Machine Learning Methods for Classifying Bad Borrowers in a Credit Portfolio

Jorge Reyes

University of Tennessee, Knoxville

COSC 522 Machine Learning

December 3, 2024

Contents

1 Introduction

Abstract

Financial Services companies chose to do a credit card dataset with the goal of creating a model that will predict whether customers are 'good' or 'bad' borrowers. The dataset contains enough features to build a classification model on the data. We will start by identifying what a 'bad' borrower is by looking at the portion of the dataset indicating whether or not a customer has missed payments throughout the lifetime of the loan. The dataset contains two files, contains information monthly status of a loan by customer for the entirety of the loan. contains fields such as employment status, annual income and other pieces of data we can use as our features to classify someone as 'good' or 'bad'. Since this is a classification problem, we will be using a supervised learning approach. We plan to use models such as Logistic Regression, Random Forest, and Gradient Boosting. The benefit to completing this project is that it would provide a bank or other financial institution with a credit scoring model that can assist in extending credit to customers and assessing risk of their loan portfolio.

2 Literature Review

Summarize existing research related to your topic. Highlight the gaps in the literature that your research aims to fill.

3 Methodology

Describe the methods used in your research. Include details about data collection, experimental design, tools, or techniques employed.

4 Results

Present the findings of your research. Use tables, figures, and charts to illustrate your results if necessary.

5 Discussion

Interpret your results and discuss their implications. Compare your findings with those of previous studies.

6 Conclusion

Summarize the key points of your research. State the main conclusions and suggest potential areas for future research.

References