Project Proposal

Hayden Strikwerda, Bryce Kastner, and Hannah George

Due: March 1, 2019

Basic Info

Title: *Fraud Detection Using Machine Learning.*

Names:

Hayden Strikwerda, *u0849929@utah.edu*, u0849929
Hannah George, *u0955984@utah.edu*, u0955984
Bryce Kastner, *u1034994@utah.edu*, u1034994

Background and Motivation

We were inspired by companies in the Salt Lake area that are creating similar programs in their day-to-day work. We felt as though this project would not only provide useful knowledge in Machine Learning, but would also be a good example of how Data Science is used every day.

Project Objectives

Questions: We would like to learn about what factors are most likely to indicate fraudulent behavior. Whether that be the location of the transaction, or how frequent the transactions are occurring. We would also like to know what would be the appropriate response to possible fraudulent transactions, like asking to answer a personal question or completely freezing the account.

Benefits:

- Improving the security of bank transactions.
- Gaining a hands-on approach to understanding the Machine Learning process.

Data

We are collecting our data from a German credit card fraud dataset by Dr. Hans Hofmann. The link to the data is listed below:

$$http://weka.8497.n7.nabble.com/file/n23121/credit\_fruad.arff$$

Ethical Considerations

Inherently, this kind of data is very personal so it is unlikely that we will be able to access the amount of information that we would ideally want for this project. But if we could get a dataset that does involve personal information, then we would need to keep that information private.

Data Processing

We do not expect to do a lot of data cleanup since our data will most likely come in easy to work with forms. Since we will most likely be working with a large dataset then we will store it like we did in homework 6. We would want to at least have the information of the time in between transactions, the amount of money displaced in the transaction, the location of the transaction, the type of transaction, and whether the transaction was fraudulent or not.

Exploratory Analysis

A good way to look at our dataset is to use a heatmap to determine which factors are more important in determining whether a transaction should be flagged as possible fraud. We could also use use a decision tree to visualize the likelihood of a transaction being fraud given a certain event occurs. Since our dataset is likely to be unbalanced we could also use a bar graph to show the imbalance between fraudulent transactions and normal transactions.

Analysis Methodology

We will create a Pandas Data Frame out of the data we collected. Then we will look at the R-squared value of the linear regression model created between fraudulence of the transaction and every other variable we have to determine which variables indicate fraud the most. Then we will create a Machine-Learning algorithm that will use the given information to determine whether a transaction should be flagged as possible fraud.

Project Schedule

March 1
- Finish Proposal.

March 7
- Peer Review.

March 8
- Clean Data / Exploratory Analysis.

March 10
- Staff Feedback.

March 22
- Finish Regression Analysis.

March 31
- Submit Project Milestone.

April 12
- Finish Machine Learning.

April 19
- Finish video and presentation.

April 21
- Submit Final Project.