**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 8 July 2024 |
| Team ID | SWTID1720092248 |
| Project Title | Revolutionizing Liver Care: Predicting Liver Cirrhosis Using Advanced Machine Learning Techniques |
| Maximum Marks | 6 Marks |

**Data Exploration and Preprocessing Template**

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | The data overview provides basic statistics, dimensions, and structure of the dataset. It includes the number of records, number of features, and data types of each feature |
| Univariate Analysis | Univariate analysis involves exploring individual variables to understand their distribution, central tendency (mean, median, mode), and dispersion (variance, standard deviation). Visualizations such as histograms and box plots are used to illustrate these statistics |
| Bivariate Analysis | Bivariate analysis examines the relationship between two variables. This includes calculating correlation coefficients and creating scatter plots to visualize potential linear or non-linear relationships between pairs of variables |
| Multivariate Analysis | Multivariate analysis explores patterns and relationships involving multiple variables simultaneously. Techniques such as principal component analysis (PCA) and multiple regression analysis are employed to understand the interactions between variables |
| Outliers and Anomalies | Identifying and treating outliers is crucial to ensure accurate analysis. This section involves detecting outliers using statistical methods (e.g., IQR, Z-scores) and deciding on appropriate treatments (e.g., removal, transformation) |
| **Data Preprocessing Code Screenshots** | |
| Loading Data |  |
| Handling Missing Data |  |
| Data Transformation |  |
| Feature Engineering |  |
| Save Processed Data |  |