Midterm Project

Lalida Kungval

03/22/2024

Introduction

The paper "Credit Approval Data Analysis Using Classification and Regression Models" delves into the crucial financial domain of credit risk assessment, employing statistical methods to predict credit worthiness. The assessment of credit approval is a pivotal aspect of the financial sector, where precise judgment of creditworthiness plays a critical role in mitigating risk while facilitating access to capital. Therefore, with the crucial consideration of developing fair and effective credit scoring models, this paper not only employs sophisticated statistical techniques, such as logistic regression and Classification and Regression Trees (CART), to predict creditworthiness but also opens a dialogue on the moral implications of these methodologies.

Summary of Method

The methodology employed in the paper includes an initial exploratory data analysis (EDA) to understand the dataset's key characteristics, followed by preprocessing steps like handling missing values and transforming data types. Data is also split into test and training sets. The study then applies logistic regression and Classification and Regression Trees (CART) to classify applicants into 'good' or 'bad' credit risks. Performance metrics, particularly accuracy, are used to evaluate the models, with results presented through confusion matrices.

Summary of Results

Logistic regression analysis identified significant predictors of credit approval, with the model achieving around 77% accuracy. For the logistic regression model, while coefficients for variables like *AgeNorm* and *Debt3* were not significant, others such as *YearsEmployedLog*, *CreditScoreLog*, and *IncomeLog* were significant predictors of credit approval. On the other hand, the CART model achieved higher prediction accuracy of 86.1%. For the CART model, *PriorDefault* is a crucial variable in classifying credit risk.

Description of Normative Consideration

The ethical dimension of automated credit scoring systems emerges as a critical normative consideration. The potential for biases and discrimination in model-driven decisions raises important questions about fairness and transparency in credit lending, highlighting the necessity for ethical scrutiny in statistical methodologies employed in financial decision-making. The philosophical aspect might not be immediately evident but can be derived from the implications of credit scoring practices on individuals and society. The ethical considerations revolve around fairness, bias in credit decisions, and the impact of these models on different demographic groups.