EC 601 Product Design Project 1

Topic: American Express – Default Prediction

Author: Tianze Li

Email: tianzeli@bu.edu

Due Date: September 18, 2022

1. **Importance of Credit Default Prediction and Background of American Express**

Nowadays, credit card has become one of the most popular payment methods. With credit cards, customers could overdraft money from their bank account without any interest, and repay the bill later. Moreover, credit cards give customers rewards in the ways of cash back, air miles, movies discounts, and so on, which definitely makes credit card more competitive than other payment methods. However, it is challenging for banks and card service companies to decide how much credit limits available to clients with diverse backgrounds. The clients’ abilities to repay the bill, and probabilities of credit default are various based on their age, gender, income, credit record, education, marital status, previous repayment status [1]. If a high credit limit is provided to a client who has less ability to pay large amount bill, then the credit card company is more likely to face the risk of financial loss due to credit default. On the contrary, if a low credit limit is provided to a client with higher spending power and better repayment ability, this will cause not only bad customer experience, but also the credit card company cannot make benefit as much as it can. In this condition, credit default prediction should be a central business for banks and credit card service companies.

American Express, founded in 1850, is a famous company headquartered at New York City with specialization in bank card services. In 2016, American Express network has 22.9% of the total dollar volume of credit card transactions in the United States [2]. Up to December 31, 2021, American Express had 121.7 million cards in active worldwide, with an average annual spending of $20392 [3]. As a super-sized bank card service company, the risks and returns of American Express company are largely depend on the performance of credit default prediction. There is a Global Decision Science team in the company responsible for managing enterprise risks, one of the main jobs of this team is creating machine learning powered models to reduce risks [4]. Comparing with other statistical and data analyzing methods, machine learning is able to process large number of data sets, develop precise models, and more flexible. Therefore, machine learning is the most common method to predict credit default.

1. **How to Use Machine Learning to Predict Default**

To learn about processes of credit default prediction, it is necessary to know how does machine learning work. The first step of machine learning is collecting data. Sufficient amount of data is indispensable to develop an accurate machine learning model. The data will be separated into two groups, one is used to train the model, the other is used to test the model and observe its performance.

For the collected raw data, it cannot be used directly as input for a machine learning model, a data processing step called feature engineering must be done. Feature engineering includes data cleaning and feature selection [5]. In the raw data, some may have different units, some may have similar properties and need to be combined together, and some may be in a wrong format. Data cleaning is a process to make sure all the data is in a unity, and can be used by machine learning model smoothly. Feature selection is a process to delete unwanted features, and creating new features which are required. Since a rich amount of data is required to make the model more accurate during model training and testing, feature engineering is usually the most time-consuming step in machine learning [6].

The next procedure, which is the critical one, is selecting an appreciate machine learning model. The strengths and weakness are different among models, so it is important to select the best one based on the user’s requirement. For example, if the user is seeking for a simple and fast operating method, then logistic regression will be a good choice, but this method is easy to underfit and it requires large amount of data [6]. Another example is decision tree, its structure is easy to understand, and relatively less data pre-processing is required, however, it doesn’t have good performance for high-dimension data, and may overfitting due to a complex tree structure [6]. Since credit default prediction is a binary classification problem [6], which classifies clients to default and not default, any machine learning methods applicable to binary classification is theoretically usable, such as KNN, decision tree, SVM, logistic regression, ensemble learning [7]. In the article *Machine Learning: Challenges, Lessons, and Opportunities in Credit Risk Modeling* by Dinesh Bacham and Dr. Janet Zhao, they concluded that random forest and boosting, two ways of ensemble learning, have the best performance in credit default prediction [8].

1. **The Usefulness and Application of Credit Default Prediction**

As mentioned previously, credit default prediction can affect how much risks the card service company will undertake, and how many profits it can make for its credit card business. Except for American Express, credit default prediction is also non-negligible for many other companies and institutions. One example is loan companies or mortgage lenders. Nowadays, home loans and auto loans are popular because many people would like to pay their houses or cars with installment plans rather than full payment in one time. Before approving a loan, lenders have to complete risk assessment to borrowers, in this condition, credit default prediction could be an appreciate method. With the prediction result, they could deny or charge higher rates for risky borrowers, and also provide more loans for genuine borrowers to maximize benefits. Moreover, credit default can also be used by credit score companies, credit scores can be determined based on people’s personal information such as payment history, amounts owed, credit history length, diverse credit accounts, and new accounts recently opened [9]. In machine learning, users just need to set required features as input variables, and then select the best model which fits the data.

In conclusion, credit default prediction by machine learning is very flexible and can be widely used. It could help financial industries reduce risk, maximize profits, and also improve clients’ experience , therefore, it is valuable for people to research and work on this project.

**4. Reference**

[1] M. Zhang, “Credit default prediction based on machine learning models,” *Medium*, 06-Sep-2021. [Online]. Available: https://medium.com/analytics-vidhya/credit-default-prediction-based-on-machine-learning-models-1717601600c9. [Accessed: 11-Sep-2022].

[2] A. McCann, “Market share by Credit Card Network,” *WalletHub*, 27-Jul-2022. [Online]. Available: https://wallethub.com/edu/cc/market-share-by-credit-card-network/25531. [Accessed: 11-Sep-2022].

[3] *AXP-20211231*. [Online]. Available: https://www.sec.gov/Archives/edgar/data/4962/000000496222000008/axp-20211231.htm. [Accessed: 18-Sep-2022].

[4] “American Express - default prediction,” *Kaggle*. [Online]. Available: https://www.kaggle.com/competitions/amex-default-prediction/overview/prizes-and-hiring. [Accessed: 18-Sep-2022].

[5] H. Patel, “What is feature engineering - importance, tools and techniques for machine learning,” *Medium*, 02-Sep-2021. [Online]. Available: https://towardsdatascience.com/what-is-feature-engineering-importance-tools-and-techniques-for-machine-learning-2080b0269f10. [Accessed: 18-Sep-2022].

[6] Zornitsa, “Credit default risk prediction with machine learning,” *Record Evolution*, 25-Feb-2022. [Online]. Available: https://www.record-evolution.de/en/blog/credit-default-risk-prediction/. [Accessed: 18-Sep-2022].

[7] T. Aditya Sai Srinivas, S. Ramasubbareddy, and K. Govinda, “Loan default prediction using Machine Learning Techniques,” *SpringerLink*, 01-Jan-1970. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-16-8987-1\_56. [Accessed: 18-Sep-2022].

[8] “Machine learning: Challenges, lessons, and opportunities in credit risk modeling,” *Machine Learning: Challenges and Opportunities in Credit Risk Modeling*. [Online]. Available: https://www.moodysanalytics.com/risk-perspectives-magazine/managing-disruption/spotlight/machine-learning-challenges-lessons-and-opportunities-in-credit-risk-modeling. [Accessed: 18-Sep-2022].

[9] Stefan.lembo-Stolba, “What affects your credit scores?,” *Experian*, 29-Jul-2022. [Online]. Available: https://www.experian.com/blogs/ask-experian/credit-education/score-basics/what-affects-your-credit-scores/. [Accessed: 18-Sep-2022].