Data Collection and Preprocessing Phase

| Date | June 2024 |
| --- | --- |
| Team ID | 739836 |
| Project Title | Predictive Modeling for H1-B Visa Approval Using Machine Learning |
| Maximum Marks | 6 Marks |

**Preparation Template for H1-B visa Approval**

The objective of this preprocessing step is to enhance data quality, promote model generalization, and improve convergence during neural network training. This will ensure robust and efficient performance across various computer vision tasks related to H1-B visa approval.

| Section | Description |
| --- | --- |
| Data Overview | There are many popular open sources for collecting the data, such as kaggle.com, UCI repository, etc. In this project we have used ‘.csv’ data related to H1-B visa applications. |
| Data Preparation | These are the general steps of pre-processing the data before using it for machine learning. |
| Handling missing values | We use techniques to handle missing values for checking and dealing with null values in the dataset. |
| Handling categorical data | Since our dataset has categorical data, we must convert it to integer encoding or binary encoding to make it suitable for machine learning algorithms. |
| Handling Outliers in Data | With the help of boxplots, outliers are visualized. We find the upper and lower bounds of numerical features using mathematical formulas to identify and handle outliers. |
| Data Preparation | |
| Collect the dataset | Please refer to the link given below to download the dataset: Customer Analytics Dataset on Kaggle |
| Importing the libraries |  |
| Loading Data | Read the Dataset:  df=pd.read\_csv("/H1-B visa dataset.csv") |
| Handling missing values | checking for Null Values:    Imputation Technique : |
| Handling Categorical values | Encoding Techniques: |
| Handling Outliers | Visualization:    Mathematical Formulas:  Handling Outliers: |