Springboard Data Science Career Track

Capstone Project 2 Proposal by Ashutosh Varshney

Problem Statement Worksheet (Hypothesis Formation)

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. In this project we will build a model to predict fraudulent transactions.

1. **Context**
   Credit card fraud is on the rise and it affects banks, customers and merchants. It is important to have a mechanism to accurately identify fraudulent transactions which will result in significant cost savings to all parties concerned.

2. **Criteria for success**
   We envision creating multiple supervised classification models which will be compared and evaluated for various performance metrics. Further, we will study interpretability of these models and analyze feature impact and feature importance.

3. **Scope of solution space**
   This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. It further identifies fraudulent transactions. Data has been anonymized for privacy.

4. **Constraints within solution space**
   a. The dataset is highly unbalanced. Negative class outweighs positive class by a huge percentage.
   b. The dataset is limited to European cardholders.
   c. The dataset contains only two days of data.

5. **Stakeholders to provide key insight**
   Key stakeholders include Credit card banks, their customers and merchants.

6. **Key data sources**
   The file creditcard.csv was obtained from Kaggle https://www.kaggle.com/mlg-ulb/creditcardfraud.