

Credit Score Classification

Kyle Jow, Jimmy Nguyen, Kyle Pickle, Jacob Tuttle, Kris Wong

December 2023

Introduction and background

A common problem in the finance and banking industry is assessing the risk involved in lending someone money. Typically, a credit score is used to determine whether a person is a stable borrower or not. This can be a difficult task with many factors to take into account when deciding someone's credit score.

Literature review

One of the first fields machine learning techniques were tested in was economics, particularly credit scoring. Researchers have utilized a variety of machine learning algorithms to predict credit scores and risk. Common machine learning algorithms used include logistic regression, support vector machines (SVM), and decision trees. This section contains relevant work by researchers.

Dumitrescu et. al. used a particular model for credit score classification with an improved logistic regression model that has non-linear decision tree effects. They created the penalised logistic tree regression, which predicted credit score more accurately than the benchmark logistic model commonly used in industry. The datasets they used to test the robustness of the model ——— Additionally, they argue that this model preserves the interpretability of logistic regression, an aspect that makes logistic regression popular in industry. Lei et.al. compare the performance of various models for predicting the probability of default.

Dataset description and exploratory data analysis

The dataset contains 27 features. The features we will be focusing on in order to classify a person's credit score into either good, standard, or poor credit will be these features: Credit Mix, Number of days past payment due date, Number of delayed payments, Outstanding debt, Credit utilization, Credit history age. We make this decision based on how traditional credit scores are determined, such as a FICO credit score.

Proposed methodology

Experimental results

Conclusion and discussion

References

Dataset

<https://www.kaggle.com/datasets/parisrohan/credit-score-classification/data>

Literature Review

Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, 297(3), 1178-1192. doi:10.1016/j.ejor.2021.06.053

Yue, Lei, and Luka Vidovic. “Machine Learning and Credit Risk Modelling.” *Machine Learning and Credit Risk Modelling | S&P Global Market Intelligence*, S&P Global Market Intelligence, 30 Nov. 2020, www.spglobal.com/marketintelligence/en/news-insights/blog/machine-learning-and-credit-risk-modelling.