Forecasting Car Park Electric Vehicles Charging Station Upgrades by Region

Problem Statement

The adoption of electric vehicles (EVs) is critical for reducing carbon emissions and combating climate change. However, one of the problems that hinders widespread EV adoption is the lack of convenient charging infrastructure. In urban areas like Singapore, where carparks play a central role in daily transportation, by identifying regions with high usage of fuel-operated vehicles, it becomes possible to forecast where carparks should be upgraded to include EV charging stations.

Objective

The primary objective of this project was to investigate the correlation between carpark availability and NO2 pollutant levels across different regions in Singapore. From which, the region of priority to upgrade carpark to equip it with EV charging infrastructure can be determined. Additionally, the project aimed to develop predictive models capable of estimating NO2 levels based on carpark availability data.

Data Sources

- PSI Pollutant Index: NO2 readings from the Singapore government API.
- Carpark Availability: Data from the Singapore government API.
- HDB Carpark Information: CSV file containing geospatial details of carparks. These datasets were chosen to provide a comprehensive view of pollutant levels and parking dynamics across time and regions.

Methodology

1. **Data Collection & Cleaning:**

   The data collection process involved gathering NO2 readings and carpark availability data from government APIs. The NO2 data provided pollutant levels at different times, while the carpark data detailed availability across various locations. Additionally, the HDB carpark dataset was processed to include latitude, longitude, and regional

classifications, enabling spatial analysis. Missing data points were addressed through interpolation techniques to ensure a continuous and reliable dataset for analysis.

2. **Feature Engineering:**

   Temporal features such as the day and hour were extracted from the datasets to capture time-based patterns. Aggregated features, including the mean availability difference for carparks and the average NO2 levels, were created to summarize trends and enhance the predictive capability of the machine learning models. These derived features provide critical insights into how temporal and spatial variations impacted carpark usage and NO2 levels.

3. **Analysis:**

   The analysis phase focused on exploring correlations between NO2 levels and carpark availability, both regionally and across the entire dataset. Patterns were identified to determine how carpark usage influenced pollutant levels, with specific attention to variations by region and time of day.

4. **Applying Machine Learning Models:**

   To predict NO2 levels, several machine learning models were developed and evaluated, including Linear Regression (LR), Support Vector Regression (SVR), Decision Tree Regression (DTR), Light Gradient Boosting Machine (LGBM) and an ensemble of the latter three models. Each model was trained and tested on the processed data to assess its accuracy and generalizability. The ensemble method emerged as the most effective approach as compared to the rest.

Insights

Significant correlations were observed between carpark availability differences and NO2 levels, with variations depending on the region and time of day. Among the models tested, the ensemble method achieved the best balance between accuracy and generalizability. From which, the based on the prediction for NO2 level by region, it can be concluded that the east region has the highest priority for upgrade. The findings aid policymakers in optimizing urban planning and implementing policies that reduces pollutants to environment.

Challenges

The project faced several challenges, including API rate limits and missing data, which required careful handling during preprocessing. Additionally, high variance in regional data posed difficulties in achieving consistent model performance.

Conclusion

The project demonstrated the feasibility of using carpark availability as a proxy for predicting NO2 levels. Future work could involve integrating additional data sources and refining model accuracy.