

***Gideon Blinick***

## **Springboard Project Proposal - Capstone 1**

**Title:** Predicting Credit Card Default

### **Background/Problem:**

Credit card use is widespread in the United States. About 70% of all Americans use at least one credit card and there were 364 million open credit card accounts in the United States at the end of 2017, according to the American Banking Association (Gonzalez-Garcia, 2018). With use of credit cards comes the risk of default, which is defined as failing to make a payment for 180 days (Konsko, 2014).

The goal of this project is to effectively predict credit card payment default using demographic factors, credit data, repayment statuses, bill statements, and history of payments. Two sub-goals accompanying that goal are determining how the probability of default payment varies by categories of different demographic variables, and determining the strongest predictors of default payment among the variables (Default of Credit Card Clients Dataset, n.d.).

### **Client Interest:**

The client for this project is, naturally, credit card companies. Such companies have a significant interest in predicting which customers will default on their payments because such defaults cost them money, and thus, they would rather not extend money to individuals with a high probability of default. A good prediction model will enable them to lend to good customers.

A good prediction model will also enable credit card companies to make early and effective interventions with existing customers who are likely to default. Such interventions may take the form of debt and financial counseling, enabling the customer to maintain good credit and enabling the company to receive payment. In less fortunate circumstances, an intervention may allow credit card companies to be paid ahead of other creditors.

In summary, credit card companies are the primary stakeholder in this project, the decision being improved is the acceptance and rejection of credit applications as well as the decision of who to extend interventions to, and improvement of the decision results in greater profit for the company.

It is worth mentioning that this problem is not too much different from other problems such as bankruptcy prediction, and that a solution to one problem might generalize easily to the other. Hence, other stakeholders to this problem could include banks and other large creditors.

### **Data Source:**

The data source for this problem is a dataset from Kaggle.com containing about 30,000 records of information on credit card clients in Taiwan from April 2005 to September 2005. It can be found, here: <https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset/home>

This dataset is originally from the University California, Irvine's Machine Learning Repository, and can also be found here: <https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients#>

### Approach to Solving the Problem:

This project will implement 6 common data mining techniques, as outlined in the paper “The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients” (Yeh & Lien, 2009). They are:

- 1) K-nearest neighbor classifiers (KNN)
- 2) Logistic regression (LR)
- 3) Discriminant analysis (DA)
- 4) Naive Bayesian classifier (NB)
- 5) Artificial neural networks (ANNs)
- 6) Classification trees (CTs)

In particular, this project will attempt to recreate the results reached by that paper and note if any differences of result are reached (and explore them if they do exist).

### Deliverables:

Deliverables for this project include project code, a project paper or report, and a presentation of the project findings in a slide deck.

## References

*Default of Credit Card Clients Dataset*. (n.d.). Retrieved from [www.kaggle.com](http://www.kaggle.com):

<https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset/home>

Gonzalez-Garcia, J. (2018, April 26). *Credit card ownership statistics*. Retrieved from

[www.creditcards.com](http://www.creditcards.com): <https://www.creditcards.com/credit-card-news/ownership-statistics.php>

Konsko, L. (2014, September 2). *I Defaulted on My Credit Card – Now What?* Retrieved from

[www.nerdwallet.com](http://www.nerdwallet.com): <https://www.nerdwallet.com/blog/credit-cards/credit-card-default-what-to-do/>

Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 2473-2480.