European Economic Letters
ISSN 2323-5233
Vol 14, Issue 1 (2024)
http://eelet.org.uk

Application of AI in Credit Analysis for SME Financing

Vidya CM
Research Scholar
International School of Management Excellence
Recognised Research Center of the University of Mysore
Email: vidya.isme20@gmail.com

Dr. Geetha Rajaram
Research Supervisor
International School of Management Excellence
Recognised Research Center of the University of Mysore
Email: rajaram.geetha@gmail.com

Abstract

Financial providers typically evaluate business statements of finances, cash flow statements, personal financial statements, and business plans as part of their traditional credit research process for any entity. This paper demonstrates how financial technology, or Fintech, may speed up determining an entity's creditworthiness, especially for SMEs. A main source of difficulty for small and medium-sized businesses seeking financing to launch or expand their operations is a deficiency of concrete data and a solid track record. The issue of information asymmetry affects SMEs. The SMEs might not be able to supply the same information because of a lack of financials, but the financial providers need additional details to complete the credit proposal. In the present situation, financial technology pitches in the form of big data, in the absence of concrete data. Every second, an enormous amount of data is gathered from multiple sources, social media platforms, online and offline transactions, and other sources. The new era of credit analysis is centered on behaviour among customers. Everything from preferences for purchased products to likes and dislikes expressed on social media platforms is captured through "Big Data." Everything is available in minutes, allowing one to determine whether or not a person deserves credit quickly. All of these are investigated using intricate mathematical calculations to examine information sequences within a vast amount of data. This study investigates how Fintech can shorten the time needed for credit examination.

Keywords: *Big data, MSMEs, Credit Analysis, Credit scoring, Fintech*

I Introduction

1.1. Market Size of SMEs in India

In India, there are about 6.3 crore SME owners. 767,734 SMEs were registered on the Udyam Registration platform as of November 26, 2021, based on information provided by the SME Ministry. According to the statistics, the SMEs contribute 45% of total industrial production and 40% of total exports. The overall input to the GDP is 37.54 %. The service sector backs to 30.05%. The manufacturing sector backs the remaining 7.09%.

1.2. Hardship faced by SMEs to avail Loan.

1.2.1. Inadequate Credit

Even if you have a great idea or a cutting-edge product, starting even a tiny business may be expensive. The majority of SME owners go to lending organisations for assistance in defraying startup costs. However, because this is a first-time loan with repayment risk, not all of these financial institutions can trust small businesses. (Chowdhury, 2011)

1.2.2. Lack of Knowledge of Technology

As most of the SMEs are based in rural or semi-urban areas, the stakeholders lack knowledge about the latest software and hardware that support their business. One of the main obstacles that small and medium-sized businesses in India must overcome is technology. Businesses can become more efficient and systematised by utilising inventory management software, ERP technologies, and artificial intelligence-based equipment and solutions. (Singh, 2016)

1.2.3. Inadequate Marketing

The third hurdle faced by SMEs in leveraging their business is good marketing tactics. SMEs may have Even great pioneering ideas, but to become successful they require a good marketing agency or a professional who would guide them to advertise their products to get greater footfall. Usually, these
services are high, and even if the SMEs can make shortcuts to marketing their products initially, in the long run, survival becomes difficult. (Quaye, 2011)

II Review of Literature

2.1 Market Research and the Ethics of Big Data Daniel Nunan University of Reading MariaLaura Di Domenico University of Surrey, 2013 (Nunan, 2013)

This paper addresses the advantages and moral dilemmas that the rise of big data in market research has highlighted. 82% of internet users are at least 15 years old, according to a survey conducted by academics. Consumers can perform several transactions anywhere they go by using smartphones and mobile internet, creating a digital trail in their wake. They also contend that there has to be an open discourse about the important ethical questions raised by the use of big data by market researchers.

2.2. A Novel Big-data-driven Credit Reporting Framework for SMEs in China. (Sun, Chunlei, Cui, Zeng, & Chang, 2016)

Three aspects of the restricted business coverage, delayed report generation, and unidimensional information approach of traditional SMEs need to be improved, according to this article. They require the following five fundamental modules for evaluation: knowledge interaction method, data gathering, data authentication, credit evaluation, and credit report generation. The authors suggest using both hard and soft data to determine a person's creditworthiness. The study also discusses the drawbacks of using massive data in terms of information supply, government initiatives to collect data and other restrictions. There is a research vacuum to discover more organised ways to combine both hard and soft

2.3 A Review of Credit Scoring Research in the Age of Big Data Ceylan Onay, Elif Ozturk, June 2017 Cited By 44 (Onay, 2018)

The paper examines the literature on credit scoring from 1976 to 2017 and delivers research findings that highlight the threats and prospects that data for credit scoring poses. In this work, it is examined how credit scoring is changing quantitatively, and it is suggested that future research should focus on data-driven reporting techniques, the use of nonfinancial data sources for credit marking, and their governing features. The study's findings illustrate both the advantages and disadvantages of incorporating big data into credit assessment systems.

2.4 On the Rise of FinTechs: Credit Scoring Using Digital Footprints

This study looks into how our internet paths, or "numerical paths," might be used to forecast loan evasions and finds that they are equally as treasured as old-style credit grades. By combining these two methods, nonpayment rates are generally reduced, credit is made available to people without an acclaimed history, and correctness is increased. This has widespread effects on laws in the numerical financial scope, consumer behavior, lenders, and the unbanked.

2.5. The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the Lending Club Consumer Platform Cited by 146 (Jagtiani, 2019) The article examines how financial technology is transforming the banking and financial sectors. The researchers have looked at how financial technology lenders use different data sources and how it affects financial adoption. The authors of this study contrasted loans made by respectable fintech lenders with loans produced by traditional banking sectors. It has been found that there is a strong correlation between a loan's performance and the interest rate differential as assessed by Lending Club.

2.6 The effect of fintech on banks' credit provision to SMEs: Evidence from China
(Sheng, 2021)

This study looks into how financial technology, or fintech, affects banks' ability to lend money to small and medium-sized businesses. The data from 2011 to 2018 from China was examined by the researchers. Based on this data, they conclude that fintech can be used as a tool to facilitate the bank loans that SMEs receive. Additionally considered in the study is how the impact of fintech varies with bank size. When compared to small banks, the study indicates that fintech has a major influence on the expansion of SME lending in large banks.

III Research Gap

Financial institutions employ a credit scoring system that determines the possibility of default risk, or the likelihood that a borrower could default on a loan in the future, to assess the creditworthiness of SMEs. To help with better decision-making, financial institutions use mathematical models known as credit scorecards to quantify the likelihood that a borrower would engage in bad credit behavior, such as loan default, bankruptcy, or delinquency. Financial entities decide whether to offer a loan or not by comparing the borrower's score with the scorecard's cut-off score. To statistically identify the significant predictors of default, the conventional method, known as financial modeling, mainly uses a variety of financial ratios and financial structures based on information obtained from the borrower's financial records.

IV Research Problem

The following research problem statements are identified from the research gap:

1. What financial and non-financial information affects the creditworthiness of SMEs?
2. What are the various Artificial intelligence tools used for the credit analysis process?
3. How can these models be assessed for the credit analysis process?

V Objectives

The following goals are set forward to facilitate finding answers to the aforementioned issues

1. To determine the industries and select SMEs in various industries.
2. Assessing the elements to determine the SME's credit risk score.
3 To assess the Artificial Intelligence tools on the SME credit scoring data.

VI Scope of the research

1. The SMEs are divided into manufacturing and service sectors. 5 sectors from manufacturing and 3 sectors from service.
2. The sectors are classified based on the list of scrips given in the BSE (Bombay Stock Exchange).
3. Out of the classified sectors, 5 SMEs are randomly selected from each sector. The selected SMEs are active enterprises as per the Scrips listed in BSE. In total 40 SMEs are taken for the study.
4. Further, each SMEs are analysed for 5 years from financial year 2018 to 2022. The analysis is based on the Annual Report of SMEs
5. For the AI models the data SME factors were taken and derived in 5000 sets.

VII. HYPOTHESES OF STUDY

The null hypothesis (H0) states that there is no discernible difference between random chance and the AI models' prediction ability for SME credit rating.

The alternative hypothesis (H1) states that the AI models' prediction performance for SME credit rating differs significantly from random chance.

VIII Research Methodology

Two main types of data can be categorised: non-financial data (like firm credit analysis reports) and financial data (like cash flow data).

8.1. Financial data classification for creditworthiness assessment for evaluation of creditworthiness

The Classical Method uses financial data and in particular, the ratios that are calculated as per the following broad category: In the below table Main criteria have been coded as X1, X2….X4, and the Sub-criteria are coded as R1, R2, R3……R14.

Main Criteria	**Sub criteria**	
Liquidity (X1)	Current ratio = Current Assets/ Current Liabilities	R1
	Liquidity ratios = Liquid Assets / Current Liabilities	R2

	Cash Ratio = (cash + marketable securities)/Current liabilities	R3
Solvency (X2)	Debt / Equity (Shareholder's funds)	R4
	Total Assets/ Long term debt	R5
	Net profit before interest and tax/ Interest on long-term funds	R6
Activity (X3)	Cost of Goods Sold/Average Inventory	R7
	Net credit sales/Average trade receivables	R8
	Net credit Purchases/ Average trade payables	R9
	Revenue from operations/ working capital	R10
	Gross profit/Turnover	R11
Profitability (X4)	Earning after Tax/ Turnover	R12
	Earning before interest, tax, and dividend (EBITDA)	R13
	EPS	R14

Figure 1.1 Classification of financial data.

Source: Roy, P. K., & Shaw, K. (2021). A multicriteria credit scoring model for SMEs using hybrid BWM and TOPSIS. *Financial Innovation*, *7*(1), 1-27

8.2. Classification of non-financial data for Evaluation of Creditworthiness

Records of commercial transactions between a business and its clients are referred to as non-financial data. There are three categories for non-financial data: business, payment history, and behavioral qualities. The Main criteria are extended with codes from the table below. Figure 1.2 as NX5, NX6….NX8 and Sub criteria are continued with codes R14, R15, R16……R32.

Main Criteria	**Sub criteria**	
Business Analysis (NX5)	Production as compared to the industry average	R15
	Sales trend – Increasing/decreasing	R16
	Impact of marketing strategies on Turnover	R17
	Bank cashflow activity profiles	R18
	Number of years in business	R19
Management Analysis (NX6)	Type of enterprise	R20
	Education and experience	R21
	Future expansion plans	R22
	Angel investors/external funding	R23
	Financial flexibility	R24
Third-Party Analysis (NX7)	The company's credit analysis reports	R25
	Personal credit reports	R26
	Quality of customers	R27
	Target customer profile	R28
	Quality of transactions	R29
Performance of account (NX8)	Credit History	R30
	Lending relationship	R31
	Risk Characteristics	R32

Figure 1.2 Classification of non-financial data

Source: Roy, P. K., & Shaw, K. (2021). A multicriteria credit scoring model for SMEs using hybrid BWM and TOPSIS. *Financial Innovation*, *7*(1), 1-27

8.3. Source for collection of data

The financial data is collected through secondary sources by accessing the Listed SMEs on the Bombay Stock Exchange, research articles, thesis, Annual reports, and any other published relevant data. Non-financial data is collected through secondary and primary sources. The primary source for the collection of data is the interview method from financial institutions.

8.4. Sampling Method and Analysis

In total there are in total 257 scrips listed in BSE SMEs, these are divided into two broad sectors Manufacturing and Service SMEs. Further, the manufacturing SMEs are sub-classified as Capital goods, Construction, Consumer durables, Fast-moving consumer goods, and textiles. The Service sector SMEs are sub-classified as Consumer service, financial service, and Health care service

Based on stratified random sampling, 5 SMEs are picked up from each sub-sector of Manufacturing and Service (a total of 8 sub-sectors). SMEs and their financials are assessed for 5 years. So total sample size is 200. The Financial and non-financial data for 40 companies are collated for 5 financial years from the Financial year 2017-18 to 2021-22.

8.5. Various Fintech models are used to access the SME credit scoring data.

Artificial Intelligence (AI) has grown in importance when it comes to Small and Medium Enterprises (SMEs) credit rating. (Agarwal, 2019) AI methods, such as deep learning and machine learning, were used to analyse massive amounts of data and spot intricate patterns that conventional credit scoring models might not have seen.

8.5.1. Logistic regression (LR)

For assessing the credit risk of SMEs, logistic regression (LR), a dependable statistical prototype often used for binary procedure problems, was an accommodating tool. This tool helped assess the likelihood that the SME would fail payments on its credit commitments since it provided widespread credit investigation based on a variety of financial and non-financial features.

One of the main merits of using logistic regression was being able to show the relation between various constraints and the possibility of nonpayment. This made it mainly suitable for the interesting task of appraising credit risk since multiple features affected the final valuation. (Bartlett, 2018)

Steps involved in the use of the Logistic regression model

Step 1: Gathering and Preparing Data: Every SME's financial and non-financial data was acquired. The Annual Reports of Listed SMEs included the information for 50 SMEs. Pre-processing was done on the aforementioned data to address outliers, inconsistent values, and missing values. The binary variable that indicates whether a loan is in default (1) or not (0) was selected as the target variable.

Step 2: Estimation and Model Training: There were training and testing sets of the data. The 20: 80 guideline was adhered to, meaning that 20% of the data sets were used for testing and 80% for training. The logistic regression model made use of the training data set. This involved estimating the coefficients (weights) of the features in the logistic regression equation:

$$P(\text{default}) = 1 / (1 + \exp(-(\beta 0 + \beta 1 * X1 + \beta 2 * X2 + ... + \beta n * Xn)))$$

- Where:
  - P(default) is the probability of loan default.
  - $\beta 0$ is the intercept term.
  - βi are the coefficients of the features (R1, R2, ..., Rn).

Step 3 Assessment and Clarification of the Model: Parameters such as correctness, accuracy, recollection, and AUC (Area Under Curve) were used to assess the model's results on the testing data. To regulate the relative implication and course of each factor’s influence on the nonpayment prospect, the features were studied. To regulate the main fundamentals prompting SME solvency, the model's outcomes were analyzed.

Step 4 Model Authentication: The accuracy of the model was analyzed using test data that was not yet observed. The model's outcome was adjusted over time and reviewed in reply to new data and alterations in the market.

8.5.2. Linear Discriminant Analysis (LDA):

Linear discriminant analysis, or LDA, was vital to evaluating credit for SMEs. Certain key features were recognized and highlighted to differentiate financially sound SMEs from others. By using a linear set of constraints, LDA efficiently acknowledged SMEs based on their soundness. This method was very accommodating since it abridged the study and absorbed the constraints that had an extreme effect on credit results, particularly when handling a huge number of financial variables. (Punniyamoorthy, 2016)

Steps Involved in Linear Discriminant Analysis

Step 1: Data Assortment: A collection of preceding SMEs with identified credit results—such as default or not—was collected. Pertinent financial and non-financial variables (R1, R2, R3……Rn) were encompassed.

Step 2 Data Cleaning and Preprocessing: Missing values in the data were dealt with (e.g., by imputation or removal). Variables were normalized or standardized if necessary to ensure comparability.

Step 3. Data Splitting:

The dataset was split into training and testing sets (e.g., 80% for training, and 20% for testing).

Step 4. Mean Vector Calculation: Mean vectors (centroids) were calculated for each class (defaulted and non-defaulted):

μ_1 became the mean of features for defaulted SMEs.

μ_2 became the mean of features for non-defaulted SMEs.

Step 5. Scatter Matrix Computation:

The within-class scatter matrix (SW) was calculated.

The between-class scatter matrix (SB) was calculated.

Step 6. Finding Linear Discriminants: The generalized eigenvalue problem was solved: **(SW^(-1) * SB) * w = λ * w.** The eigenvectors (w) corresponding to the largest eigenvalues (λ) were selected. These became the linear discriminants maximizing class separation.

- SW^(-1): Inverse of matrix S multiplied by matrix W.
- SB: Matrix S multiplied by matrix B.
- w: A vector (non-zero).
- λ: A scalar value (eigenvalue).

Step 7. Data Projection: The data was projected onto the new subspace defined by the linear discriminants. A decision boundary (e.g., linear) was used to create a classification rule based on the projected data.

Step 8. Performance Evaluation: The classification rule was applied to the testing set to assess model accuracy. Metrics like accuracy, precision, recall, F1-score, and ROC curves were used for evaluation. The model was used to assess the creditworthiness of new SMEs. Their features were projected onto the discriminant space and the decision rule was applied to classify them as likely to default or not.

8.5.3. K Nearest Neighbor (KNN):

K-Nearest Neighbors (KNN) made a significant addition to the credit analysis of small and medium-sized enterprises (SMEs). It was used to assess SMEs' creditworthiness based on how similar they were to other similar enterprises in the featured vector. In the past, KNN has proven to be an effective tool for SME credit analysis due to its efficiency and ease of use. (Sadok, 2022)

Steps involved in the KNN model.

Step 1. Data Preparation: A dataset of past SMEs with known credit outcomes, like defaults or no defaults, was collected. Important financial and non-financial features indicating financial health and creditworthiness (R1, R2, R3,….Rn) were chosen. Missing values, outliers, and discrepancies were dealt with, and variables were scaled or normalized for consistency.

Step 2. Training and Prediction: The number of nearest neighbors (K) to consider for predictions was chosen. This crucial parameter was determined through experimentation and cross-validation. For each new SME, distances to every data point in the training set were calculated using metrics like Euclidean or Manhattan distance. The K data points closest to the new SME, based on the chosen distance metric, were identified. Predictions for the new SME's creditworthiness were made based on the K nearest neighbors labels (e.g., defaulted or not).

Step 3. Model Evaluation: Using a testing set of SMEs with known outcomes, the model accuracy, precision, recall, F1-score, and ROC curve were evaluated. This assessed the KNN model's effectiveness for credit analysis. If necessary, the K value and other parameters were adjusted based on the evaluation results to improve the model's performance.

Step 4. Credit Analysis and Scoring: The trained KNN model was applied to new SMEs seeking credit. Their feature values were used to predict their creditworthiness (e.g., credit score) using the KNN approach. Informed credit decisions, such as approving or denying loan applications, were made based on the predicted credit score and other relevant information.

IX Data Interpretation and Analysis

The data of 40 listed SMEs for 5 years was augmented through the following methods to increase the sample size to 5000 sets. Out of these 5000 sets, 1000 sets were considered training sets and the remaining as test data sets. The following are the methods used for augmentation.

9.1. Generating Synthetic Data:

GANs: The Generative adversarial networks GANs have realistic synthetic data points that mimic actual loan applicants by learning the underlying distribution of the current financial data. **VAEs, or variational autoencoders**: The VAEs as a quick substitute for GANs, highlighting their advantages over GANs in terms of interpretability and control over the properties of the generated data. **GAN-tabular, or Generative Adversarial Nets for Tabular Data**: Present GAN-tabular, a specific architecture created to manage tabular data that is often used in credit research

9.2. Feature Engineering:

Extract new features from existing data, like financial ratios, risk scores, or delinquency indicators. This can enrich your dataset and provide the model with more information to learn from.

Financial Ratios: Describe how the model can acquire meaningful features by extracting important ratios such as the debt-to-income, current, and loan-to-value ratios.

Deficiency Markers: Talk about how crucial it is to capture risk indicators by including past due dates and other derived aspects, such as the amount of missed payments or days past due. **Other Sources of Data**: Examine briefly the possibility of combining data from other sources, such as public records, social media posts, or transaction data, keeping in mind privacy and legal requirements.

9.3. Data Smoothing:

Techniques like SMOTE can oversample the minority class (defaulters) in an imbalanced dataset, reducing bias and ensuring the model learns from both positive and negative cases.

Techniques for Smoothing Data:

SMOTE: Describe how SMOTE balances the dataset and lessens bias by creating new data points based on existing ones, hence oversampling the minority class (defaulters).
ADASYN: Present Adaptive Synthetic Sampling, a more sophisticated method that produces a more nuanced distribution by taking the density of data points into account while oversampling.

9.4. Results and Analysis

The following results are obtained by applying the ML models:

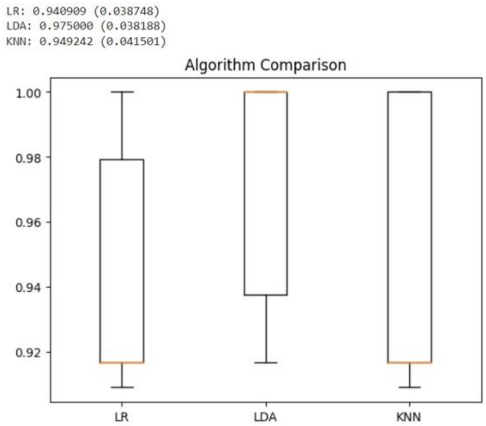

Figure 1.3 Results of ML models run on Collab.
Primary source

The above result shows an algorithm comparison for an SME dataset, where 4000 data points were used for testing. Three AI models were used: Logistic Regression (LR), Linear Discriminant Analysis (LDA), and K-nearest neighbors (KNN).

The average score of the three models is 0.99. The individual scores are:

- LR: 0.940909 (with a standard deviation of 0.038748)
- LDA: 0.975000 (with a standard deviation of 0.038188)
- KNN: 0.949242 (with a standard deviation of 0.041501)

Based on the scores, LDA appears to be the best-performing model, followed by KNN and then LR. However, it is important to consider the standard deviations as well. The standard deviation is a measure of variability, so a lower standard deviation indicates that the model's performance is more consistent. In this case, LDA has the lowest standard deviation, which suggests that it is the most reliable model.

Overall, all three models achieved a high accuracy score, suggesting that they are all good candidates for this task. However, LDA appears to be the best accomplishment model based on its average score and low standard deviation.

9.5. Results of each model giving the results of default and non-default percentages.

Logistic regression model results

Where Positive (POS) is non-default (ND) and negative (NEG) is the default (D) and ND is nondefault

Method	Logistic Regression
Test Data	4000
Correct Prediction	3764
Wrong prediction	236
ACTUAL / Computed	ACTUAL / Computed
NEG/POS(D/ND) : False Positive	POS/POS (ND/ND)
36	3178
NEG/NEG(D/D)	POS/NEG(ND/D) : False Negatives
586	200
NP	PP
0.9%	79.5%
NN	PN
14.7%	5.0%

Table 1.1 Analysis of result run on Logistic regression model

The table shows the results of a logistic regression model that is used to predict whether a customer will default on a loan. The model is applied to a test dataset of 4000 SMEs. The "ACTUAL/Computed" section of the table shows the number of customers who were correctly classified and the number of customers who were misclassified. For example, the "NEG/POS (D/ND)" cell shows that there were 3178 customers who did not default (NEG) but were predicted to default (POS) by the model. These are false negatives. The "NP" (Negative positive) and "PP" (Positive, positive) values show the percentage of customers who were correctly classified as not having defaulted (NP) and having defaulted (PP), respectively. In this case, 90.9% of SMEs were correctly classified as not having defaulted, and 79.5% of SMEs were correctly classified as having defaulted. The "NN" Negative, negative, and "PN" Positive negative values show the percentage of customers who were misclassified. In this case, 9.1% of SMEs who did not default were misclassified as having defaulted, and 5.0% of SMEs who did default were misclassified as not having defaulted. Overall, the model is doing a good job of predicting whether SMEs will default on a loan.

Linear discriminant Analysis model results
Where Positive (POS) is non-default (ND) and negative (NEG) is the default (D) and ND is nondefault

Method	LDA
Test Data	4000
Correct Prediction	3900
Wrong prediction	100
ACTUAL / Computed	ACTUAL / Computed
NEG/POS(D/ND) : False Positive	POS/POS (ND/ND)
16	3294
NEG/NEG(D/D)	POS/NEG(ND/D) : False Negatives
606	84
NP	PP
0.4%	82.4%
NN	PN
15.2%	2.1%

Table 1.2 Analysis of result run on the LDA model

NEG/POS (D/ND): This row shows the number of false positives, which are customers who were predicted to default (POS) but did not default (NEG). In this case, there were 16 false positives for the D group and 3294 for the ND group. NEG/NEG (D/D) and POS/NEG (ND/D): These cells show the number of correctly classified customers for each group. For example, 606 customers in the D group and 84 in the ND group were correctly classified as not having defaulted (NEG/NEG).NP and PP: These values represent the percentage of customers who were correctly classified. NP is 90.9% for non-defaulters and PP is 79.5% for defaulters. NN and PN: These values represent the percentage of customers who were misclassified. NN is 9.1% for non-defaulters who were misclassified as defaulters and PN is 5.0% for defaulters who were misclassified as non-defaulters.

K Nearest Neighbor results

Method	KNN
Test Data	4000
Correct Prediction	3797
Wrong prediction	203
ACTUAL / Computed	ACTUAL / Computed
NEG/POS(D/ND) : False Positive	POS/POS (ND/ND)
32	3207
NEG/NEG(D/D)	POS/NEG(ND/D) : False Negatives
590	171
NP	PP
0.8%	80.2%
NN	PN
14.8%	4.3%

Table 1.3 Analysis of result run on the KNN model

NEG/POS (D/ND): This row shows the number of false positives, which are customers who were predicted to default (POS) but did not default (NEG). In this case, there were 16 false positives for the D group and 3294 for the ND group. NEG/NEG (D/D) and POS/NEG (ND/D): These cells show the number of correctly classified customers for each group. For example, 606 customers in the D group and 84 in the ND group were correctly classified as not having defaulted (NEG/NEG). NP and PP: These values represent the percentage of customers who were correctly classified. NP is 90.9% for non-

defaulters and PP is 79.5% for defaulters. NN and PN: These values represent the percentage of customers who were misclassified. NN is 9.1% for non-defaulters who were misclassified as defaulters and PN is 5.0% for defaulters who were misclassified as non-defaulters.

X Conclusion

Financial institutions can handle massive datasets more effectively thanks to FinTech's streamlining of credit analysis processes. By eliminating manual labour, automated procedures increase operational effectiveness and speed up decision-making. Fintech's advanced algorithms improve risk assessment by examining various data sources. Financial institutions can reduce possible credit risks and make well-informed judgments with the aid of real-time risk monitoring Adopting Fintech lowers costs by reducing overhead and expediting the credit assessment process. Financial firms are better able to manage budgets and maximise resources.

Credit ratings and financial statements, among other static historical data, are major sources of reliance for traditional credit analysis. FinTech provides access to a wide range of dynamic, real-time data. Cash flow, payments, and purchases all

reveal information about an SME's operations and financial standing. E-commerce activity, supply chain data, and social media involvement can all help to provide a more complete picture of the SMEs. For SMEs with little credit history, using machine learning to leverage alternative data can produce more accurate creditworthiness ratings.
Increasing SMEs' access to financing can stimulate innovation, entrepreneurship, and the creation of new jobs, all of which support economic expansion. By promoting financial inclusion for underbanked or underserved groups, less reliance on traditional credit scoring can encourage more people to engage in the formal economy By bridging economic gaps, customized finance options for SMEs in neglected areas can support regional development. **:** Both the creation of jobs and the decrease in poverty can be indirectly attributed to SMEs, who employ a large number of people and have more access to financing. Giving SMEs and communities the chance to build their businesses and become financially included can help them rise to the economic ladder.

References

1. Gavalas, D., & Syriopoulos, T. (2014). Bank Credit Risk Management and Rating Migration Analysis. *International Journal of Financial Studies*.
2. Hurley, M., & Adebayo, J. (2016). CREDIT SCORING IN THE ERA OF BIG. *Yale Journal of Law and Technology,* Pages From - To.
3. Yi Man, S. (2016). *Applying Big Data to accurately determine.*
4. Agarwal, S. A. (2019). Financial inclusion and alternate credit scoring: Role of big data and machine learning in Fintech. Indian School of Business.
5. Bartlett, R. M. (2018). Consumer-lending discrimination in the era of fintech. *University of California, Berkeley.*
6. Chowdhury, T. A. (2011). An appraisal of the problems and prospects of small and medium enterprises (SMEs) financing in Bangladesh: A study on selected districts. *Center for Research and Training East West University*.
7. E. Khandani, A., A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*.
8. *Estuate*. (2017). Retrieved from www.estuate.com.
9. Godbillon-Camus, B., & Christophe Godlewski. (January 2006). Credit Risk Management in Banks: Hard Information, Soft Information and Manipulation. *SSRN Electronic Journal*.
10. https://ibsintelligence.com/ibsi-news/5-top-fintech-lending-companies-for-msmes-in-india/. (n.d.).
11. https://www.mbarendezvous.com/more/msme-indian-economy/. (n.d.).
12. https://www.smeraonline.com/newweb/index.php/about/. (n.d.).
13. https://www.tatacapital.com/blog/business-loan/top-challenges-faced-by-sme-msme-financing-in-india/. (n.d.).
14. Last Name, F. M. (Year). *Book Title.* City Name: Publisher Name.
15. *newgenapps*. (2017). Retrieved from www.newgenapps.com.
16. Punniyamoorthy, M. &. (2016). Identification of a standard AI-based technique for credit risk analysis. *Benchmarking: An International Journal,* 23(5), 1381-1390.
17. Quaye, D. N. (2011). The effect of microfinance institutions on the growth of small and medium scale enterprises (SMEs): A case study of selected SMEs in the Kumasi metropolis. *Doctoral dissertation)*.
18. Sadok, H. S. (2022). Artificial intelligence and bank credit analysis: A review. *Cogent Economics & Finance*, , 10(1), 2023262.
19. Schweitzer, M. &. (2017). Is' Fintech Good for Small Business Borrowers? Impacts on Firm Growth and Customer Satisfaction.
20. Singh, C. &. (2016). Finance for micro, small, and medium-sized enterprises in India: Sources and challenges.
21. Sun, Y., C. L., Cui, X., Zeng, X., & Chang, X. (2016). A Novel Big-data-driven Credit Reporting Framework for SMEs in China. *2016 International Conference on Identification, Information, and Knowledge in the Internet of Things.* Bejing: Research Gate publication.
22. *zestfinance*. (2009). Retrieved from https://www.zestfinance.com/zaml.
23. Gavalas, D., & Syriopoulos, T. (2014). Bank Credit Risk Management and Rating Migration Analysis. *International Journal of Financial Studies*.Hurley, M., & Adebayo, J. (2016). CREDIT SCORING IN THE ERA OF BIG. *Yale Journal of Law and Technology*, Pages From - To.Yi Man, S. (2016). *Applying Big Data to accurately determine.*

24. x`xE. Khandani, A., A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*.
25. *Estate*. (2017). Retrieved from www.estuate.com.
26. Godbillon-Camus, B., & Christophe Godlewski. (January 2006). Credit Risk Management in Banks: Hard Information, Soft Information and Manipulation. *SSRN` Electronic Journal*.
27. https://ibsintelligence.com/ibsi-news/5-top-fintech-lending-companies-for-msmes-in-india/. (n.d.).
28. https://www.mbarendezvous.com/more/msme-indian-economy/. (n.d.).
29. https://www.smeraonline.com/newweb/index.php/about/. (n.d.).
30. https://www.tatacapital.com/blog/business-loan/top-challenges-faced-by-sme-msme-financing-in-india/. (n.d.).
31. Last Name, F. M. (Year). *Book Title.* City Name: Publisher Name.
32. *newgenapps*. (2017). Retrieved from www.newgenapps.com.
33. Sun, Y., C. L., Cui, X., Zeng, X., & Chang, X. (2016). A Novel Big-data-driven Credit Reporting Framework for SMEs in China. *2016 International Conference on Identification, Information, and Knowledge in the Internet of Things.* Bejing: Research Gate publication.
34. *zestfinance*. (2009). Retrieved from https://www.zestfinance.com/zaml.