**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 29 April 2024 |
| Team ID | 737820 |
| Project Title | EcoForecast: AI-Powered Prediction of Carbon Monoxide Levels |
| Maximum Marks | 6 Marks |

**Data Exploration and Preprocessing**

In the data exploration phase, various statistical techniques are employed to understand the distribution and characteristics of the carbon monoxide data. Preprocessing involves cleaning the data, handling missing values, and possibly scaling or transforming features to prepare it for input into machine learning algorithms.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | The data typically consists of time-stamped measurements of carbon monoxide levels across various locations.  It may also include associated meteorological data such as temperature, humidity, wind speed, and other factors influencing CO levels. |
| Univariate Analysis | Exploration of individual variables (mean, median, mode, etc.).  Mean: The average CO level over a period of time.  Median: The middle value of the CO levels when arranged in order.  Standard Deviation: Measures the amount of variation or dispersion in CO levels |
| Bivariate Analysis | Relationships between two variables (correlation, scatter plots).  A scatter plot can visually represent the correlation between two variables by showing their relationship on a graph, where a positive correlation is indicated by an upward trend, a negative correlation by a downward trend, and a lack of correlation by a random pattern of points. |
| Multivariate Analysis | EcoForecast uses AI algorithms to provide real-time, accurate predictions of carbon monoxide levels through multivariate analysis that identifies patterns and relationships among multiple variables, such as traffic density, weather conditions, industrial emissions, and geographical data, to offer comprehensive insights for targeted interventions. |
| Outliers and Anomalies | Identification and treatment of outliers in EcoForecast's carbon monoxide level predictions involves monitoring for data points significantly deviating from expected ranges and employing statistical methods such as z-score or IQR-based outlier detection, then adjusting or removing these points to maintain accuracy. |
| **Data Preprocessing Code Screenshots** | | |
| Loading Data | 2024-04-29 |
| Handling Missing Data | 2024-04-30  WhatsApp Image 2024-04-30 at 20.29.08_ee4cc64f |
| Data Transformation | WhatsApp Image 2024-04-30 at 20.31.33_22a14a3f. |
| Feature Engineering |  |
| Save Processed Data |  |