**Data Collection and Preprocessing Phase**

| Date | June 2024 |
| --- | --- |
| Team ID | Team-739764 |
| Project Title | Auto Insurance Fraud Detection Using Machine Learning |
| Maximum Marks | 6 Marks |

**Preparation Template**

The images will be preprocessed by resizing, normalizing, augmenting, de noising, adjusting contrast, detecting edges, converting color space, cropping, batch normalizing, and whitening data. These steps will enhance data quality, promote model generalization, and improve convergence during neural network training, ensuring robust and efficient performance across various computer vision tasks.

| **Section** | **Description** |
| --- | --- |
| Data Overview | There are many popular open sources for collecting the data. E.g.: kaggle.com, UCI repository, etc. In this project we have used .csv data. |
| Data Preparation | These are the general steps of pre-processing the data before using it for machine learning |
| Handling missing values | We use Handling missing values For checking the null values |
| Handling categorical data | As we can see our dataset has categorical data we must convert the categorical data to integer encoding or binary encoding |
| Handling Outliers in Data | With the help of boxplot, outliers are visualized. And here we are going to find upper bound and lower bound of numerical features with some mathematical formula. |
| Data Preparation | |
| Collect the dataset | Please refer to the link given below to download the dataset.  Link: https://www.kaggle.com/datasets/buntyshah/auto-insurance-claims-data |
| Importing the libraries |  |
| Loading Data | We use the code df=pd.read\_csv("/content/Train.csv")  For reading the dataset |
| Handling missing values |  |
| Handling Categorical values |  |
| Handling Outliers |  |