**Model Development Phase Template**

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | SWTID1720078183 |
| Project Title | Predictive Modeling for Fleet Fuel Management using Machine Learning |
| Maximum Marks | 5 Marks |

**Feature Selection Report**

Below is a detailed report on each feature from our dataset, including a brief description, selection status, and reasoning.

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **Description** | **Selected (Yes/No)** | **Reasoning** |
| ‘distance’ | Distance traveled during the trip | No | Data type is ‘object’, conversion needed; lacks direct correlation to consumption. |
| ‘consume’ | Fuel consumption during the trip | Yes | Target variable for the analysis and model prediction. |
| ‘speed’ | Average speed during the trip | Yes | Directly influences fuel consumption, significant predictor. |
| ‘temp\_inside’ | Inside temperature during the trip | Yes | Affects comfort and possibly the use of air conditioning, impacting fuel consumption. |
| ‘temp\_outside’ | Outside temperature during the trip | Yes | Influences engine performance and potential AC use, impacting fuel consumption. |
| ‘specials’ | Special conditions (rain, snow, etc.) | No | Many missing values and is categorical; difficult to standardize for modeling. |
| ‘gas\_type’ | Type of gas used (Petrol or Diesel) | Yes | Different gas types have different efficiencies, crucial for accurate modeling. |
| ‘AC’ | Air conditioning usage (binary) | Yes | Significant impact on fuel consumption due to additional engine load. |
| ‘rain’ | Rain conditions during the trip (binary) | Yes | Affects driving conditions and potentially fuel consumption. |
| ‘sun’ | Sunny conditions during the trip (binary) | No | Limited direct impact on fuel consumption; can be combined with other weather conditions. |
| ‘refill  liters’ | Amount of gas refilled in liters | No | Few non-null values; insufficient data to be reliable for modeling. |
| ‘refill gas’ | Type of gas refilled | No | Few non-null values; insufficient data to be reliable for modeling. |

**Reasoning:**

* **Target Variable (consume):** Essential for model prediction, hence selected.
* **Numeric Variables (speed, temp\_inside, temp\_outside):** These variables are directly measurable and impact fuel consumption. Their selection is based on logical correlations

with the target variable. Missing values for temp\_inside and temp\_outside were filled with the mean to ensure completeness.

* **Categorical Variables (gas\_type, AC, rain):** These features are converted to numeric or dummy variables for analysis. They have a significant impact on fuel consumption.
* **Dropped Variables (distance, specials, sun, refill liters, refill gas):** These features either have too many missing values, are categorical with difficult standardization, or have limited direct impact on the target variable.