Development of Classification Models for Predicting Credit Limit Approval of Business Customers

Author : Napapha Phonthip Advisors : Assistant Professor Dr. Phimphaka Taninpong and Dr. Salinee Thumronglaohapun
Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand
E-mail : Napapha_ p@cmu.ac.th

Abstract

This study aims to analyze the general characteristics of business customers applying for credit limit approval, develop classification models using Decision Tree and Binary Logistic Regression techniques to predict credit approval outcomes, and compare the performance of these models. The sample consisted of 300 business customers, and the dataset was divided into training and testing sets at a ratio of 90:10. The dataset contains fifteen independent variables and a dependent variable (approval status). The results revealed that key factors influencing credit limit approval were registered capital, latest shareholders' equity, the presence of a guaranteed letter, shared credit utilization, and net profit. The decision tree model achieved balanced classification performance with an accuracy of 82.76% and an F1-score of 82.62%, whereas the Binary Logistic Regression model produced lower accuracy of 72.41% and F1-score of 60.00%. This findings indicate that the decision tree model outperforms Binary Logistic Regression and is more suitable for supporting decision-making in credit limit approval processes.

Introduction

Credit provision plays an important role in supporting business operations and stimulating economic growth by increasing financial liquidity for entrepreneurs and enabling business expansion. For lending organizations, appropriate credit approval is an essential tool for managing financial risk and strengthening long-term business relationships with customers. However, credit approval decisions may sometimes be influenced by external factors such as company reputation, business size, or existing relationships, which may increase the risk of future credit losses. Therefore, this study aims to develop classification models for predicting credit limit approval for business customers using financial and company-related variables.

Objectives

- To examine the general characteristics of business customers applying for credit limits using descriptive statistics, comparing approved and non-approved groups.
- To develop classification models for predicting credit limit approval using Decision Tree and Binary Logistic Regression.
- To compare the predictive performance of the two classification models.
- To identify key factors influencing credit limit approval for business customers.

Methodology


1.Data Collection

2.Data Preparation

3.Model Development

4.Model Evaluation

5.Summary

Variables

Dependent variable

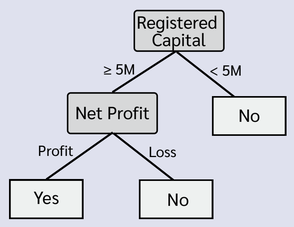
Credit Approval Status (0 = No, 1 = Yes)

Independent variables

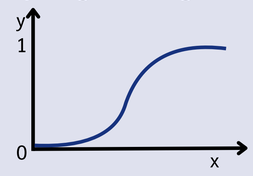
X_1 = Registered Capital
X_2 = Years of Registration
X_3 = Previous Year Equity
X_4 = Latest Year Equity
X_5 = Previous Year Revenue
X_6 = Latest Year Revenue
X_7 = Previous Year Net Profit
X_8 = Latest Year Net Profit
X_9 = Liquidity Ratio
X_{10} = D/E Ratio
X_{11} = Net Profit Margin
X_{12} = Shared Credit Line
X_{13} = Bank Guarantee (BG)
X_{14} = Quotation
X_{15} = Existing Customer

Models

Decision Tree



Binary Logistic Regression



$$logit(P) = \ln\left(\frac{P}{1-P}\right) = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n$$

Evaluation

Confusion Matrix

		Actual Values: Positive (+)	Actual Values: Negative (-)
Predicted Values	Positive (+)	TP	FP
Predicted Values	Negative (-)	FN	TN

Evaluation Metrics

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

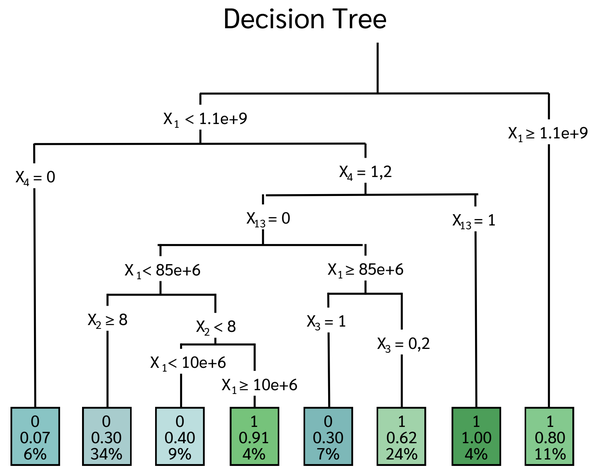
$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Results

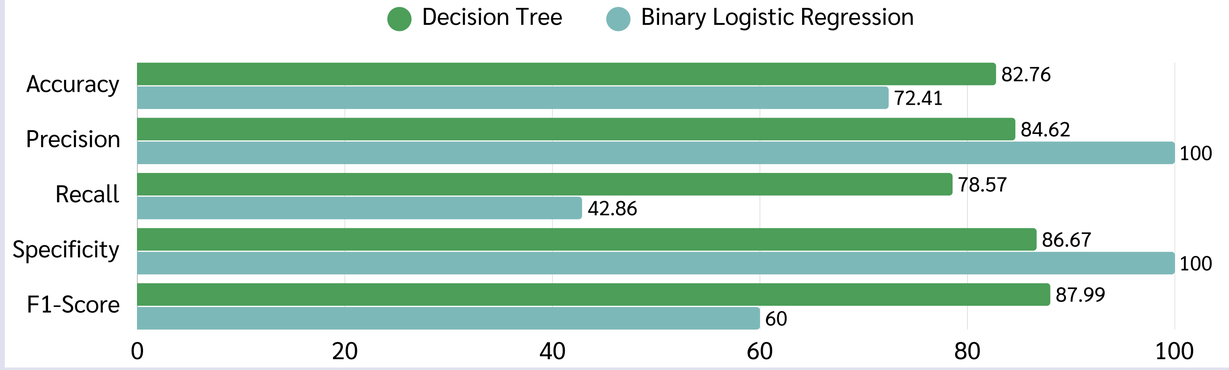
Binary Logistic Regression

Variable	Estimate (β)	S.E.	Wald	df	p-value	Exp (β)
Constant	-2.55	0.981	6.763	1	0.0093	0.078
X_1	2.285×10^{-10}	1.036×10^{-10}	4.866	1	0.0274	1.00
X_{3_1}	-1.5	0.984	2.323	1	0.128	0.223
X_{3_2}	-0.184	1.008	0.033	1	0.856	0.832
X_{4_1}	2.822	1.288	4.799	1	0.028	16.82
X_{4_2}	2.463	1.357	3.295	1	0.07	11.74
X_{6_2}	-1.019	0.479	4.522	1	0.033	0.361
X_{8_1}	1.076	0.514	4.382	1	0.036	2.933
X_{8_2}	1.084	0.575	3.559	1	0.059	2.956
X_{12_1}	2.01	0.759	7.015	1	0.008	7.465
X_{13_1}	3.14	1.026	9.356	1	0.002	23.1
X_{14_1}	0.555	0.367	2.293	1	0.13	1.742
-2 Log Likelihood (Residual deviance) = 316.79						
Chi-square (Null vs Model) = 58.60, df = 11						
Cox & Snell $R^2 \approx 0.136$, Nagelkerke $R^2 \approx 0.181$						

Decision Tree



Comparison of Decision Tree and Binary Logistic Regression



Conclusion

This study examined the factors influencing business credit approval and compared the performance of Decision Tree and Binary Logistic Regression models. The results indicate that key factors affecting credit approval include registered capital, shareholders' equity, net profit, joint credit usage, and the presence of a bank guarantee. Firms with stronger financial stability and higher liquidity are more likely to receive credit approval. The Decision Tree model demonstrated better classification performance, achieving an accuracy of 82.76% and an F1-score of 82.62%, indicating balanced prediction between approved and non-approved cases. In contrast, Binary Logistic Regression effectively identified statistically significant factors and showed high precision and specificity, but its lower recall indicates limitations in detecting all approved cases. Overall, the Decision Tree model is more suitable as a decision-support tool for credit approval, while Binary Logistic Regression is useful for analyzing significant factors and preliminary credit risk assessment.

Limitations

- The dataset contains a limited number of samples, which may affect model generalization.
- Some potential influencing variables could not be included due to data availability constraints.

References

- National Credit Bureau. (n.d.). *Frequently asked questions (FAQ).*
- Wongwiwat, P., Chairojwattana, A., & Phaphan, W. (2023). *Machine learning approaches for credit risk assessment in SMEs.*