

# **Machine Learning Engineer Nanodegree**

## **Capstone Project Proposal**

**Medehal Priyanka**

**February 2nd, 2019**

### **Proposal:**

#### **Credit Approval Prediction**

#### **Domain Background:**

##### **History:**

According to historian Jonathan Kenoyer, The concept of using a valueless instrument to represent banking transactions dates back 5000 years. The charge-plate, developed in 1928 was an early predecessor of the credit card. The first universal credit card was introduced by Diners Club, Inc., in 1950. Presently cooperating merchants of the credit card issuer throughout the world would pay service charge in range of 4-7% of total billings.

When you apply for a credit card, you may be asked to disclose your income as part of the application. The card issuer will likely look at your overall income relative to your debt also known as debt-to-income ratio, To make sure the company doesn't offer you more credit than they think you can reasonably payback

There are many more factors that are considered while issuing a credit card . Here I am going to predict whether a credit card can be approved or not.

##### **Applications:**

This project is predicting the approval of a credit card. And it will help the credit card issuers whether to approve a credit card or not.

##### **Problem Statement:**

The main aim of my project is to predict the credit card approval. For doing this I selected the dataset from UCI (<https://archive.ics.uci.edu/ml/datasets/Credit+Approval>). So my goal is to predict the credit card approvals. Here I am using classification models to find the accuracy of each model and select the best model with high accuracy to predict the approvals. Here the input parameters are training data that we took and the output will be whether to approve the credit card or not.

##### **Dataset Information:**

This file concerns credit card applications. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data. This dataset is interesting because there is a good mixture of attributes – Categorical, integer and real valued attributes. There are also few missing values. On a total there are 15 attributes. The

total number of instances in the datasets is 690. The characteristics of the dataset are multivariate. Here there are mainly two classes: + (approving credit card) and - (not approving credit card). There are 307 +'s and 383 -'s which are closely balanced.

For the best result we will split the data into training set and testing set. On a whole we will assign 70% of the data to training set and 30% of the data to testing set.

### Attributes:

A1: b, a.  
A2: continuous.  
A3: continuous.  
A4: u, y, l, t.  
A5: g, p, gg.  
A6: c, d, cc, i, j, k, m, r, q, w, x, e, aa, ff.  
A7: v, h, bb, j, n, z, dd, ff, o.  
A8: continuous.  
A9: t, f.  
A10: t, f.  
A11: continuous.  
A12: t, f.  
A13: g, p, s.  
A14: continuous.  
A15: continuous.  
A16: +, - (class attribute)

### Solution Statement:

Here, I am trying to predict the credit approval from the selected dataset for predicting the credit card approval we will use the different classification models. Then we will find the accuracy score for each model. I explore the dataset by using read\_csv and for visualization which helps me to better understand the solution, I used matplotlib.pyplot.

### Benchmark Model:

Benchmark model is a model which we will take as reference and achieve the best result than the benchmark model in our project we will take naive bayes as a benchmark model. Using naive bayes model we will achieve an accuracy of 62.74. Now we will try and achieve the better accuracy than the benchmark model.

### Evaluation Metrics:

I want to use accuracy score as evaluation metric for prediction of credit approval. Here the dataset classes (+ and -) are closely balanced, so we can use accuracy as an evaluation metric. Here I am predicting the accuracy score for the selected models. Here we will select a model whose accuracy score is greater than all the other models and we treat it as the best.

### Project Design:

The project is composed of different steps as follows:

**Pre-processing:**

The first task is to read the data and perform visualizations on it to get some insights about the data after. Reading the data, clean the data that is removing the unwanted data or replacing null values with some constant values or removing duplicates.

After data exploration I split the data into training and testing sets. After splitting the data we will apply classifying models and then predict the accuracy score for the selected model.

**Training the data:**

Here I will use the classification models like random forest, decision tree, svm, logistic regression, knn, naive bayes. After training the data we will test all the models with testing data. After that we will find out the accuracy score for all the models.

Finally, I will declare the model with highest accuracy score as the best model for detecting the credit card approval.