Data Science and Engineering (DSE) Record, Volume 6, issue 1.

Estimation of the Credit Rating of the Listed Companies in The Stock Exchange of Thailand Based on Financial Statement

Natt Ungcharoen [1] and Karn Patanukhom [2]

[1] Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand
[2] Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand
natt_u@cmu.ac.th

Abstract. Nowaday, there are over 921 listed companies on the Stock Exchange of Thailand, with a total market capitalization of 17,430,644.71 billion THB as of the end of 2023. These listed companies can issue bonds (debt securities) for public sale, providing Thai investors with diverse financial investment options. In 2023, more than 4,753,851 billion THB was raised through initial bond offerings. Despite stringent oversight by the Securities and Exchange Commission of Thailand (SEC), some companies have faced financial failures, leading to delisting and defaults on bond payments, which have significantly harmed numerous investors. Most companies that defaulted on bond payments lacked credit ratings from credit rating agencies, which are crucial for investors to assess the risk of financial failure.
As of August 2024, only 175 listed companies on the Stock Exchange of Thailand had received credit ratings from Tris Rating Co., Ltd. This highlights the importance of analyzing and estimating credit ratings for listed companies based on their financial statements to support Thai investors in evaluating financial investments. The findings of this research aim to provide a valuable tool for investors in analyzing investments in financial instruments issued by listed companies. The result of study show in a tabular format including machine learning model performance and training parameters.

Keywords: Credit Rating, Financial Statement, Machine Learning.

1 Introduction

The Stock Exchange of Thailand (SET) currently lists over 921 companies. By the end of 2023, the total market capitalization of the SET was 17,430,644.71 million THB. The trend for listed companies shows significant growth, partly because they can raise substantial capital from the public and access financing options beyond those available to unlisted companies, such as issuing bonds (debt securities) for public sale. In 2023, the primary market saw over 4,753,851 million THB raised through bond issuance, increasing investment options for Thai investors.

However, despite strict regulation by the SEC, some listed companies have experienced financial failure, resulting in delisting from the stock exchange and significant harm to retail investors. From 1975 to April 2024, 298 listed companies were delisted, and 20 were at risk of delisting. In 2023, six companies defaulted on bond payments, totaling 16,363 million THB. Five out of these six defaulting companies were Non-rated (lacked a credit rating). Financial failure can lead to a company's inability to meet obligations, potentially resulting in bankruptcy and causing damage to owners, investors, and stakeholders, possibly impacting the overall economy.

Credit rating indicates a company's ability to repay debt and reflects risk. It can change based on performance, industry trends, and economic conditions. Companies wishing to raise funds through long-term debt instruments (bonds) offered to the public generally require a credit rating to be able to sell them. Companies that choose not to obtain a credit rating face greater restrictions on offering long-term debt instruments. Many listed companies that do not issue debt instruments for public sale do not need a credit rating. Credit rating agencies assess an entity's ability to repay debt, indicating the level of default risk; a high rating implies low risk. Ratings can be for the entire organization (Company Rating) or specific debt instruments (Issue Rating). Assessment considers organizational structure, financial information, and business plans to evaluate financial status and earning potential. In Thailand, two SEC-certified credit rating agencies are TRIS Rating and Fitch Rating Thailand.

This independent study aims to create a Machine Learning model capable of estimating credit ratings for companies listed on the Stock Exchange of Thailand using financial statement data. It also seeks to identify variables that influence company credit ratings. The scope of the research includes credit rating data from TRIS Rating for companies listed on SET and mai between April 30, 1975, and August 31, 2024, qualitative data (market and industry group), and quarterly financial statement data for rated companies from April 30, 1975, to June 30, 2024. The study involves building the model and process, and comparing the model's estimated ratings with actual TRIS ratings.

2 Literature Review

2.1 Deep learning approaches provide effective methods for corporate credit rating predictions

Napasorn Thavichaigarn [1] studied corporate credit rating prediction using data from TRIS Rating and 17 financial ratios. The research compared three machine learning approaches: Support Vector Machine (SVM), Linear Regression, and Deep Neural Network (DNN). The 17 financial variables included stock price change rate, return on equity, total assets, enterprise value, operating cash flow, investment cash flow, financing cash flow, net cash flow, earnings per share, net profit, fixed asset turnover ratio, total asset turnover ratio, net profit margin, current ratio, quick ratio, cash cycle, and debt service coverage ratio. The study found that the Deep Neural Network method provided results comparable in effectiveness to other machine learning approaches for predicting corporate credit ratings.

2.2 Feature selection significantly improves accuracy in credit rating prediction models

Zeyu Huang, Savio Pereira, and Meiqi Liu [2] investigated corporate credit rating prediction using machine learning techniques with a focus on feature selection methods. The study compared five models and two feature selection approaches: The Permutation Feature Selection and The Null Important Feature Selection. The research focused on aviation and energy industries to eliminate sector-specific risk factors, using 48 financial ratios categorized into three main risk groups: Business Risk (including profit margin, return, and efficiency ratios), Financial Risk (including leverage, coverage, and debt profile ratios), and Other Risks (including liquidity and valuation ratios). The findings revealed that proper feature selection significantly improved prediction accuracy, with financial structure (leverage) ratios emerging as the primary predictive factors, while business risk ratios showed minimal impact on credit rating predictions.

2.3 XGBoost demonstrates superior performance in corporate credit rating prediction models

Pamuk, Mustafa, and Matthias Schumann [3] conducted research comparing multiple machine learning methods for credit rating prediction based on annual financial statements. The study evaluated four approaches: Neural Network, XGBoost, Logistic Regression, and Decision Tree. The research utilized seven financial variables: Equity Ratio, Short-term Debt Ratio, Current Ratio/Working Capital Ratio, Return on Assets (ROA), Return on Equity (ROE), Asset Coverage Ratio, and 2nd Degree Liquidity. The findings revealed that

XGBoost outperformed the other methods, achieving prediction accuracy levels between 75% and 89% for corporate credit rating forecasts.

2.4 Altman's EM-Score Model provides a robust framework for credit rating prediction in emerging markets

Edward I. Altman [4] developed the Emerging Market Credit Scoring System for Corporate Bonds (Altman's EM-Score Model) to evaluate and predict credit ratings for non-listed companies and firms in emerging markets. The model applies a weighted equation with four key financial variables: Z-Score = $3.25 + 6.56X_1 + 3.26X_2 + 6.72X_3 + 1.05X_4$, where X_1 represents working capital to total assets ratio, X_2 is retained earnings to total assets ratio, X_3 is EBIT to total assets ratio, and X_4 is book value of equity to total liabilities ratio. The constant 3.25 is derived from the median Z-Score of bankrupt US companies, establishing a baseline for the lowest credit rating (D). The model classifies Z-Score ranges into specific credit ratings from AAA to D, categorized into three risk zones: Safe Zone (BBB and above), Gray Zone (BB+ to B), and Distress Zone (B- and below). Further validation studies using data from Mexico, Argentina, Italy, China, Singapore, and Malaysia confirmed the model's accuracy and appropriateness for emerging market applications.

3 Data and Methodology

3.1 Data

The study used data from listed companies on the SET and mai markets that received credit ratings from TRIS Rating Co., Ltd. The data collection period for credit ratings was from January 1, 2019, to May 8, 2024. For financial statements, quarterly data from January 1, 2019, to March 31, 2024, were collected. Data sources included the Stock Exchange of Thailand website (set.or.th) for listed company lists, the TRIS Rating website (trisrating.com) for credit ratings, and the SETSMART website (setsmart.com) for quarterly financial statements. The financial statement data were further processed into financial ratios. Initial data collection identified 105 companies with TRIS ratings (103 on SET, 2 on mai), totaling 2765 records. The initial ratings were distributed across 13 distinct levels & 4 group levels, following the classification framework from Professor Dr. Anya Khanthavit's [5] "ACE Portfolio Investment" academic framework as show in table 1. The distribution for each distinct levels and 4 group levels are visualize in Figure 1 & 2 respectively.

Table 1. comparing grades from the academic lecture on “ACE Portfolio Investment” by Professor Dr. Anya Khanthawit with the ratings by TRIS Rating.

Category	Grade	TRIS Rating
Investment Grade	1	AAA
		AA+
		AA
		AA-
Investment Grade	2	A+
		A
		A-
Investment Grade	3	BBB+
		BBB
		BBB-
Non-investment Grade	4	-
Non-investment Grade	5	BB+
		BB
		BB-
Non-investment Grade	6	B+
		B
		B-
		CCC+
		CCC
		CCC-
		CC
		C
		D

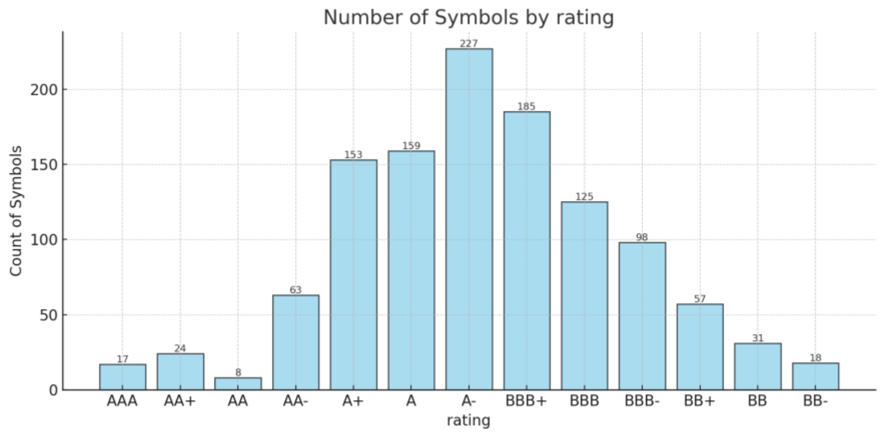

Fig 1. The distribution of 13 distinct credit rating level

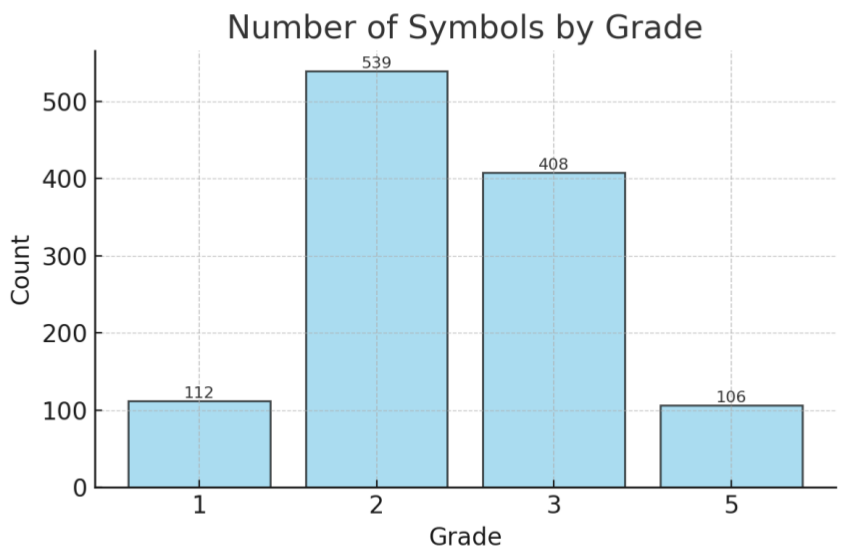

Fig 2. The distribution of 4 group level

1) Explained variable: Y: Ordinal Credit rating level from Tris Rating Co., Ltd. Of both 13 levels & 4 groups

2) Explanatory variable: Calculated financial ratio as showed in Table 2.

Table 2. Explanatory variable from fiancial ratio calculation

Vairable	Financial Ratio	Calculation
X1	Working Capital to Total Assets Ratio	(Current Assets - Current Liabilities) / Total Assets
X2	Retained Earnings to Total Assets Ratio	Retained Earnings / Total Assets
X3	EBIT to Total Assets Ratio	Earnings Before Interest and Taxes (EBIT) / Total Assets
X4	Book Value of Equity to Total Liabilities Ratio	Book Value of Equity / Total Liabilities
X5	Current Ratio	Current Assets / Current Liabilities
X6	Quick Ratio	(Current Assets - Inventories) / Current Liabilities
X7	Inventory Turnover Ratio	Cost of Goods Sold / Average Inventory
X8	Inventory Period	365 / Inventory Turnover
X9	Accounts Receivable Turnover	Revenue / Average Accounts Receivable
X10	Collection Period	365 / Accounts Receivable Turnover
X11	Accounts Payable Turnover	Cost of Goods Sold / Average Accounts Payable
X12	Payment Period	365 / Accounts Payable Turnover
X13	Cash Conversion Cycle	Inventory Period + Collection Period - Payment Period
X14	Gross Profit Margin	(Gross Profit / Revenue) x 100
X15	Operating Profit Margin	(Operating Profit / Revenue) x 100
X16	Net Profit Margin	(Net Profit / Revenue) × 100
X17	Earnings Per Share (EPS)	Net Profit / Number of Shares
X18	Return on Assets (ROA)	(Net Profit / Total Assets) × 100
X19	Return on Equity (ROE)	(Net Profit / Shareholders' Equity) × 100
X20	Total Asset Turnover Ratio	Revenue / Total Assets
X21	Fixed Asset Turnover Ratio	Revenue / Fixed Assets
X22	Debt to Equity Ratio	Total Liabilities / Shareholders' Equity
X23	Debt to Total Assets Ratio	Total Liabilities / Total Assets
X24	Interest Coverage Ratio	EBIT / Interest Expense
X25	Debt Service Coverage Ratio	(EBITDA) / (Current Assets - Trade and Other Payables)
X26	Cash Flow from Operating Activities	Data from Financial Statements (Not Calculated)
X27	Cash Flow from Investing Activities	Data from Financial Statements (Not Calculated)
X28	Cash Flow from Financing Activities	Data from Financial Statements (Not Calculated)
X29	Net Cash Flow	Data from Financial Statements (Not Calculated)
X30	Enterprise Value	Data from Financial Statements (Not Calculated)
X31	Market Capitalization	Data from Financial Statements (Not Calculated)
X32	Equity to Total Assets Ratio	Shareholders' Equity / Total Assets

X33	Current Liabilities to Total Assets Ratio	Current Liabilities / Total Assets
X34	Ratio of assets net of current liabilities to total liabilities	(Total Assets - Current Liabilities) / Total Liabilities

3.2 Methodology

This research aims to develop a machine learning model for predicting credit ratings of companies listed on the Stock Exchange of Thailand using financial statement data. The methodology employs a structured approach to identify key financial variables influencing credit ratings. The study utilizes two primary data sources: financial statements of SET-listed companies and credit ratings from TRIS Rating Limited spanning from January 1, 2019, to May 8, 2024. The research process includes systematic sample selection with specific inclusion criteria (companies listed on SET and mai markets) and exclusion criteria (companies without TRIS ratings during the study period). The implementation follows a comprehensive ten-step procedure: qualifying company selection, data cleaning, credit rating collection, qualitative data gathering (including market registration and industry classification), quarterly financial data collection spanning January 2019 to March 2024, algorithm development for credit rating prediction, machine learning model construction, model accuracy verification, result analysis, and documentation. This methodical approach ensures robust model development while identifying the most significant financial variables affecting corporate credit ratings.

1) Feature Selection and Correlation Analysis.

One of the objectives of this study is to identify the variables that influence the credit ratings of companies as published by TRIS Rating. This objective aligns with a key data science methodology known as feature selection, which aims to identify relevant variables that affect outcomes while excluding irrelevant ones. This process enhances the model's predictive accuracy by reducing noise and preventing overfitting.

In this research, the selection of influential variables was carried out using **Pearson's Correlation Coefficient**. The correlation coefficients between each financial ratio and the credit rating (ordinally encoded as ordinal_rating) are illustrated in the heatmap shown in Figure 3. Variables X1 to X34 represent financial ratio.

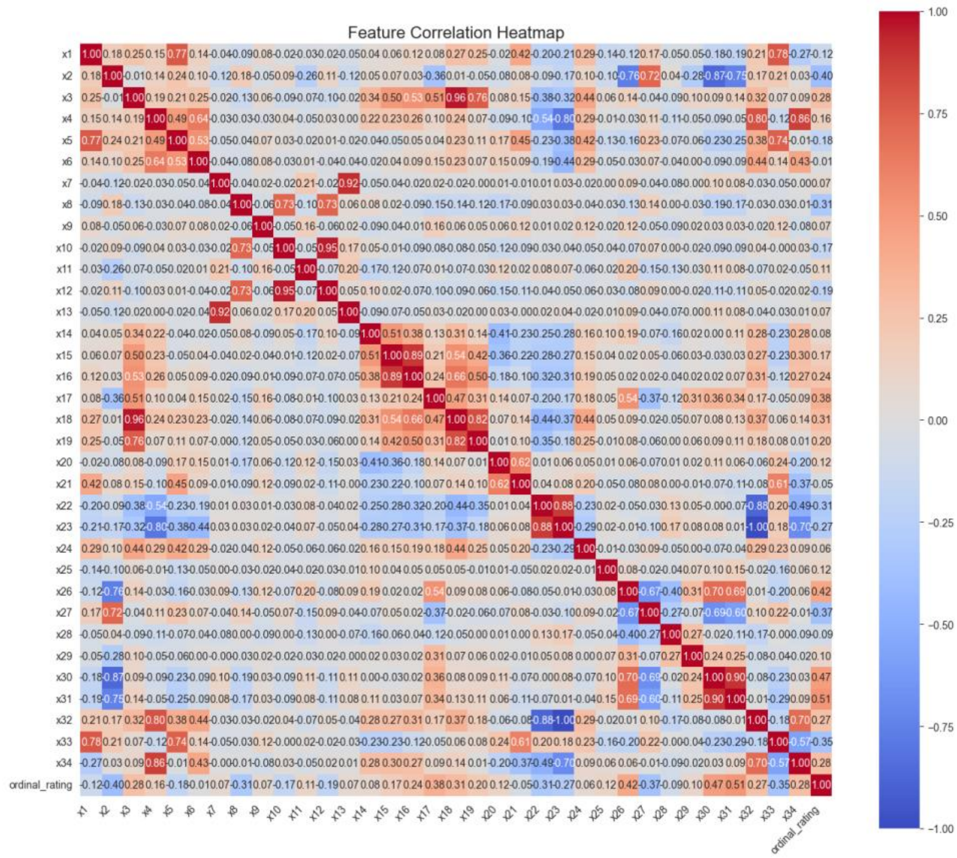

Fig 3. The Correlation heatmap on variables & credit ratings

The individual correlation coefficients between each variable and the credit rating are reported in Table 3 Feature selection for model training was based on Pearson's correlation coefficient, with a threshold of absolute value greater than 0.3. Only variables with a coefficient greater than 0.3 or less than -0.3 were retained for further analysis. The selected variables that met this criterion are listed in Table 4.

Table 3. Individual correlation coefficients between each variable and the credit rating

Features	Correlation coefficient
x34	0.28332
x3	0.278085
x32	0.2732

Features	Correlation coefficient
x16	0.240356
x19	0.19594
x15	0.16605
x4	0.15522
x20	0.123678
x25	0.11664
x11	0.105078
x29	0.098044
x14	0.078715
x7	0.067644
x13	0.066906
x9	0.066557
x24	0.064501
x6	-0.012738
x21	-0.051572
x28	-0.086679
x1	-0.120037
x10	-0.165641
x5	-0.182535
x12	-0.189157
x23	-0.2732
x22	-0.305065
x8	-0.312104
x33	-0.353586
x27	-0.374499
x2	-0.401965
x34	0.28332
x3	0.278085
x32	0.2732
x16	0.240356
x19	0.19594

Table 4. selected variables

Features	Correlation coefficient
x31	0.51239
x30	0.471731
x26	0.423293
x17	0.37784
x18	0.310697
x22	-0.305065
x8	-0.312104
x33	-0.353586
x27	-0.374499
x2	-0.401965

2) Machine learning model training, Evaluation, and Experimental Design

To evaluate the effectiveness of machine learning algorithms in predicting corporate credit ratings, the study was designed to explore multiple combinations of preprocessing and modeling configurations. Two datasets were prepared: one with 13 distinct credit rating levels and another simplified into 4 group levels. Each dataset was split into training and test sets using three different methods: random sampling, company-based grouping, and quarterly-based grouping. Feature scaling was applied using four techniques—None, MinMax, Robust, and Standard scaling—to assess their influence on model performance. Additionally, Synthetic Minority Over-sampling Technique (SMOTE) was optionally applied to address class imbalance issues.

Five machine learning models were trained: Logistic Regression, Decision Tree, Random Forest, XGBoost, and Artificial Neural Network (ANN). For each configuration, hyperparameters were optimized using a grid search method. The resulting models were evaluated using accuracy scores an confusion matrices to ensure both overall and class-level prediction effectiveness. This experimental design enables a comprehensive analysis of how preprocessing choices and model selection affect the accuracy of credit rating predictions

The experimental design summarize in Table 5.

Table 5. Experimental design summary

No.	Parameter	Options / Description
1	Dataset	2 datasets: 1) 13-class credit rating, 2) 4-class credit rating

No.	Parameter	Options / Description
2	Data Splitting Methods	3 splitting methods: 1) Random split 2) Split By Company 3) Split By Quarter
3	Data Scaling Methods	4 Scaling methods: 1) None 2) Min-Max Scaler 3) Robust Scaler 4) Standard Scale
4	Oversampling	SMOTE applied 1) No 2) Yes
5	Machine Learning models	5 Machine learning models: 1) Logistic Regression 2) Decision Tree 3) Random Forest 4) XGBoost 5) ANN
6	Hyperparemeter Tuning	Grid Search used to find best model parameters
7	Evaluation Metrics	1) Accuracy 2) Confusion Matrix

4 Result

The performance of the machine learning models was evaluated using metrics such as **accuracy** and the **confusion matrix**. Table 6 presents the average accuracy of each model across all conducted experiments. Among the five models evaluated, the **Random Forest** model achieved the highest average accuracy (0.5565), followed by **XGBoost** (0.5224), **Decision Tree** (0.4863), **Artificial Neural Network (ANN)** (0.4469), and **Logistic Regression** (0.3404).

Table 6. Average accuracy of models on every experiments

Model	Average accuracy
Random forest	0.5565
Xgboost	0.5224
Decision tree	0.4863
Artificial Neural network	0.4469
Logistic Regression	0.3404

To determine the most suitable model for predicting ranking outcomes, a comparative analysis was conducted between the top two performing models—**Random Forest** and **XGBoost**. The best-performing experiment from each model was selected based on maximum achieved accuracy. The detailed results, including data preprocessing methods such as scaling, splitting strategy, and application of SMOTE, are presented in Table 7.

Table 7. Best-performing experiments of Random Forest and XGBoost by dataset

Model	Scaling method	Splitting method	SMOTE	Dataset	Accuracy
Random forest	Standard scale	Random split	Yes	4-classes	0.8412
	None	Random split	Yes	13-classes	0.5751
XGBoost	Standard scale	Random split	Yes	4-classes	0.8455
	Standard scale	Random split	Yes	13-classes	0.5236

The **XGBoost** model yielded the highest accuracy of **84.55%** on the 4-class dataset. This result was obtained using **random splitting, standard scaling (StandardScaler)**, and **SMOTE** for addressing class imbalance. The optimal hyperparameters for this XGBoost configuration were: learning_rate = 0.2, max_depth = 5, and n_estimators = 300.

In contrast, for the **13-class dataset**, the **Random Forest** model performed better, achieving an accuracy of **57.51%.** This experiment used a similar setup to that of XGBoost but did not apply any feature scaling.

5 Conclusion

This research aimed to develop machine learning models capable of predicting the credit ratings of publicly listed companies in the Stock Exchange of Thailand, using financial ratios and rating score evaluated by TRIS Rating Co., Ltd. The ultimate goal was to provide investors with an additional decision-support tool that enhances investment efficiency and confidence.

Through experimentation with multiple machine learning models—including Random Forest, XGBoost, Decision Tree, Artificial Neural Network, and Logistic Regression—it was found that the most accurate model for the **4-level rating classification** was **XGBoost**, achieving an accuracy of **84.55%**. This was based on a training configuration that included random data splitting, standardization with StandardScaler, and class balancing using SMOTE. In contrast, when attempting a **13-level classification** to match TRIS Rating's full rating scale, the best-performing model was **Random Forest**, which achieved an accuracy of **57.51%**, deemed insufficient for practical deployment.

Despite its reduced granularity, the 4-level rating classification model provides a meaningful approximation of a company's creditworthiness trend and is therefore viable for use in preliminary investment assessments.

The study also analyzed how different data-splitting strategies—random split, year-based split, and company-based split—affected model performance. It was observed that random and year-based splits yielded similar distributions between training and testing datasets, resulting in higher model performance. In contrast, the company-based split produced inconsistent distributions, leading to reduced prediction accuracy.

Limitations encountered during this research include the relatively small and imbalanced dataset, particularly the underrepresentation of certain credit rating levels. This data imbalance likely impacted the models' ability to learn and predict accurately.

6 Policy Recommendations

6.1 Support Open Financial Data Initiatives

Regulatory bodies and financial institutions should promote open access to historical financial and credit rating data. Enhanced data availability would empower researchers and financial technology developers to build more reliable predictive tools, fostering innovation in investment analytics.

6.2 Promote Collaboration between Academia and Industry

Government agencies and private financial institutions should encourage partnerships that enable academic researchers to work with real-world financial data under secure and privacy-respecting conditions, thus bridging the gap between theoretical research and practical application.

References

1. Napasorn Thavichaigarn. (2022). Corporate Credit Rating Prediction Using Deep Learning. *Chulalongkorn University Theses and Dissertations (Chula ETD).* 5808.

2. Zeyu Huang, Savio Pereira, Meiqi Liu. (2024). *Credit Rating Prediction Using Machine Learning Techniques*. Available at: URL: https://fsc.stevens.edu/credit-rating-prediction-using-machine-learning-techniques/. Accessed May 10, 2024.
3. Pamuk, Mustafa, and Matthias Schumann. (2003). Opening a New Era with Machine Learning in Financial Service? Forecasting Corporate Credit Rating Based on Annual Financial Statements. *International Journal of Financial Studies.* 11: 96.
4. Edward I. Altman. (2005). An Emerging Market Credit Scoring System for Corporate Bonds. *Emerging Market Review*. 2005 (6). 331-323.
5. Anya Kanthawit, Ph.D. (2016). *ACE ECAIs: The Mapping of Risk Rating Scales*. Academic lecture on *ACE Portfolio Investment*, March 8, 2016. Faculty of Commerce and Accountancy, Thammasat University, in collaboration with the Stock Exchange of Thailand.