**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 JULY 2024 |
| Team ID | 739811 |
| Project Title | Detection Of Autistic Spectrum Disorder: Classification |
| 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 | #Structure of the data:      #Descriptive Statistical:  Descriptive analysis is to study the basic features of data with the statistical process. Here pandas has a worthy function called describe. With this describe function we can understand the unique, top and frequent values of categorical features. And we can find mean, std, min, max and percentile values of continuous features. |
| Univariate Analysis | Visual analysis is the process of using visual representations, such as charts, plots, and graphs, to explore and understand data. It is a way to quickly identify patterns, trends, and outliers in the data, which can help to gain insights and make informed decisions.  Univariate Analysis:  In simple words, univariate analysis is understanding the data with a single feature. We have displayed three different types of graphs and plots.  For simple visualizations we can use the matplotlib. pyplot library. Here the plt. figure() command is used to determine the size of the plot.  We have histogram for all features of the dataset which include phosphorus, humidity, temperature as well . The histogram shows the distribution of nitrogen fertilizers for crop. |
| Outliers and Anomalies | There is no Outliers in our project. |
|  |  |
| Outliers and Anomalies | There is no Outliers in our project. |
| **Data Preprocessing Code Screenshots** | |
| Loading Data | #Loading the data |
| Handling Missing Data | For checking the null values . isnull() function is used. To sum those null values we use. sum() function. From the below image we found that there are no null values present in our dataset. So we can skip handling the missing values step. |
|  |  |
|  |  |