Business Understanding & problem Discovery

Business Problem Introduction:

The use of credit cards in todays world has become one of the primary payment methods. According to data from Nov. 1, 2019 by the American Bankers Association, there were 374 million open credit card accounts in the U.S. as of the middle of 2019.[[1]](#footnote-1) With that many credit cards accounts in the United State alone, credit card scores are a common risk control method in the financial industry where it uses personal information and data acquired by applicants to predict potential risk when issuing a credit card to an individual. Using historical data, the credit card provider can predict whether an applicant might default on their card and essentially, if the bank should issue the credit card to the applicant in the first place.

Refined Problem Statement:

Understanding what makes a good applicant who is approved for the credit card is the fundamental problem that needs to be solved. The task at hand is to develop a machine learning model to see if potential applicants for credit cards will be approved or not based on their information and to understand the factors and features to what makes a good applicant.

Data Requirements:

The key data requirements for this project is historical data for credit card applications as to better understand the key variables in seeing which features have the largest impact on credit card approve. The current data set that contains historical data of credit card applications is will be acquired from the public data domain of Kaggle. [[2]](#footnote-2)

Data Assumptions and limitations:

The primary assumption for this data set is that is proved a wide enough range of applicants and all the applicants differ enough as to provide a good population sample size in order to develop a model that can predict the approval for new credit card applicants. It will be assumed that this data provides a perfect scaled representation of an unknown countries population as to create a simulated perfect model. The potential limitations of these assumptions are that this unknown country might have different credit approval factors that differ from Canada or the United States or England etc. This means that the model would not be able to be ‘plugged and played’ in all countries and model adaptation is needed for it to work in each country. Furthermore, regarding if the data used for the model is a perfect scaled representation of a population then the model might be not be able to predict outliers and irregularities or infer potential missing data.

Testing:

Testing will involve separating the data set and using cross validation methods of testing and training to better adapt the model to guarantee performance.

1. chttps://www.creditcards.com/credit-card-news/ownership-statistics/#:~:text=Using%20the%20U.S.%20Census%20Bureau,a%20charge%20card%20or%20both. [↑](#footnote-ref-1)
2. https://www.kaggle.com/rikdifos/credit-card-approval-prediction [↑](#footnote-ref-2)