Project Proposal

MET CS – 677

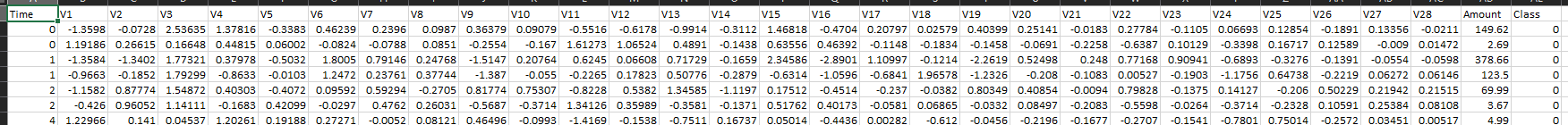
Adetya Aggarwal

Project – For the term project I will be using a credit card dataset which is highly imbalanced. This is a classification problem dataset where fraud and legitimate transitions need to be classified respectively. The motivation to choose this as a class project was to work with financial data and also to learn to tackle highly imbalanced dataset and solve the problem using SMOTE, oversampling and other techniques.

Dataset – Credit Card Fraud Dataset

Source - <https://www.kaggle.com/mlg-ulb/creditcardfraud>

Task – To classify fraud and legitimate transactions

Sample - 

Attributes – there are 31 attributes. The first feature is time feature with gives the seconds between transactions. the amount feature provides the transaction amount. The class feature provides information regarding transaction type. 1 in case of fraud and 0 for legal transactions. All other attributes are masked to save user’s personal information and no information is available about them.