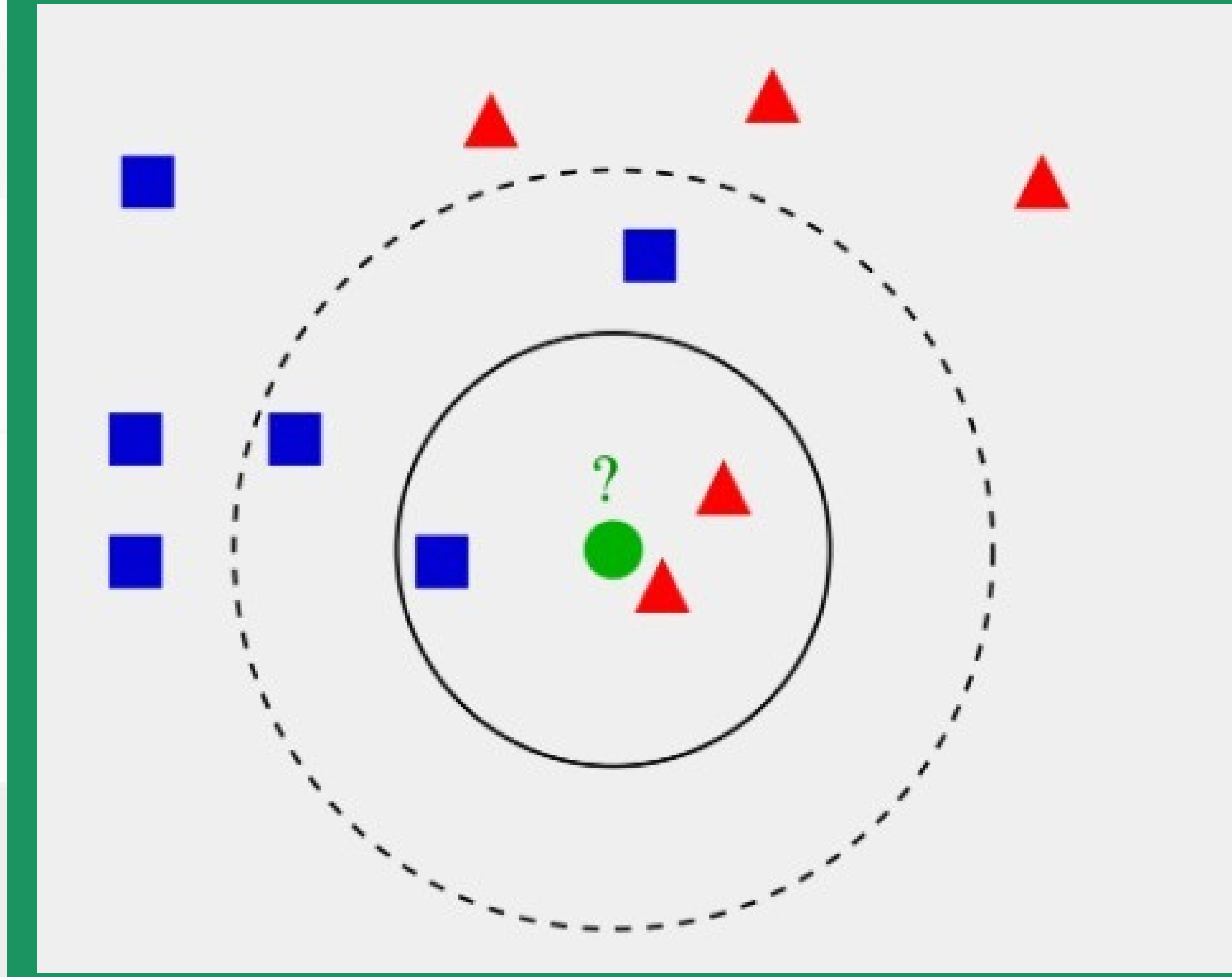
- Detect outliers using Z-score & IQR
- Remove or cap extreme values to maintain data integrity

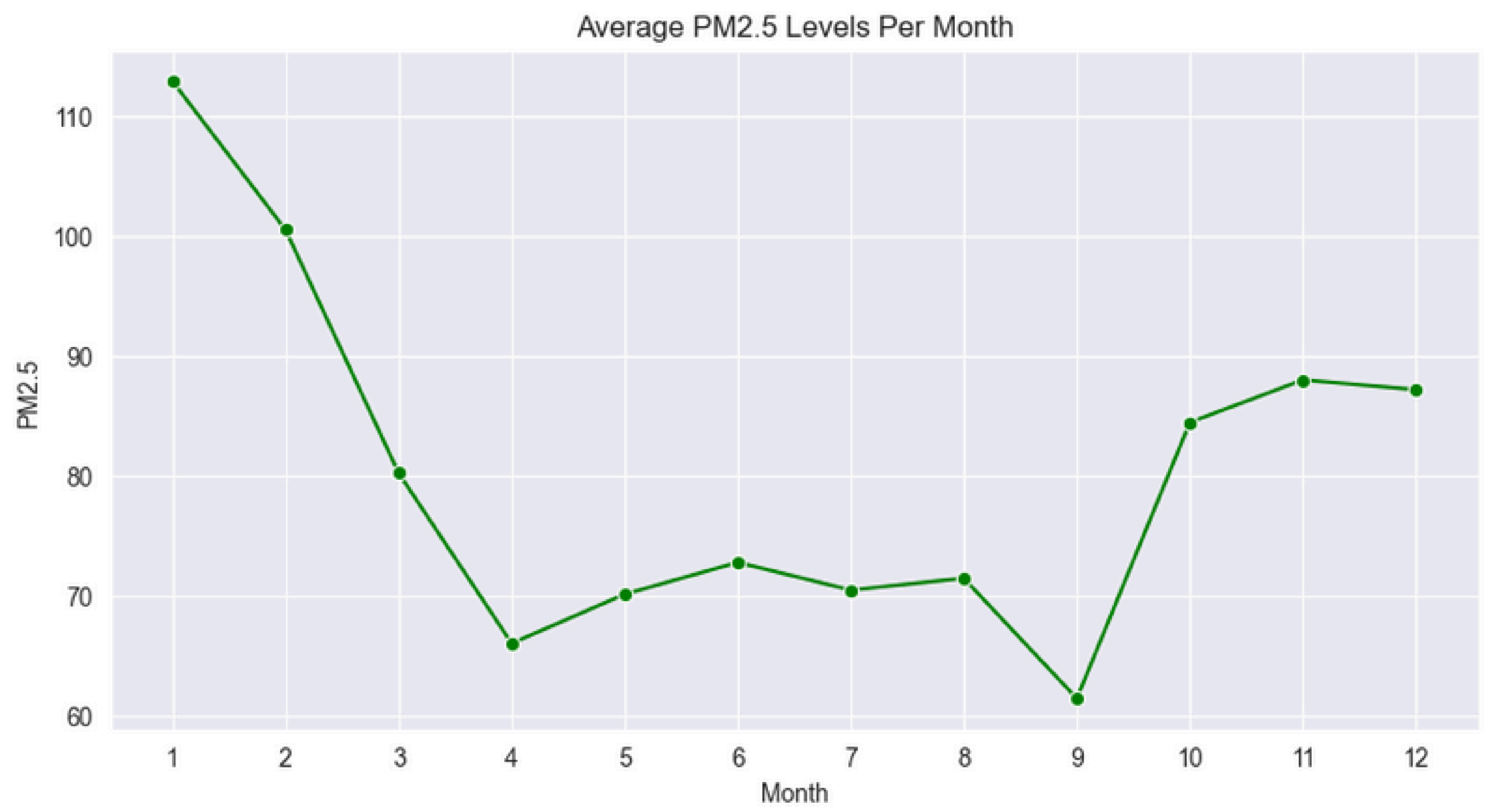
Skewness Correction

- Use Box-Cox & log transformation
- Helps normalize skewed data for better model performance

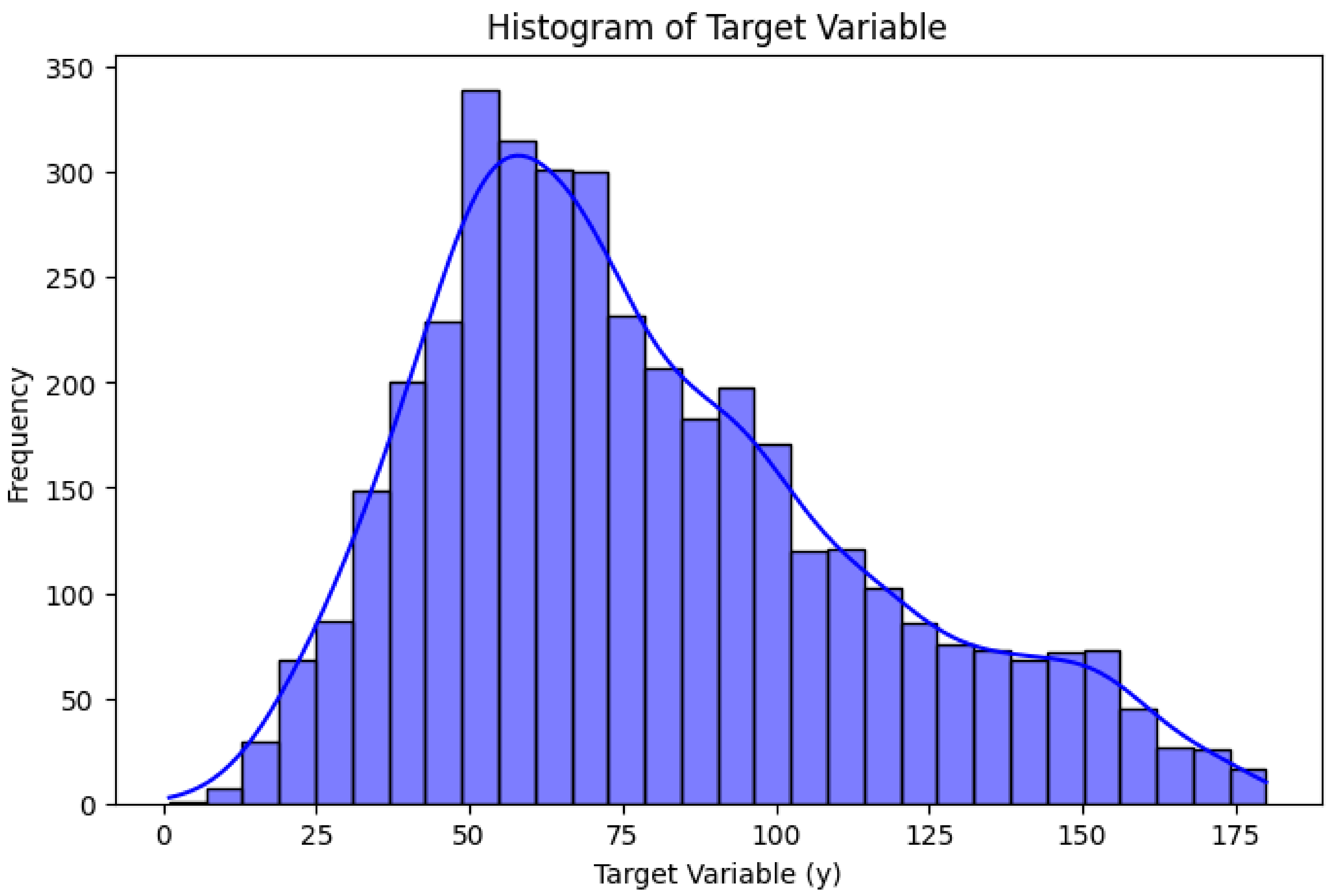
Null Value Handling (KNN Imputer)

- Fills missing values using K-Nearest Neighbors (KNN)
- Preserves patterns by imputing based on similar data points





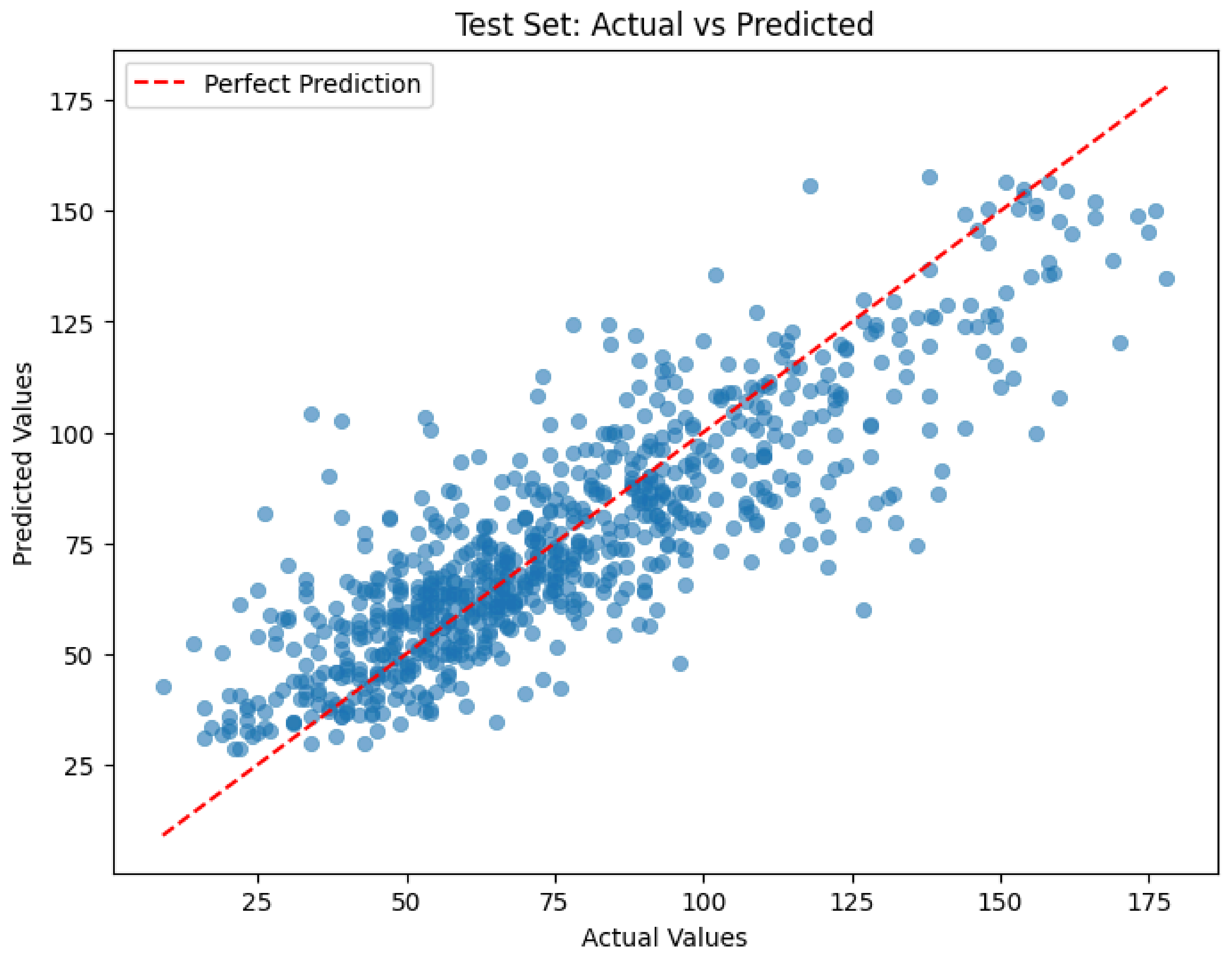
Data Analysis

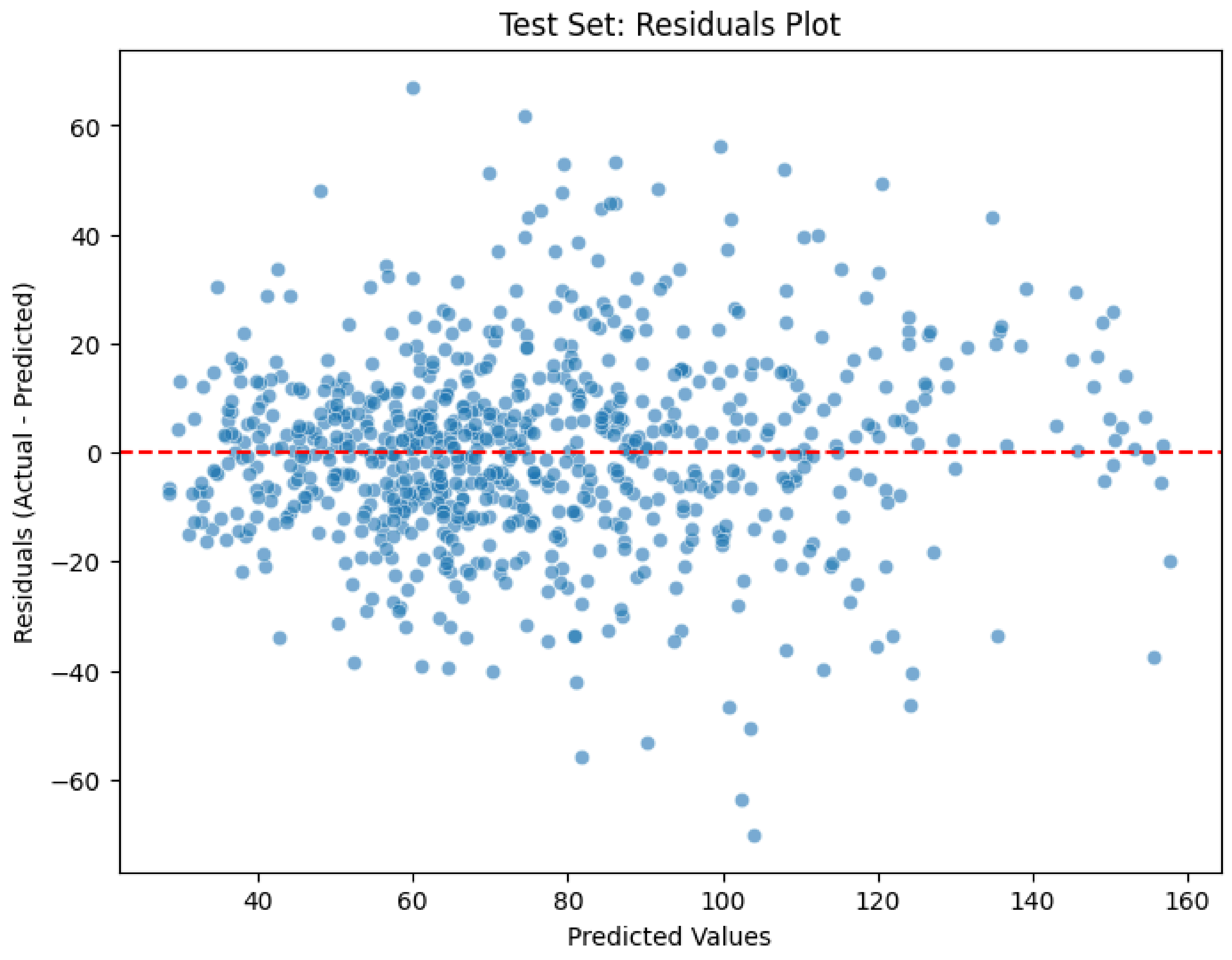


Increase in PM2.5 in Winters: Graphs show a rising trend during winter, likely due to temperature inversions and low wind speeds.

High Pollution at High Temperatures: Elevated PM2.5 levels suggest sources like vehicular emissions and industrial activities.

Skewed PM2.5 Distribution: Most values lie between 50-100 (Moderate AQI), with occasional high pollution spikes.

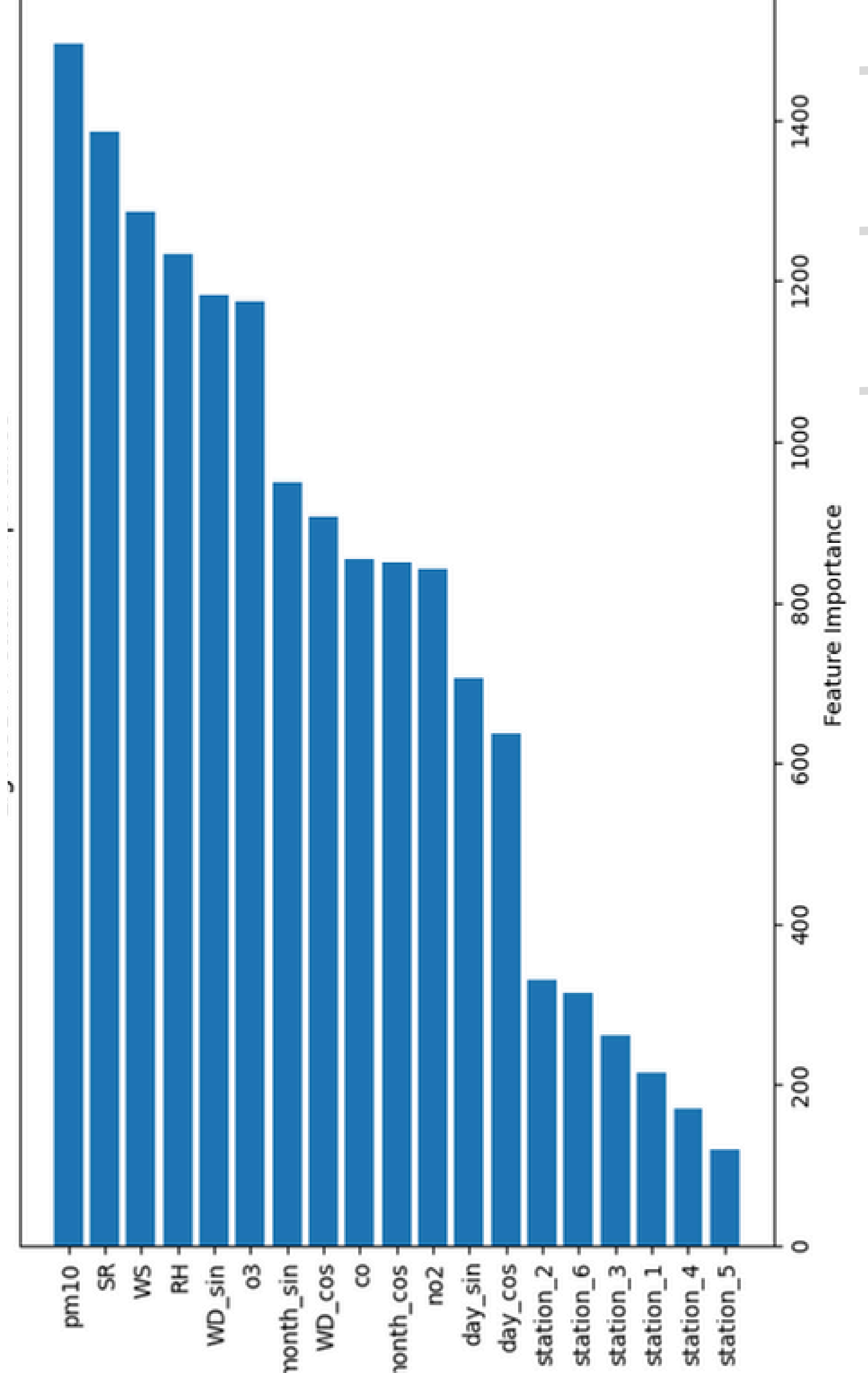

Test Set: Actual vs Predicted
Perfect Prediction
Predicted Values
Actual Values
25
50
75
100
125
150
175


Test Set: Residuals Plot
Residuals (Actual - Predicted)
Predicted Values
60
40
20
0
−20
−40
−60
80
100
120
140
160

- The model demonstrates strong predictive performance with a testing R^2 of 0.7221, testing RMSE of 17.4554 and a testing MAE of 13.0465, indicating moderate error with room for optimization.

- Residuals show some spread, particularly for higher predicted values, suggesting potential model improvement areas.

- Feature importance analysis shows mostly metrological factors as the top influencing factors in prediction.



03.

MODELING

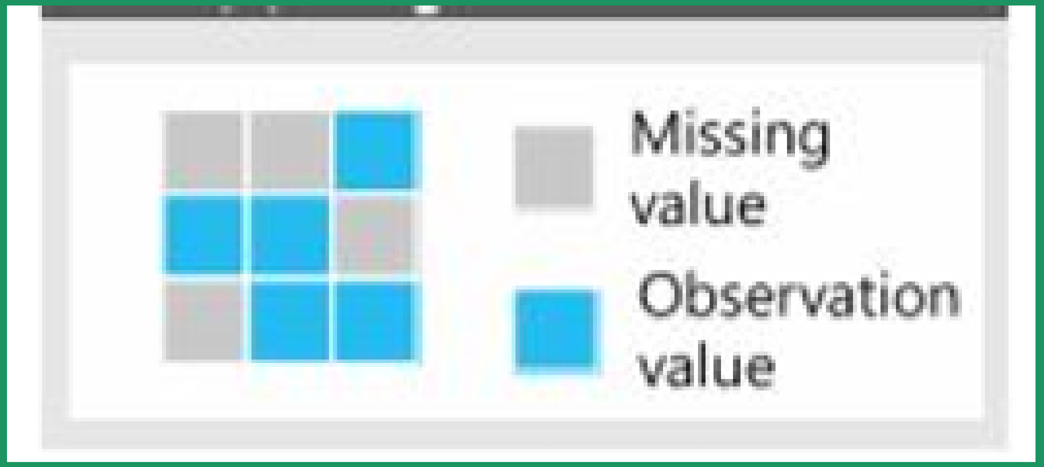
Temporal Forecasting

- As mentioned before, we use multiple features and ground station data.
- We plan to use Graph Neural Networks (GNN) for the forecasting

Source : LINK

- Graph neural networks (GNN) are connectionist models that capture the dependence of graphs via message passing between the nodes of graphs.
- Unlike standard neural networks, graph neural networks retain a state that can represent information from its neighborhood with arbitrary depth

Spatial Gap Filling



Such missing values could be predicted using previous spatial patterns

- Spatial Gap Filling method gives us the values at the intermediate locations.
- Kriging leverages spatial correlation between observed pollutant levels at different monitoring stations to estimate values at unmonitored locations.
- Kriging helps refine air quality maps by providing continuous pollutant concentration estimates across a region, overcoming the limitations of sparse monitoring networks.

04.

Health Recommendations

Why Do we need Health Recommendations?

- **Protects Public Health** – Air pollution causes respiratory & cardiovascular diseases.
- **Personalized Warnings** – Sensitive groups (asthmatics, elderly, children) need specific guidance.
- **Prevents Severe Incidents** – Example: Asthma patients can avoid attacks by staying indoors & using masks.

User-Specific Health Recommendations

1️⃣ Rule-Based Filtering

- Matches user health conditions with AQI levels.
- Provides instant alerts and precautionary measures.

2️⃣ ML-Based Risk Prediction

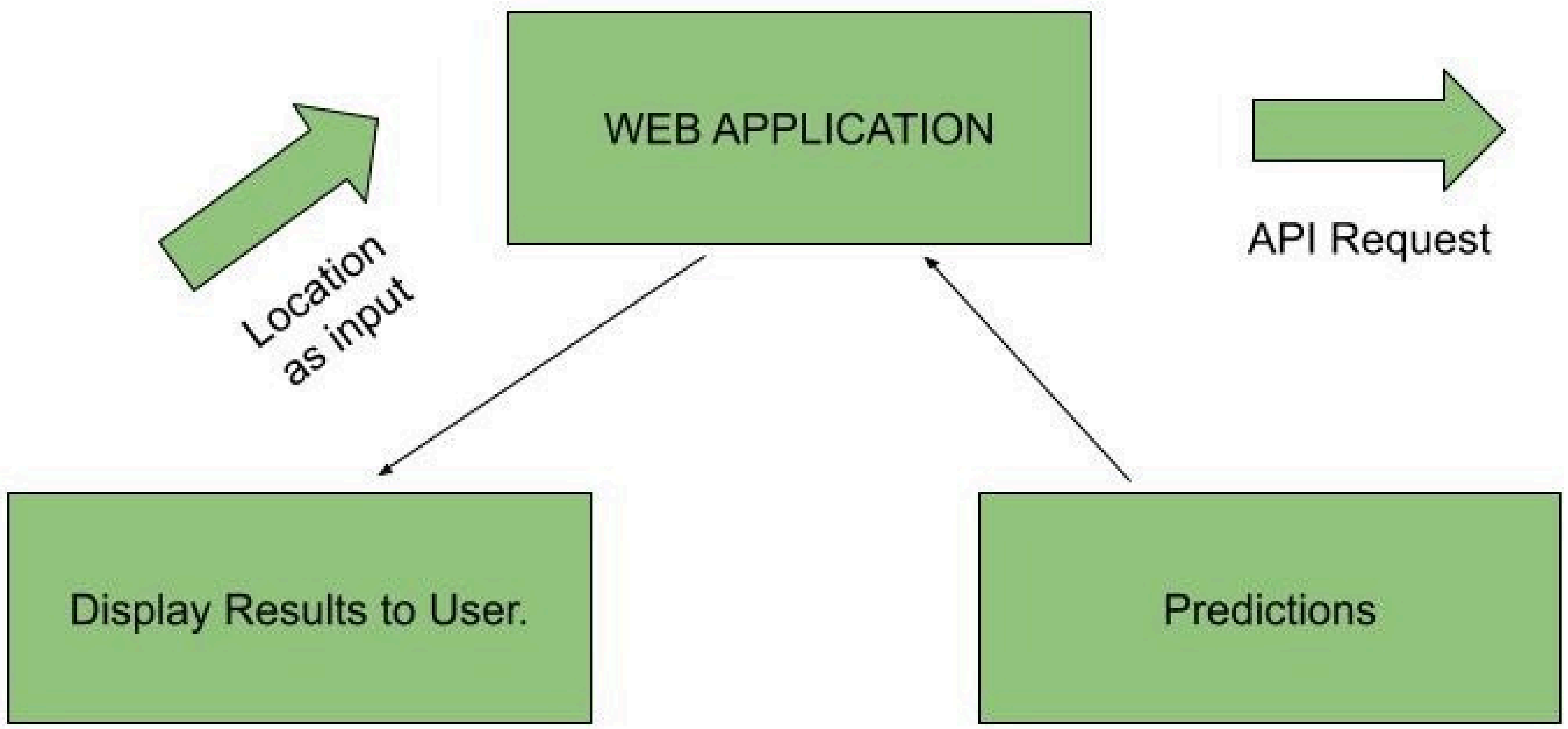
- Uses historical AQI trends and health impact data.
- Dynamically adjusts personalized health recommendations.

3️⃣ Geospatial Health Risk Mapping

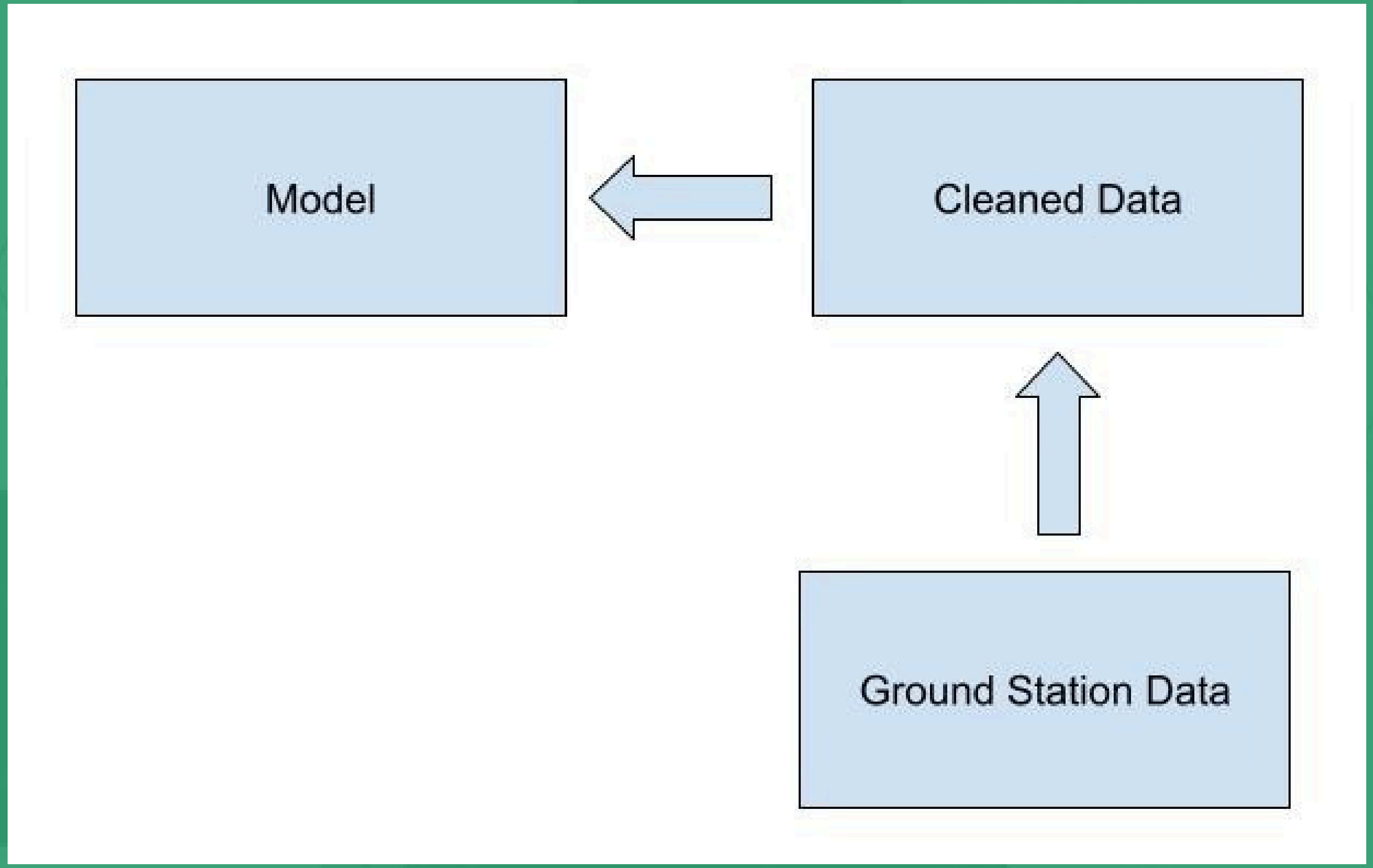
- Overlays pollution hotspots, user location, and vulnerability zones.
- Provides location-based health warnings.

WORKFLOW

Frontend



Backend



05.

CHALLENGES AND OUTCOMES

Challenges

- The main challenge of this project is estimation of pollutants data for the Spatial Gap Filling.
- Cleaning the ground station data as it contains many repeated values, null data and various data types.

END PRODUCT

Web Application

➢Implement a functionality over a selected region to let the user select the location.

➢ Getting the predictions of pollutants through API calls from backend and displaying to the user.

06.
TIMELINE

Work Done Till Now

- Gone through the research papers, proposing various methodologies.
- Collected the ground station data.
- Cleaned the data.
- Data Analysis of few features.
- Built some basic models to understand the data.
- Finalized the plan of action and models to use.

Future Work

- Predict the values for various stations using GNN model.
- Spatial Gap Filling using the predictions of the GNN model.
- Displaying the forecast to the user through web application.
- Sharing users important tips based on health recommendations.

Timeline

EVAL 1
Progress as mentioned above.

EVAL 2
GNN predictions and spatial gap filling

EVAL 3
Basic Web app

EVAL 4
Complete Website with health recommendations

Our Team

Jyothish Varma - S20220010127

Nandhitha Shaganti - S20220010198

Hemanth Chakravarthy - S20220010095

Thank You