Midpoint Report

Team Members
Swetha Venkatesan
Sree Vandana Nadipalli

Project Objective:

The topic we want to explore is "Data Breaches" and our focus is on automating the process of identifying the breach. And to do so, we are going to implement and generate a model which could do so, using Machine Learning algorithms. The problem we are going to work is "Credit Card Fraud Detection"

Project Approach:

Our Approach involves following steps:

- Gathering the data and visualizing the data to get a brief understanding of the data that we are going to work with.
- Deciding on what type of models to use, that would best detect the fraud based on the data we have.
- Create multiple models and assess the accuracy of each model
- Compare the results, observations, and precision/accuracy of different models.

Our goal in this project is to construct models to predict whether a credit card transaction is fraudulent. It's a supervised learning approach. We have created visualizations to help us understand the structure of the data and unearth any interesting patterns. As it helps in deciding the suitability of a model. Implement a few models and analyze the performance among the models constructed.

Team Structure:

We are a team of two. We both will be finding the data source and visualize the data that we gathered and discussing what models would be best to implement based on the data we have.
We will split the implementation of the decided models and gather the outputs of both the implementations. We will finally evaluate the observations of the models we implemented.
We both will be contributing to the Final Presentation Slides.

Project Progress:

1. Setup the environment for our project (Python 3.8, TensorFlow 1.4 and Keras 2.0)
2. Data Gathering and Data Visualization -
    a. Transaction class distribution
    b. Amount per transaction by class
    c. Time vs Amount of transactions

d. Time density Transaction
e. Transaction Amounts (Box-Cox Transformed)
f. V1-V28 means, standard deviation, skewnesses

3. Deciding the models
a. Logistic Regression
b. Support Vector
c. Random Forest

So, far we have decided only on these models but we are exploring other models such as autoencoders, neural networks, adaBoostClassifier, XGBoost etc.

Project Milestone - Remaining

- Designing the model and gathering output's
- Analyze/Evaluate the observations and accuracies of the models
- Code Submission
- Final Presentation

Midpoint Meeting

- Date - 07/23/2020 Time - 02:00 PM
- *Points discussed in Meeting*