International Journal of Science and Business 2025, 47(1), 30-43
ISSN: 2520-4750 (Online) 2521-3040 (Print)
Journal: ijsab.com/ijsb

Small and Medium Sized Enterprise Credit Customer's Insolvency Prediction by using Two Group Discriminant Analysis: A Case Study of Bangladesh

Md. Abdul Maleque [1] , Md. Abdul Matin [2] , & Subrata Deb Nath [3,*]

[1] Deputy General Manager, Janata Bank PLC, Dhaka-1000, Bangladesh.
[2] General Manager, Janata Bank PLC, Dhaka-1000, Bangladesh.
[3] Senior Principal Officer, Janata Bank PLC, Dhaka-1000, Bangladesh.
* Corresponding author: Subrata Deb Nath (subratadebnath81@gmail.com)

Abstract
This study develops a discriminant function to predict the creditworthiness of Small and Medium Enterprises (SMEs) in Bangladesh, aiming to distinguish between default and non-default borrowers. Data were collected from 20 SME credit customers of a large commercial bank, evenly split between defaulters and non-defaulters. Six independent variables related to financial and socio-economic characteristics were analyzed to build the predictive model. The discriminant analysis identified Loan Position Against Portfolio (LPAP), loan amount, and number of employees as the most significant factors influencing credit risk classification. The estimated discriminant function demonstrated statistical significance at the 1% level, indicating a strong model fit. When applied to the dataset, the model correctly classified 95% of the original sample cases and maintained a 70% accuracy rate under cross-validation, confirming its robustness and practical utility. The function enables calculation of a discriminant score (Z-score) for new loan applicants, which can be used to predict their likelihood of default. Positive scores indicate higher default risk, while negative scores suggest lower risk. Implementing this discriminant function can improve credit risk management by providing a systematic, data-driven tool to assist banks in loan approval decisions, risk-based pricing, and resource allocation. The model offers a cost-effective approach to reduce non-performing loans and enhance portfolio quality. However, the study is limited by its small sample size and scope, which calls for further research with larger and more diverse datasets. Future studies could also explore the inclusion of additional variables and advanced modeling techniques to improve predictive accuracy and adaptability across different economic conditions.

ARTICLE INFO
Research paper
Received: 11 May 2025
Accepted: 03 July 2025
Published: 05 July 2025
DOI: 10.58970/IJSB.2603

CITATION
Maleque, M. A., Matin, M. A. & Deb Nath, S. (2025). Small and Medium Sized Enterprise Credit Customer's Insolvency Prediction by using Two Group Discriminant Analysis: A Case Study of Bangladesh, *International Journal of Science and Business*, 47(1), 30-43.

***Keywords**: SME Credit Risk, Discriminant Analysis, Loan Default Prediction, Financial Variables, Credit Scoring, Bangladesh Banking.*

1. Introduction

Small and Medium Enterprises (SMEs) are widely recognized as critical drivers of economic growth, employment generation, and innovation across the globe. Their role is particularly significant in developing economies, where they often serve as the backbone of industrial activity and act as catalysts for inclusive development. Ayyagari, Beck, and Demirgüç-Kunt (2007) estimate that SMEs account for more than 95% of registered firms worldwide, contributing up to 60% of employment and a substantial share of GDP, depending on the country context. In their cross-country analysis, Beck, Demirgüç-Kunt, and Levine (2005) further found that SME development is closely associated with reductions in income inequality and poverty, although the mechanisms vary across financial and institutional settings. In Bangladesh, SMEs hold immense importance for the national economy. They contribute approximately 25% to GDP and employ about 40% of the total labor force (Ahmed, 1999). These enterprises span diverse sectors including manufacturing, agriculture, trade, and services. While the government and development partners have taken steps to support SME development, the sector continues to face considerable constraints—chief among them being limited access to finance, a lack of formal credit history, and regulatory burdens (Chowdhury & Alam, 2017). The inability of many SMEs to secure timely and adequate financing often results in liquidity shortages and, ultimately, financial distress or insolvency. From a financial sector perspective, lending to SMEs poses a distinctive challenge. Unlike large corporations, SMEs typically lack audited financial statements, credit ratings, or adequate collateral, making it difficult for banks to assess their creditworthiness using conventional methods. Berger and Udell (2006) argue that asymmetric information and high transaction costs make SME financing riskier and costlier for financial institutions, particularly in underdeveloped financial systems like that of Bangladesh. This situation necessitates the development and use of structured, data-driven credit evaluation models that can effectively manage and mitigate credit risk.

Credit risk—defined as the potential that a borrower will fail to meet its obligations—is a fundamental risk faced by banks. Credit allocation, which involves evaluating loan applications, is among the most critical processes in banking. Traditionally, this process relied heavily on expert judgment, which, while valuable, can also be subjective, inconsistent, and susceptible to bias. With the growing complexity of banking operations and the expansion of retail and SME lending, statistical tools and predictive modeling techniques have become indispensable in enhancing the accuracy and efficiency of credit decisions (Thomas, Crook, & Edelman, 2002). Among various quantitative methods, discriminant analysis has emerged as a widely used and well-established technique for predicting borrower behavior. Altman's (1968) pioneering work introduced discriminant analysis to the finance domain through his Z-score model, which utilized financial ratios to predict corporate bankruptcy. His model demonstrated that statistical classification techniques could outperform intuitive approaches in identifying financially distressed firms. This breakthrough led to a wave of research applying discriminant analysis to personal and business credit scoring. Wiginton (1980) compared logit and discriminant models and found that discriminant analysis was particularly effective in situations where borrowers could be distinctly classified into good and bad credit categories. Hand and Henley (1997) further reinforced the relevance of discriminant analysis, showing that it remains a reliable method when historical data are limited and the sample size is relatively small—conditions often encountered in SME lending scenarios in developing countries. Their review concluded that discriminant and logistic models remain robust foundations for most consumer credit scoring systems, especially when augmented with socio-demographic and behavioral data.

In the context of developing economies, studies such as that by Dinh and Kleimeier (2007) have shown that integrating variables like age, education, employment history, residential status, and income into credit scoring models significantly improves prediction accuracy. Their work, which developed a credit scoring model for Vietnam's retail banking sector, is particularly relevant for Bangladesh, as both countries share similar financial market structures and SME constraints.

Despite the global progress in risk modeling, such practices are still at a nascent stage in Bangladesh, especially in the domain of SME finance. Most commercial banks continue to rely heavily on manual screening and qualitative assessments. These methods often fail to capture the nuanced financial behavior of SMEs and overlook early signs of insolvency. A structured statistical model, such as one based on discriminant analysis, offers the advantage of objectivity, scalability, and transparency—traits that are vital in a rapidly expanding credit market.

The issue of insolvency is particularly important in the SME segment. Insolvency among SMEs not only results in financial losses for lenders but also disrupts supply chains, causes job losses, and affects broader economic stability. Therefore, early prediction of insolvency risk allows banks to adopt preventive strategies—such as risk-based pricing, enhanced monitoring, or credit restructuring—to minimize losses and support borrower sustainability. This study aims to develop a discriminant function to predict the financial status of SME borrowers in Bangladesh, distinguishing between those likely to default and those likely to remain solvent. The model is built using data collected from a large commercial bank in Bangladesh, focusing on six key variables: loan amount, project finance percentage, number of employees, owner's education, net wealth, and experience. These variables were selected based on prior empirical studies (e.g., Altman, 1968; Wiginton, 1980; Dinh & Kleimeier, 2007) and adjusted for the Bangladeshi context based on data availability and expert consultations.

2. Literature Review

Small and Medium Enterprises (SMEs) are universally acknowledged as crucial drivers of economic growth, employment generation, and poverty alleviation, especially in developing countries such as Bangladesh. The literature on SME finance and credit risk management has grown substantially over the last several decades, addressing both the role of SMEs in economic development and the challenges related to their access to finance and insolvency risks. This literature review synthesizes prior empirical and theoretical research on SME financing, credit risk prediction methods, and the application of discriminant analysis and related techniques for insolvency forecasting, drawing on foundational works as well as recent studies pertinent to the Bangladeshi context. The seminal work of Altman (1968) established the use of financial ratios combined with discriminant analysis as a pioneering method for predicting corporate bankruptcy. Altman's Z-score model utilized key financial ratios to distinguish between bankrupt and non-bankrupt firms, demonstrating the power of discriminant analysis as an effective tool in credit risk evaluation. This model laid the groundwork for subsequent research focused on applying quantitative techniques for insolvency prediction in various sectors and countries. Altman's approach underscored the importance of using multiple financial indicators simultaneously to achieve high predictive accuracy, a concept that has been adapted and expanded in SME credit risk assessment. Expanding beyond corporate bankruptcy, the literature emphasizes the critical role SMEs play across economies globally. Ayyagari, Beck, and Demirgüç-Kunt (2007) provide a comprehensive overview of SMEs worldwide, highlighting their contributions to innovation, job creation, and economic dynamism. Their cross-country analysis reveals systemic barriers that SMEs face, notably in accessing formal credit markets due to perceived higher risks and information asymmetries. Similarly, Beck, Demirgüç-Kunt, and Levine (2005) offer evidence on the relationship between SMEs, economic growth, and poverty reduction, underscoring how improved SME finance can enhance inclusive growth in emerging economies. Their findings reinforce the importance of reliable credit risk assessment tools to facilitate SME financing. Berger and Udell (2006) further elaborate a conceptual framework that encompasses both demand- and supply-side constraints faced by SMEs in obtaining finance. They argue that the informational opacity of SMEs and the high transaction costs of lending to smaller borrowers compel financial institutions to rely heavily on quantitative risk assessment models. The authors advocate for a more nuanced understanding of SME finance that integrates traditional financial ratios with behavioral and qualitative factors, a theme echoed by subsequent research on credit scoring and risk prediction.

In the context of Bangladesh, SMEs constitute a substantial portion of the economy, but face unique challenges in finance, infrastructure, and regulatory environments. Chowdhury and Alam (2017) investigate these constraints at the micro-level in Khulna City, identifying limited access to formal credit as a critical bottleneck for SME growth. Their findings suggest that banks' risk aversion towards SME lending partly stems from inadequate credit risk assessment mechanisms, which often fail to capture the multifaceted nature of SME business operations. This calls for tailored credit scoring models that can accurately differentiate between solvent and insolvent SME borrowers in Bangladesh. A rich body of literature has focused on developing and refining statistical techniques to predict credit risk and financial distress. Discriminant analysis, logistic regression, and machine learning methods have been widely employed in consumer and SME credit risk modeling. Wiginton (1980) was among the early researchers to apply discriminant analysis to consumer credit behavior using demographic and economic variables. His study demonstrated that employment status, living arrangements, and occupation types were significantly related to credit risk ratings, indicating the importance of socio-economic factors in credit evaluation beyond pure financial metrics. Grablowsky (1975) contributed to this field by starting with a broad set of 36 variables encompassing behavioral, financial, and demographic characteristics, and through sensitivity analysis narrowed the model to 13 key predictors for consumer credit risk. Despite certain statistical assumption violations, his discriminant model achieved an impressive 94% accuracy in classifying credit risk, underscoring the method's practical viability. This study highlighted the balance between model complexity and predictive power, a consideration critical to SME credit scoring where data availability can be limited. Hand and Henley (1997) provide a thorough review of statistical classification methods in credit scoring, including discriminant analysis and logistic regression. They identify a broad range of borrower characteristics—such as time at current address, homeownership, income, age, and employment history—as typical predictors distinguishing problematic from regular customers. Their work emphasizes that while traditional discriminant analysis assumes equal covariance matrices among groups, logistic regression often provides a more flexible alternative, especially when assumptions are violated. Dinh and Kleimeier (2007) applied logistic regression to develop a credit scoring model for Vietnam's retail banking sector. Their research identified time with the bank, gender, number of loans, and loan duration as the most important predictors of credit risk. This study illustrates how demographic and behavioral data, combined with loan performance history, improve the accuracy of credit risk models in emerging markets. Such empirical findings resonate with Bangladesh's SME credit environment, suggesting similar predictive variables may be relevant. Nasir Uddin (2013) provides a pertinent case study of consumer credit customers' financial distress prediction in Bangladesh using two-group discriminant analysis. Employing thirteen demographic, socio-economic, and loan-related variables, Uddin's model delivered a faster and more cost-effective credit disbursement process with improved accuracy. This study demonstrated that applying quantitative credit scoring models to the Bangladeshi market can significantly enhance credit management efficiency, enabling risk-based pricing and reducing non-performing loans. More broadly, Thomas, Crook, and Edelman (2002) offer a comprehensive overview of credit scoring methodologies, discussing their application across consumer and SME lending contexts. Their work highlights the evolution of credit scoring from early discriminant analysis to modern machine learning techniques, reflecting the ongoing efforts to improve credit risk prediction accuracy and adapt to changing financial environments. Wiginton (1980) compared logit and discriminant models for consumer credit risk, concluding that while both methods have merits, discriminant analysis remains a robust tool under certain data conditions. This foundational research supports the ongoing use of discriminant functions in credit risk evaluation, particularly when sample sizes are small and variables meet necessary statistical assumptions. The collective insights from these studies emphasize the critical need for robust, data-driven credit risk assessment models tailored to SMEs. In developing countries like Bangladesh, where informal business practices and limited financial documentation are common, credit risk models must integrate both financial ratios and socio-economic variables to capture borrower risk comprehensively. Finally, Ahmed (1999) provides an essential overview of SME development in Bangladesh, highlighting institutional and financial challenges SMEs face. The report underscores the importance of enhancing credit risk evaluation tools as a prerequisite for expanding SME access to finance, which is vital for broader economic growth and poverty alleviation. Moreover, the literature strongly supports the application of discriminant analysis and similar quantitative methods to predict SME credit risk and insolvency, particularly when combined with relevant demographic and socio-economic variables. The evidence from various global and local studies illustrates that accurate insolvency prediction models contribute not only to better credit decision-making but also to the overall sustainability of SME sectors in emerging economies. This study builds upon these foundations by applying two-group discriminant analysis to SME credit customers of a large commercial bank in Bangladesh, aiming to develop a reliable function for insolvency prediction that addresses local market characteristics and data constraints.

3. Methodology

3.1 Data Collection

This research focused on variables extracted from loan application forms used by a major commercial bank in Bangladesh. Accessing such confidential information is often challenging; however, through certain permissions and cooperation from the bank branch, a total of 40 completed loan applications were initially obtained. Out of these, 20 applications were excluded due to insufficient or incomplete information, resulting in a final sample size of 20 loan applications. These comprised 10 cases where borrowers defaulted and 10 cases where borrowers maintained good standing. The combined dataset of these 20 cases was designated as the analysis sample, while an additional subset was reserved as the holdout or validation sample for model verification purposes.

3.2 Data Analysis

To fulfill the research objectives, the study employed the direct method of discriminant analysis to examine the dataset. This approach involves including all predictor variables simultaneously in the model, without preliminary screening based on their individual discriminatory strength. An alternative strategy, known as stepwise discriminant analysis, selects variables progressively according to their ability to distinguish between groups. For this study, however, the direct method was preferred due to the consideration of multiple demographic and socio-economic factors of the borrowers. Data processing and statistical analysis were conducted using SPSS software.

3.3 Variable Description

The study categorized variables into dependent and independent groups. The dependent variable indicated the financial status of the borrower: a value of 1 represented a borrower in good standing, whereas a value of 2 represented a borrower in default. The independent variables included those related to the loan specifics and the borrower's demographic and socio-economic background. Specifically, loan-related predictors comprised loan amount, percentage of project finance, number of employees, educational level, net wealth, and years of experience.

4. Results and Discussions

4.1 Conducting the Discriminant Analysis

Discriminant analysis is a widely used parametric statistical method designed to differentiate between two or more groups based on predictor variables. It is particularly popular in credit risk modeling to classify borrowers as either good credit risks or bad credit risks. This technique helps in creating a discriminant function—a linear combination of the independent variables—that maximizes the distinction between groups. The process of conducting discriminant analysis generally involves several key steps: formulating the discriminant function, estimating coefficients, assessing the significance of predictors, interpreting the model, and validating the predictive accuracy. In this study, discriminant analysis was applied to classify Small and Medium Enterprises (SMEs) into default and non-default categories based on various financial and socio-economic factors. The objective was to identify which variables most effectively discriminate between the two groups, thereby assisting in improving credit evaluation processes.

4.2 Group statistics

The initial stage of the analysis involves comparing the two groups—default and non-default SMEs—by evaluating their group means and standard deviations across the independent variables. This comparison helps identify which variables show significant differences between the groups and can thus serve as potential discriminators in the model. Table 1 presents the descriptive statistics for each variable, segregated by group. As observed, there are clear differences in means between the default and non-default groups for several key variables. For instance, the average loan amount for non-default SMEs (3,424.29) is considerably higher than that of the default group (1,326.95), suggesting that businesses with larger loan amounts in this sample are less likely to default. Similarly, the number of employees is greater among non-default SMEs (mean of 480.33) compared to defaulting ones (255.70), possibly indicating that larger operational scale correlates with better credit performance. The variable "Loan Position Against Portfolio" (LPAP) also shows a substantial gap between groups, with non-default borrowers having a mean LPAP of 2,824.53 versus 1,087.06 for defaulting borrowers. This may reflect differences in the relative risk or portfolio exposure attributed to each loan. Net worth follows the same pattern, although the wide standard deviation suggests considerable variability within the groups. Conversely, variables such as Education and Owner's Experience do not exhibit marked differences between default and non-default groups. Education levels are relatively similar (16.22 vs. 15.30), and owner experience shows negligible variation (16.22 vs. 15.10), indicating that these socio-economic characteristics might not be strong discriminators for credit default in this context. By examining these group statistics, it becomes evident which variables could be most influential in the discriminant function, guiding further statistical testing and model refinement.

Table 1: Group statistics

Group		**Mean**	**Std. Deviation**
1	Loan	3424.29	2705.42
	Emp	480.33	360.63
	Edu	16.22	2.488
	LPAP	2824.53	1803.39
	NW	1699.88	7957.90
	Exp	16.22	12.03
2	Loan	1326.95	333.19
	Emp	255.700	191.10
	Edu	15.30	2.35
	LPAP	1087.06	536.48
	NW	529.04	1446.06
	Exp	15.10	5.23
Total	Loan	2320.43	2113.32
	Emp	362.10	298.90
	Edu	15.73	2.40
	LPAP	1910.07	1543.94
	NW	1,112.56	5522.57
	Exp	15.63	8.85

4.3 Tests of equality of group means

To further assess the ability of each independent variable to differentiate between default and non-default SME groups, tests of equality of group means were conducted using Wilks' Lambda and associated F-statistics. Wilks' Lambda is a key multivariate test statistic used in discriminant analysis to evaluate whether group means on a given variable are statistically different. The value of Wilks' Lambda ranges from 0 to 1, where values closer to 0 indicate greater differences in group means, and values approaching 1 suggest little or no difference between the groups. In this context, a smaller Wilks' Lambda signifies that the variable contributes significantly to distinguishing between the groups, making it a valuable predictor in the discriminant function. Conversely, variables with a Wilks' Lambda value near 1 have limited discriminatory power. Generally, a Wilks' Lambda value below 0.95 is considered acceptable to indicate meaningful group differences. Variables exceeding this threshold may be less impactful and could be considered for exclusion without significantly affecting the overall model. The results, as presented in Table 2, indicate that several variables demonstrate statistically significant differences between default and non-default groups. Specifically, the loan amount (Wilks' Lambda = 0.741, F = 5.948, p = 0.026), number of employees (Wilks' Lambda = 0.851, F = 2.968, p = 0.103), loan position against portfolio (Wilks' Lambda = 0.667, F = 8.497, p = 0.010), net worth (Wilks' Lambda = 0.957, F = 0.761, p = 0.395), education level (Wilks' Lambda = 0.961, F = 0.687, p = 0.419), and owner's experience (Wilks' Lambda = 0.996, F = 0.072, p = 0.792) were examined. Among these, loan amount and loan position against portfolio stand out as having significant p-values below the conventional 0.05 threshold, indicating their strong ability to differentiate between the two groups. Number of employees shows a moderate effect with a p-value slightly above 0.10, suggesting it may still hold some relevance. On the other hand, education level, net worth, and owner's experience display high Wilks' Lambda values and insignificant p-values, implying that these variables may not play a crucial role in the discriminant model. The F-statistic and corresponding p-values provide further insight into the statistical significance of these group mean differences. A lower p-value associated with the F-test indicates stronger evidence against the null hypothesis of equal means, reinforcing the variable's importance in group classification. Therefore, variables like loan amount and loan position against portfolio are essential components of the discriminant function due to their significant role in distinguishing default status. Moreover, these tests guide the selection of variables for the discriminant analysis, ensuring that the model focuses on predictors that meaningfully contribute to classifying borrowers. Variables with high Wilks' Lambda and non-significant F-tests may be excluded to enhance the model's efficiency without compromising predictive accuracy.

Table 2: Tests of equality of group means

Tests of Equality of Group Means					
	Wilks' Lambda	F	df1	df2	Sig.
Loan	.741	5.948	1	17	.026
Emp	.851	2.968	1	17	.103
Edu	.961	.687	1	17	.419
LPOP	.667	8.497	1	17	.010
NW	.957	.761	1	17	.395
Exp	.996	.072	1	17	.792

4.4 Structure matrix

The structure matrix presents the correlations between each independent variable and the discriminant function. These correlations are instrumental in identifying which variables contribute most strongly to distinguishing between the default and non-default borrower groups. The higher the absolute value of the correlation coefficient, the more significant the role of that variable in separating the two groups. As shown in Table 3, the variable with the highest correlation with the discriminant function is Loan Position Against Portfolio (LPAP), with a coefficient of 0.509. This indicates that LPAP is the most influential predictor in determining whether an applicant is likely to default or not. Following LPAP, Loan Amount (0.426) and Number of Employees (0.301) also exhibit relatively strong correlations, suggesting they play a substantial role in group discrimination. Conversely, variables such as Net Worth (–0.152), Education (0.145), and Owner's Experience (0.047) display comparatively low correlations. This suggests these variables have limited discriminating power in this particular dataset. While they may still offer some marginal insight into the classification, their contribution is minimal relative to the top three variables. This ranking of variables by discriminant strength is valuable when interpreting the results and making decisions about feature selection for credit risk models. Variables with larger absolute correlation values are prioritized for their predictive power, while those with weaker correlations may be reconsidered or excluded in future analyses to streamline the model.

Table 3: Structure matrix

Variables	Function 1
LPAP	.509
Loan	.426
Emp	.301
NW	-.152
Edu	.145
Exp	.047

4.5 Canonical discriminant function coefficients (unstandardized coefficients)

One of the central objectives of this study is to derive the discriminant function, which classifies loan applicants into either *default* or *non-default* groups based on their characteristics. The unstandardized discriminant function coefficients are particularly important because they allow for the formulation of a discriminant equation using the original scale of measurement for each variable, making the model practical for real-world use by banks or financial institutions. These coefficients serve as weights that multiply each independent variable to produce a discriminant score (Z). This score is then used to classify applicants based on their proximity to the group centroids for defaulters and non-defaulters. The discriminant function estimated from the analysis is as follows:

$$Z = -.26237 + .00038\ \text{Loan} + .00193\ \text{Employee's}$$
$$-\ .26555\ \text{Education} + 00135\ \text{LPAP} + 00031\ \text{Net}$$
$$-.00594\ \text{Experience}$$

To classify a new loan applicant, the values from the application form are substituted into this equation. The resulting Z score is then interpreted as follows:

- If Z is positive, the applicant is likely to belong to the default group. The greater the distance from zero in the positive direction, the higher the credit risk, indicating that the borrower may require rejection or a higher risk premium if credit is extended.
- If Z is negative, the applicant is predicted to fall into the non-default (or good) group. The more negative the score, the lower the credit risk, suggesting that the borrower is likely reliable, and the bank may consider offering better credit terms or lower interest rates.

This approach not only allows for binary classification but also enables risk-based pricing, as banks can use the Z score magnitude to scale interest rates according to perceived borrower risk.

Table 4: Canonical discriminant function coefficients

Variables	Function 1
Loan	.00038
Emp	.00193
Edu	-.26555
LPAP	00135
NW	.00031
Exp	-.00594
(Constant)	.26237

4.6 Group Centroids

An important interpretive element of discriminant analysis is the examination of group centroids, which represent the average discriminant scores (Z-values) for each group based on the predictive model. These centroids provide insight into how distinctly the groups are separated by the discriminant function. In a two-group discriminant analysis such as this, the centroids serve as reference points for classifying new observations. Each centroid is calculated by substituting the group mean values of the predictor variables into the unstandardized discriminant function. The resulting values indicate the central tendency of each group within the discriminant function's space. In this study:

The centroid for non-default clients is 1.385, and
The centroid for default clients is −1.246.

These values suggest a clear separation between the two groups on the discriminant dimension. The larger the absolute difference between the centroids, the stronger the discriminant function is in distinguishing the two categories. To classify a new loan applicant, their individual values for the predictor variables (e.g., loan amount, number of employees, education, etc.) are substituted into the discriminant function:

$$Z=-0.26237+0.00038(\text{Loan})+0.00193(\text{Employees})-0.26555(\text{Education})+0.00135(\text{LPAP})+0.00031(\text{Net Worth})-0.00594(\text{Experience})$$

$$Z=-0.26237+0.00038(\text{Loan})+0.00193(\text{Employees})-0.26555(\text{Education})+0.00135(\text{LPAP})+0.00031(\text{Net Worth})-0.00594(\text{Experience})$$

The resulting Z-score is then compared to the group centroids. If the Z-score is closer to the non-default centroid (1.385), the applicant is classified as low risk or non-default. Conversely, if the Z-score is closer to the default centroid (−1.246), the applicant is classified as high risk or default. This logic underpins a risk-sensitive loan approval process. Moreover, the distance between an applicant's Z-score and the centroids can be used to fine-tune loan terms, such as interest rates or collateral requirements, thereby enabling risk-based pricing. The group centroids are summarized below in Table 5. The relatively large separation between these centroid values (a gap of over 2.6 units) confirms that the discriminant function possesses meaningful predictive power. This strengthens the model's utility in supporting credit decision-making processes for SMEs in Bangladesh.

Table 5: Functions at Group Centroids

Group	Function
1	1.385
2	-1.246

4.7 Casewise statistic

Casewise statistics offer a comprehensive view of how well the discriminant function performs in classifying individual loan applicants into the correct group—either default (Group 2) or non-default (Group 1). Each observation is assessed based on whether the predicted group matches the applicant's actual group membership, along with the associated probabilities and the calculated discriminant scores. These results, summarized in Table 6, help evaluate both the predictive strength and practical applicability of the discriminant model. For each case, Table 6 presents the actual group, the predicted group, the probability of group membership, and the corresponding discriminant function score (Z-score). The table also provides the squared Mahalanobis distance of each observation from both group centroids. These distances offer insights into how far a given applicant is from the "center" of each group on the discriminant axis. The Z-score, calculated using the estimated function, represents the linear combination of predictor variables that determines the applicant's position relative to default risk. The results from Table 6 show that in the majority of cases, the model accurately classified applicants. For instance, Case 1 belongs to the non-default group (Group 1) and was correctly predicted as such with a high probability of 0.993 and a discriminant score of 1.969, indicating low risk. Similar accuracy is seen in Case 10, where the model predicted a non-default status with perfect classification confidence and a high Z-score of 3.361. These strong positive Z-scores suggest a strong alignment with non-default characteristics. Conversely, the model also performed well in predicting default clients. Case 11, which belongs to the default group (Group 2), was predicted correctly with a high level of certainty and a significantly negative Z-score of −2.511.

Table 6: Casewise statistic

	Case Number	Actual Group	Highest Group: Predicted Group	P(D>d | G=g): p	P(D>d | G=g): df	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid	Second Highest Group: Group	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid	Discriminant Scores: Function 1
Original	1	1	1	.559	1	.993	.341	2	.007	10.335	1.969
	2	1	1	.581	1	.882	.305	2	.118	4.321	.833
	3	1	2**	.332	1	.713	.940	1	.287	2.760	-.277
	4	1	1	.636	1	.991	.224	2	.009	9.634	1.858
	5	1	1	.203	1	.527	1.622	2	.473	1.842	.111
	6	1	1	.258	1	.998	1.277	2	.002	14.143	2.515
	7	1	1	.218	1	.555	1.516	2	.445	1.958	.153
	8	1	1	.275	1	.643	1.190	2	.357	2.371	.294
	9	1	1	.679	1	.990	.171	2	.010	9.267	1.798
	10	1	1	.048	1	1.000	3.907	2	.000	21.227	3.361
	11	2	2	.206	1	.999	1.601	1	.001	15.179	-2.511
	12	2	2	.370	1	.751	.804	1	.249	3.007	-.349
	13	2	2	.292	1	.666	1.110	1	.334	2.488	-.193
	14	2	2	.584	1	.993	.299	1	.007	10.098	-1.793
	15	2	2	.491	1	.838	.475	1	.162	3.768	-.557
	16	2	2	.973	1	.967	.001	1	.033	6.745	-1.213
	17	2	2	.933	1	.962	.007	1	.038	6.486	-1.162
	18	2	2	.436	1	.996	.607	1	.004	11.626	-2.025
	19	2	2	.645	1	.991	.212	1	.009	9.553	-1.706
	20	2	2	.768	1	.936	.087	1	.064	5.456	-.951
Cross-validated[b]	1	1	1	.000	6	.974	55.524	2	.026	62.800	
	2	1	1	.659	6	.773	4.133	2	.227	6.579	
	3	1	2**	.015	6	1.000	15.757	1	.000	40.786	
	4	1	2**	.000	6	.998	282.081	1	.002	295.061	
	5	1	2**	.526	6	.864	5.136	1	.136	8.838	
	6	1	1	.433	6	.999	5.912	2	.001	19.350	
	7	1	2**	.420	6	.888	6.028	1	.112	10.166	
	8	1	2**	.066	6	.981	11.821	1	.019	19.666	
	9	1	1	.000	6	.899	46.316	2	.101	50.691	
	10	1	1	.011	6	1.000	16.493	2	.000	41.528	
	11	2	2	.771	6	.999	3.297	1	.001	17.420	
	12	2	2	.465	6	.505	5.640	1	.495	5.684	
	13	2	1**	.690	6	.503	3.904	2	.497	3.930	
	14	2	2	.951	6	.991	1.625	1	.009	10.925	
	15	2	2	.261	6	.556	7.705	1	.444	8.151	
	16	2	2	.415	6	.929	6.076	1	.071	11.212	
	17	2	2	.950	6	.947	1.630	1	.053	7.383	
	18	2	2	.758	6	.995	3.395	1	.005	14.165	
	19	2	2	.993	6	.988	.753	1	.012	9.592	
	20	2	2	.149	6	.780	9.469	1	.220	12.002	

This pattern is consistent across other default cases, such as Cases 14 and 19, which had Z-scores of −1.793 and −1.706 respectively, indicating a strong alignment with default risk profiles. Despite the model's overall high classification accuracy, a few misclassifications occurred. Notably, Case 3, a non-default client, was incorrectly classified as default, with a borderline Z-score of −0.277, which may explain the misjudgment. The proximity of this score to zero indicates an ambiguous risk position, highlighting the challenge in cases that do not clearly align with one group. The value of these findings lies in the model's ability to assist credit officers in making informed lending decisions. When the discriminant score of an applicant is significantly negative, it suggests a strong match with the non-default profile, implying lower credit risk and potentially more favorable loan terms. Conversely, a strongly positive score signals a high risk of default, suggesting that the loan should either be denied or issued with a higher interest rate to compensate for the risk. Overall, the casewise statistics as presented in Table 6 confirm the robustness of the discriminant function. The high rate of accurate classifications and the clarity of separation between group centroids validate the model's effectiveness in distinguishing between good and bad credit applicants. This analysis supports the practical implementation of discriminant scoring in SME lending decisions.

4.8 Classification Results

The final step in evaluating the discriminant model involves assessing its classification accuracy. This is achieved by applying the estimated discriminant function to both the original (analysis) sample and the holdout (cross-validated) sample to determine how well the function predicts group membership. By substituting the values of the predictor variables into the discriminant equation, Z-scores are generated for each case, which are then compared to the group centroids to predict whether a case belongs to the default or non-default group. Table 7 provides a summary of the classification results for both the original and cross-validated samples. In the original analysis, the model demonstrates a strong predictive capability, correctly classifying 95% of the cases. Specifically, all 10 clients from the default group (Group 2) were correctly classified, and 9 out of 10 clients from the non-default group (Group 1) were also accurately identified. This high level of accuracy indicates that the model effectively distinguishes between high-risk and low-risk borrowers based on the selected variables. However, when cross-validation is performed—where each case is classified by a function derived from all other cases except the one being classified—the accuracy decreases to 70%. In this validation approach, 5 out of 10 non-default clients and 9 out of 10 default clients were correctly classified. While the drop in classification accuracy during cross-validation is expected due to the more rigorous evaluation method, a 70% accuracy still reflects a reasonably good model, especially considering the limited sample size. These findings suggest that the model has practical utility in credit risk assessment, particularly in identifying default-prone clients. The disparity between the original and cross-validated results also highlights the importance of further testing with larger datasets to improve generalizability and reduce overfitting. Overall, the classification results, as shown in Table 7, validate the discriminant model's potential as a decision-support tool for SME loan assessments. The ability to correctly classify a majority of applicants provides a foundation for implementing risk-based lending strategies and enhancing the efficiency of credit evaluation processes.

Table 7: Functions at Group Centroids

			Predicted Group Membership		
		Group	1	2	Total
Original	Count	1	9	1	10
		2	0	10	10
	%	1	90.0	10.0	100.0
		2	.0	100.0	100.0
Cross-validated	Count	1	5	5	10
		2	1	9	10
	%	1	50.0	50.0	100.0
		2	10.0	90.0	100.0

a. 95.0% of original grouped cases correctly classified.
b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.
c. 70.0% of cross-validated grouped cases correctly classified.

4.9 Histogram of Z Values of Status

The Z scores calculated for each observation in the analysis sample are displayed in Figure 1 as histograms for the non-default (Group 1) and default (Group 2) clients. The histogram for Group 1 shows most Z values are well above zero, indicating accurate classification of non-defaulters. Similarly, the histogram for Group 2 reveals mostly negative Z scores, aligning with the expected profile of defaulters. This clear separation in Z score distributions demonstrates the model's strong ability to distinguish between the two groups.

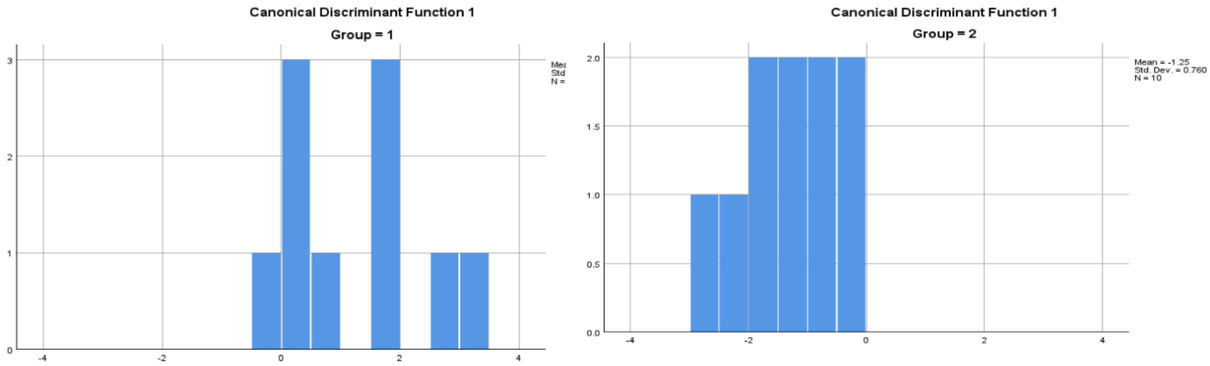

Figure 1: Histogram of Z Values of Status (Histogram of Z values of Group-1 & Group 2)

5. Discussion

This study applied discriminant analysis to classify Small and Medium Enterprises (SMEs) into default and non-default groups using financial and socio-economic predictors. The results demonstrate that certain variables significantly contribute to distinguishing between the two groups, which has important implications for credit risk assessment. The group statistics (Table 1) showed marked differences between defaulting and non-defaulting SMEs in terms of loan amount, number of employees, and loan position against portfolio (LPAP). Specifically, non-default SMEs tended to have higher loan amounts, more employees, and greater LPAP values, suggesting that larger-scale operations and better portfolio positioning are associated with lower default risk. Conversely, variables such as education level and owner's experience exhibited minimal differences, indicating limited usefulness in predicting default status in this context.

This interpretation was reinforced by tests of equality of group means (Table 2), where loan amount and LPAP showed statistically significant differences, highlighting their strong discriminative power. The non-significant results for education, net worth, and owner experience imply that these factors may not be reliable predictors of credit risk within this SME sample. The structure matrix (Table 3) further confirmed that LPAP, loan amount, and number of employees hold the greatest discriminating power among the studied variables. The discriminant function coefficients (Table 4) provide a practical tool to classify applicants. Positive Z scores indicate higher default risk, while negative values suggest safer credit profiles. This function enables lenders not only to classify but also to rank applicants by risk level, allowing risk-based pricing and targeted credit decisions. The clear separation between group centroids (Table 5) with values of 1.385 for non-defaulters and −1.246 for defaulters further validates the model's ability to differentiate between groups effectively. Casewise statistics (Table 6) illustrate that the model performs well at the individual level, accurately predicting most applicants' default status. The few misclassifications—especially cases with borderline Z scores—highlight the inherent challenges in credit risk modeling, where some applicants naturally fall near decision boundaries. Nonetheless, the overall accuracy provides confidence in the model's practical utility for SME credit evaluation.

Classification results (Table 7) confirm high predictive accuracy (95%) for the original sample, although accuracy drops to 70% in cross-validation, reflecting typical performance declines under stricter validation. Despite this decrease, the model's ability to correctly identify a majority of default cases remains valuable for risk management, particularly in emerging markets with limited data availability. Finally, the histograms of Z values (Figure 1) visually illustrate the model's discriminative power, showing a distinct separation of scores between default and non-default groups. This reinforces the robustness of the discriminant function and its effectiveness as a decision-support tool. Overall, this discriminant analysis model offers a statistically sound and interpretable framework for SME credit risk assessment in Bangladesh. By focusing on key financial variables, lenders can better identify high-risk applicants, optimize loan approval processes, and implement risk-adjusted pricing strategies. Future research should consider larger and more diverse samples, incorporate additional variables such as cash flow and market conditions, and explore nonlinear or machine learning methods to further enhance predictive accuracy.

6. Conclusion

This study demonstrated the effectiveness of discriminant analysis in classifying Small and Medium Enterprises (SMEs) into default and non-default groups based on financial and socio-economic variables. The analysis identified loan amount, loan position against portfolio (LPAP), and number of employees as significant predictors of credit risk, while education, net worth, and owner's experience had less influence. The discriminant function achieved high classification accuracy, particularly in the original sample, indicating its practical utility for credit risk assessment in the SME sector in Bangladesh. These results suggest that the model can serve as a valuable tool for lenders by enabling more informed, data-driven decisions that reduce credit risk and improve loan portfolio quality.

7. Applications

The discriminant analysis model developed in this study offers practical applications for financial institutions and credit managers involved in SME lending. By utilizing the discriminant function, lenders can efficiently classify loan applicants into risk categories, enabling more objective and data-driven credit decisions. This can streamline the loan approval process, reduce subjective biases, and enhance risk-based pricing strategies by adjusting interest rates according to the borrower's risk level. Additionally, the model can support ongoing portfolio management by identifying potentially risky borrowers early, allowing institutions to take preventive measures such as increased monitoring or loan restructuring. Furthermore, the insights from the model can help policymakers and regulatory bodies understand risk factors prevalent among SMEs, guiding the design of targeted credit support programs to promote sustainable growth in the SME sector.

8. Limitations and Future Research Directions

One limitation of this study is the relatively small sample size, which may affect the generalizability of the results. To address this, future research should involve larger, more diverse datasets encompassing SMEs from various industries and geographic regions. Expanding the sample size will improve the robustness of the model and its applicability across different contexts. Another limitation lies in the selection of predictor variables, as some potentially important financial and behavioral factors were not included. Future studies should incorporate additional variables such as cash flow metrics, credit history, profitability ratios, and macroeconomic indicators to build a more comprehensive risk assessment model. This could enhance the accuracy and explanatory power of the discriminant function. The current model assumes linear relationships between predictors and the outcome, which might oversimplify the complex nature of credit risk. Future research could explore nonlinear and interaction effects by applying advanced techniques such as machine learning algorithms (e.g., random forests, support vector machines, or neural networks). These methods may better capture intricate patterns in the data and improve classification performance. Finally, the model's predictive accuracy declined during cross-validation, indicating possible overfitting to the sample data. Future studies should emphasize model validation through repeated cross-validation, bootstrapping, or testing on external datasets to ensure the model's stability and reliability in practical applications. Developing dynamic models that update risk predictions over time based on changing borrower conditions could also increase the model's relevance and usefulness in real-world credit risk management.

Author contributions

All authors equally involved in this research.

Financial support

It is declared that we don't receive any monetary help in any form from any individual body or organization.

Conflict of interest

No conflict of interest.

References

Ahmed, M. U. (1999). *The small and medium enterprises (SME) development in Bangladesh: An overview*. Asian Development Bank.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23*(4), 589–609. https://doi.org/10.2307/2978933

Ayyagari, M., Beck, T., & Demirgüç-Kunt, A. (2007). Small and medium enterprises across the globe. *Small Business Economics, 29*(4), 415–434. https://doi.org/10.1007/s11187-006-9002-5

Beck, T., Demirgüç-Kunt, A., & Levine, R. (2005). SMEs, growth, and poverty: Cross-country evidence. *Journal of Economic Growth, 10*(3), 199–229. https://doi.org/10.1007/s10887-005-3533-5

Berger, A. N., & Udell, G. F. (2006). A more complete conceptual framework for SME finance. *Journal of Banking & Finance, 30*(11), 2945–2966. https://doi.org/10.1016/j.jbankfin.2006.05.008

Chowdhury, M. S., & Alam, Z. (2017). Constraints of SMEs in Bangladesh: Evidence from Khulna City. *International Journal of Business and Economic Sciences Applied Research, 10*(1), 19–25. https://doi.org/10.25103/ijbesar.101.03

Dinh, T. H. T., & Kleimeier, S. (2007). A Credit Scoring Model for Vietnam's Retail Banking Market. *International Review of Financial Analysis*, 16(5), 571-495. http://dx.doi.org/10.1016/j.irfa.2007.06.001

Nasir Uddin, D. (2013). *Consumer Credit Customers' Financial Distress Prediction by Using Two-Group Discriminant Analysis*. URL: http://dx.doi.org/10.5539/ijef.v5n6p55

Glen, J. J. (2001). *Classification Accuracy in Discriminant Analysis: A Mixed Integer Programming Approach*. The Journal of Operational Research Society, 52(3), 328. http://dx.doi.org/10.1057/palgrave.jors.2601085

Grablowsky, B. J. (1975). Discriminant analysis in consumer credit risk evaluation. *Journal of Financial and Quantitative Analysis, 10*(3), 541–553. https://doi.org/10.2307/2329960

Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 160*(3), 523–541. http://dx.doi.org/10.1111/j.1467-985X.1997.00078.x

Thomas, L. C., Crook, J. N., & Edelman, D. B. (2002). *Credit scoring and its applications*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898718317

Wiginton, J. C. (1980). A Note on the *Comparison of Logit and Discriminant Models of Consumer Credit Behavior*. Journal of Financial and Quantitative Analysis, 15(3), 757-770. http://dx.doi.org/10.2307/2330408

Published by

IJSAB International