PROJECT PROPOSAL

CS-584 MACHINE LEARNING

Mohit Kulpe, A20372614

1 INTRODUCTION:

In today's world credit history has become one of the most important factor. Without good credit history financial institutions deny to lend loan. Some financial Institutions also remain in dilemma to whether approve credit card or not for client. Since I am interested in financial sector I decided to implement a Machine learning approach to determine whether a client will be defaulter or not. I found a suitable dataset online, of Taiwan credit card clients.

2 PROBLEM:

Financial Institutions need to take best decision regarding loans. In case if financial institute lends loan to client who is not capable of repaying principle along with interest then the institute will suffer a loss. This Machine Learning model will help financial institutions to make better decision whether to lend loan or not.

3 TECHNIQUE:

This is supervised learning technique of Machine Learning. Goal of this project is to Classify defaulter & non defaulter clients, So basic technique I am going to use is Classification Algorithms.
Logistic Regression algorithm: It is used to predict probability based on given set of independent variables.
Decision Tree: Used to split population in different groups.
Support Vector Machine: This algorithm splits features & graphically plots it.
Naïve Bayes: This algorithm is based on Naïve theorem which is used to classify things.

After implementing these techniques I will evaluate their performance by using Recall Precision metrics and will try to find best fit algorithm. Also will experiment with combinations of above techniques to get best accuracy.

Then after finding best match and combination I will try to increase accuracy by fitting different features (predictors), by making different combinations of features (predictors) & by performing interactions with features (Predictors).

3 DATASET:

This dataset contains information on default payments, demographic factors, credit data, history of payment, and bill statements of credit card clients in Taiwan. There are total 24 variables, 1 response (defaulter Yes/No) & 3000 records.

Input: Limit balance, Demographic information (Sex, Education, Age, Marriage), Repayment status of last 5 months (Pay), Bill amount of last 5 months (Bill_Amt) & 5 Previous payment (Pay_Amt).

Output: Defaulter Yes or No.

Source: UCI Machine Learning Repository
URL: https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients

4 REFERENCE:

https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients
https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset
http://www.ijcaonline.org/archives/volume145/number7/ajay-2016-ijca-910702.pdf