[Credit Card Fraud Detection using Machine   
Learning Approach]

[Shanti Mutyala]

**Data Science Capstone Project   
Data Acquisition and Pre-Processing Report**

Date:

[09/02/2023]

Team Members:

Name: Shanti Mutyala..

[The purpose of this report is to describe the data of your project. It includes three major sections: Data Sources, Data-Processing, and Appendix]

**Identifying Data**

**Data Sources:**

**Credit Card Fraud Dataset:** I acquired the dataset from the following site manually [www.kaggle.com/datasets/mlg-ulb/creditcardfraud?resource=download](http://www.kaggle.com/datasets/mlg-ulb/creditcardfraud?resource=download).

**Student Exam Performance Dataset :**  I acquired the dataset from the following site manually

<https://www.kaggle.com/datasets/spscientist/students-performance-in-exams>

**Acquisition Process:**

Both the data were readily available and manually downloaded.

Further processing of them was done by unzipping the files and reading through the pandas package of python.

**Issues:**

No, there were major issues found in fetching the database.

The only issue was the long time required to process the Credit Card Database as the data size was huge and processing them required a longer time.

**Data-Processing**

1. I inspected null values for each dataset.
2. I scaled the Amount and Time data using Robust Scaler as I wanted scaling to be less affected with outliers.
3. I used various sampling methods such as oversampling and undersampling as well for the Credit Card Fraud Dataset due to an imbalance dataset.

**Appendix**

**Sample Credit Card Fraud Data Description :**

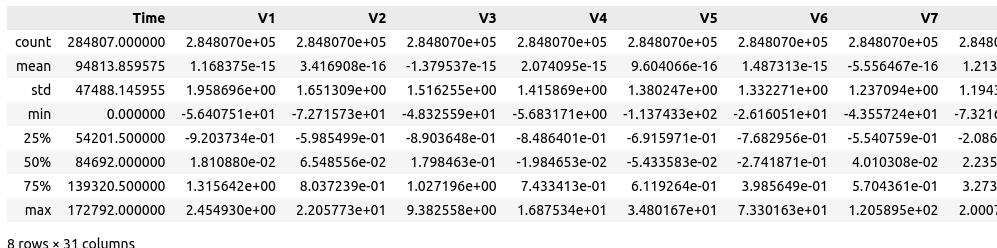


Table of Contributions

The table below identifies contributors to various sections of this document.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Section** | **Writing** | **Editing** |
| **1** | **Data Sources** | **Shanti** | **Shanti** |
| **2** | **Data Pre-Processing** | **Shanti** | **Shanti** |
| **3** | **Appendix** | **Shanti** | **Shanti** |

**Grading..**

The grade is given on the basis of quality, clarity, presentation, completeness, and writing of each section in the report. This is the grade of the group. Individual grades will be assigned at the end of the term when peer reviews are collected.