Machine Learning Final Project Proposal

Teammates:
1. Aditya Sawwalakhe
2. Shubham Patwa
3. Harshal Rasal

Problem statement:

Credit cards have been a huge success in recent years. They are issued without background checks to increase their market share and hence, there is a good chance of customers defaulting the payment or overusing the credit card in the sense that debt cannot be repaid. Thus for the bank, identifying risky and non-risky customers is very crucial. In this project, we will use different machine learning algorithms for coming up with solutions whether the customer will default or not.

We are using the dataset on the UCI website, and taking consideration of various attributes of customers including education, payment history, gender, marital status, etc.

Tentatively we will use the following algorithms for data preprocessing, features selection, training model over the dataset.

Tentative Algorithms list:

1. Logistic Regression
2. SVM
3. Neural Networks
4. Random Forest
5. KNN

We will use the above algorithms over the dataset and evaluate each by plotting the confusion matrix and calculate precision, recall, and F1 score and decide which algorithm is the best.

We will probably also use ensemble methods with the above algorithms.

Dataset: https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients

Research Papers:

Application of Machine Learning Algorithms in Predicting Credit Card Default Payment

https://escholarship.org/uc/item/9zg7157q

Comparison of Several Data Mining Methods in Credit Card Default Prediction
https://www.scirp.org/pdf/IIM_2018092515040783.pdf

The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients
https://bradzzz.gitbooks.io/ga-seattle-dsi/content/dsi/dsi_05_classification_databases/2.1-lesson/assets/datasets/DefaultCreditCardClients_yeh_2009.pdf

Task Responsibilities:

Shubham Patwa:
1. Data preprocessing
2. Implementing SVM and Logistic Regression
3. Plotting confusion matrices
4. Concluding

Aditya Sawwalakhe:
1. Preprocessing and features selection over the dataset.
2. Using Neural Networks and Random Forest for the provided dataset
3. Implementing various visualization techniques to understand the dataset.

Harshal Rasal :
1. Implementing KNN
2. helping out other teammates with presentation and final report
3. Deriving different metrics from the confusion metrics.