**Data Collection and Preprocessing Phase**



|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | 739687 |
| Project Title | SMS SPAM DETECTION |
| Maximum Marks | 6 Marks |

**Data Exploration and Preprocessing Report**

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

|  |  |  |
| --- | --- | --- |
| **Section** | **Description** | |
| Data Overview | Dimension:  14999 rows × 4 columns  Descriptive statistics: | |
| Univariate Analysis |  | |
|  | |  |
| Bivariate Analysis | | To find the relation between two features we use bivariate analysis. You can use seaborn package to plot visualisation uisng two variables of the datase |

|  |  |
| --- | --- |
| Multivariate Analysis |  |

|  |  |
| --- | --- |
| **Loading Data** |  |

|  |  |
| --- | --- |
| **Handling missing**  **Values** |  |
| **Feature**  **Engineering**      **Save processed data** | Attached the codes in final submission          - |
|  |  |
|  |