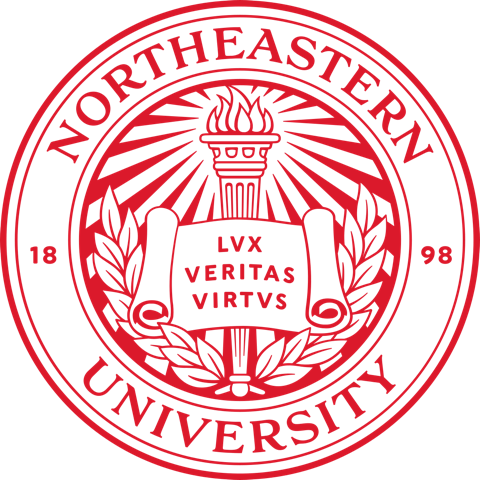
Credit Card Approval Prediction using Data Mining Techniques

**Course:** ALY 6040

**Instructor’s Name:** Prof. Dr. Justin Grosz



**Submitted By:**

Akshay Kirit Kalavadia

Sunil Raj Thota

Jyothi Chandrakanth

Li-Ching Hwang

Nissy Asish Gantepalli

**Date:** 09/26/2021

**Introduction**

According to Robert Scott, on average a household in the U.S. has 8 credit cards in total. By the start of 2020, the total credit card balance in the US is $893 billion and increasing each day. Over the past two decades, the credit card system has been widely used as a mechanism to propel the world economy toward spectacular growth. But, issuing a credit card to risky customers can lead to financial crises which can eventually lead to situations like 2008. Hence, in this project, we will use data mining applications to evaluate a credit card application that would help credit lending institutions decide whether to accept or reject an application. We would use information submitted by applicants to predict the likelihood of defaulting on credit card payments.

One of the most important factors we consider while making this decision is the credit score of the applicant. A credit score can accurately measure the magnitude of risk. Based on the historical data of the applicants, we can develop a model which would provide regulators with a prediction of a client as a ‘good’ or ‘risky’ customer.

**An Overview of the Dataset**

**Include both the number of records, number of variables, and a list of specific variables of interest with the data type of each.**

For the project, we will combine two datasets. The first one is the application record dataset, there are 18 variables and 438557 observations. Another one is the credit record dataset having 3 variables and 1,048,575 observations. These columns included Client number, Gender, Is there a car, Is there a property, Number of children, Annual income, Income category, Education level, Marital status, Way of living, Birthday, Start date of employment, Is there a mobile phone, Is there a work phone, Is there a phone, Is there an email, Occupation, and Family size. Client number, Record month, and Status are included in the credit record data. We think all the variables are interesting, except having the phone and email. Our main analysis items are annual income and whether there is a property, the number of children, and status. The type of annual income and number of children columns are numerical variables, and the property column is a factor variable.

Here is the list of variables that we are interested in to do Exploratory Data Analysis and perform Data Mining Techniques accordingly. This includes;

|  |  |
| --- | --- |
| **Features** | **Type** |
| ID | INTEGER |
| CODE\_GENDER | FACTOR |
| FLAG\_OWN\_CAR | FACTOR |
| FLAG\_OWN\_REALTY | FACTOR |
| CNT\_CHILDREN | INTEGER |
| AMT\_INCOME\_TOTAL | NUMERIC |
| NAME\_INCOME\_TYPE | FACTOR |
| NAME\_EDUCATION\_TYPE | FACTOR |
| NAME\_FAMILY\_STATUS | FACTOR |
| NAME\_HOUSING\_TYPE | FACTOR |
| DAYS\_BIRTH | INTEGER |
| DAYS\_EMPLOYED | INTEGER |
| OCCUPATION\_TYPE | FACTOR |
| CNT\_FAM\_MEMBERS | NUMERIC |

**Describe your initial interest in this dataset?**

According to data released by the Federal Reserve Bank of Atlanta in August 2019, 75.5% of consumers have at least one credit card which allows the cardholder to purchase by borrowing funds that can be paid back to the credit card company later.

The credit card industry is huge, with many companies providing services to hundreds of millions of credit cardholders. With so many customers it’s really hard for credit card companies to manage everything. There are so many decisions to optimize, from determining whom to issue a credit card, how much credit to offer, what benefits to offer and when, how to protect from fraud, and much more.

Machine learning can help Credit Card Issuers to make better sense of the exponentially increasing transactional and behavioral data. It also helps in predicting the spending behavior of the consumers accurately. This interested us in selecting the Credit card Dataset where we can apply ML Algorithms in issuing Credit to the consumers by considering their Credit Score, Income, Credit History, etc.,

**What (in general terms) you are hoping to learn?**

We aspire to foster and combine data analysis concepts, data mining techniques, and machine learning models such as regression, classification, and clustering algorithms. Also, we would like to explore and go ahead by analyzing data and interpreting the insights in a visual format by following data visualization best practices. We are also keen on gaining the necessary knowledge and skills to employ advanced data strategies in our future careers. We will be going through necessary research papers, case studies, white papers, blogs, articles, and courses to uncover data mining methodologies and try to implement them in our projects.

Also, we will develop our leadership skills by following all the necessary standards and take responsibility for implementing the business objectives. We will use all our analytical skills, critical thinking, problem-solving, and data-driven decision-making strategies to provide recommendations to the necessary stakeholders.

**Make sure you define and can clearly define what you are trying to predict**

The goal of this project is to forecast credit card acceptance based on a variety of factors that influence a customer's application such as, If the applicant owns a vehicle or property, the applicant's yearly income, and the type of house they live in, such as rented or owned. The duration of the employment of an applicant, the kind of work they perform and their educational qualification, and the Gender. Additionally, the objective is also to assess whether or not the customer is a prospective defaulter.

**References:**

[1] Credit card Approval Prediction. Kaggle. (n.d.). Retrieved September 26, 2021, from https://www.kaggle.com/rikdifos/credit-card-approval-prediction?