

DATS 6103

INTRODUCTION TO DATA MINING

Prof. Edwin Lo

Project Proposal on

Credit Card Fraud Detection

By,

TEAM V

Aveline Mariya Shaji, Kyuri Kim, Satya Bharath Reddy Duvvi

Research Proposal

- **RESEARCH TOPIC**

It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase. The objective of this project is to do data visualization to understand the data and then move ahead with Machine Learning techniques based on the understanding from Exploratory Data Analysis. The entire project will use different libraries from Python and we expect to come with an accurate model for prediction.

- **SMART Questions**

What factors cause fraud credit card transactions?

- **Understanding Data**

The dataset is taken from kaggle. The dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. It contains only numeric input variables which are the result of a PCA transformation.

The link to our dataset is: <https://www.kaggle.com/mlg-ulb/creditcardfraud>

- **GitHub Repository**

We are a group of three and will be working on this project. Our collaborations can be seen at our GitHub Repository.

The link to our repo is : https://github.com/orangekim28/Team5_DataMining6103