Journal of Computer Science

Research Article

Predicting Financial Statement Failure of Listed Firms in Market for Alternative Investment (mai) using Machine Learning

[1] **Anupong Sukprasert,** [2] **Pattanapong Kumsin and** [2] **Salakjit Ninlaphay**

[1] *Computer Department, Mahasarakham Business School, Mahasarakham University, Mahasarakham, Thailand*
[2] *Accountancy Department, Mahasarakham Business School, Mahasarakham University, Mahasarakham, Thailand*

Article history
Received: 20-04-2025
Revised: 14-06-2025
Accepted: 26-07-2025

Corresponding Author:
Pattanapong Kumsin
Accountancy Department,
Mahasarakham Business School,
Mahasarakham University,
Mahasarakham, Thailand
Email: Ae162557@gmail.com

Abstract: The purpose of this research is to investigate the financial ratios of companies listed on the Market for Alternative Investment (mai) in Thailand, as well as the likelihood of these companies' financial statement failure. It also intends to create a financial statement failure predicting model for companies listed on the mai and assess the performance of various forecasting models using machine learning techniques. This study predicts financial failure among firms listed on Thailand's Market for Alternative Investment (mai) using machine learning. Financial ratios from 127 companies (2019–2023) were analyzed using Z-Score-based financial distress indicator to label distress. three machine learning techniques: Logistic Regression, Deep Learning and k-Nearest Neighbors technique. were developed using Cross Industry Standard Process for Data Mining (CRISP-DM) framework, k-Nearest Neighbors outperformed others with 96.54% accuracy, 96.34% recall, 95.70% precision, and 96.00% F-measure, making it a practical tool for identifying high-risk firms. These findings highlight k-Nearest Neighbors model's superior ability to handle the dataset's complexity and non-linearity, making it the most effective technique for predicting financial statement failure.

Keywords: Financial Distress Prediction, Altman Z-score, Logistic Regression, Deep Learning, k-Nearest Neighbors

Introduction

After the COVID-19 crisis, Thailand's economy has progressively recovered, with the industrial and service sectors driving growth. This recovery is being fueled by rising private investment and exports (Bank of Thailand, 2024). Despite persistent global uncertainties, financial risk assessment remains essential for investors. The Market for Alternative Investment (mai) in Thailand is an appealing option because of its rapid expansion, strong service standards, and use of current technologies, notably quick and fast online trading. Supporting and educating new investors is critical for building understanding of the mai market, which is ideal for people looking to grow their wealth. However, investors should proceed with caution, as many firms still face financial challenges following the pandemic. Preliminary investigation revealed that some financial ratios of market for Alternative Investment listed companies improved during the crisis period, prompting a deeper analysis. This led to the present study, which applies machine learning techniques to examine financial ratios and identify patterns indicative of financial distress. Although investing in a dynamic market like the mai offers long-term growth potential, evaluating firm vulnerabilities and adopting diversification strategies remain essential to mitigate future risks (World Bank, 2023).

The Market for Alternative Investment (mai) in Thailand has expanded rapidly, with increased trade volume and high-quality services attracting both domestic and global investors. This expansion has increased confidence in the Thai market and solidified its reputation in Asia. This rise has been fueled by using contemporary technology, particularly convenient and fast online trading. Furthermore, helping new investors understand the market has been critical in making the mai an appealing option for investment and wealth growth (Finspace, 2024) However, The economic disruption from COVID-19 had a significant adverse impact on mai-listed companies. Lockdowns and travel restrictions severely affected the tourism and service sectors. Numerous small businesses closed, leading to widespread job losses and deteriorating financial indicators such as lower liquidity ratios and higher debt-to-equity ratios, thereby elevating bankruptcy (Bradley & Stumpner, 2021).

Financial ratios are derived from a company's financial data, encompassing revenue, expenses, assets, and liabilities. This study will employ these ratios to evaluate the company's financial performance, overall health, operational efficiency, and potential financial risks (Dechanubeksa, 2024; Apichaimongkol, 2023; Varnaprux *et al.*, 2022; Kim *et al.*, 2023, Srijunpetch *et al.*, 2019). This study analyzes eight key financial ratios current assets to total assets, retained earnings to total assets, EBIT to total assets, net profit margin, gross profit margin, net profit ratio, debt-to-equity ratio, and quick ratio to predict potential financial difficulties in mai-listed companies. The Altman Z-score model is used to forecast the likelihood of firm failure.

The model uses discriminant analysis techniques to forecast financial failure, with financial data serving as independent variables to help limit investment decision risks. According to Altman (1968) financial ratios can reliably forecast insolvent enterprises (94% accuracy) and non-bankrupt companies (97% accuracy) (Altman *et al.*, 1998). Significant indicators include retained earnings to total assets, return on assets, and total debt to total assets. Despite widespread use of the Z-score model and conventional machine learning techniques in financial prediction studies, most existing research has concentrated on companies listed on the Stock Exchange of Thailand (Elewa, 2022; Panigrahi, 2019; Matejić *et al.*, 2022). In contrast, studies focusing specifically on mai-listed companies typically smaller and more vulnerable to post-pandemic shock remain scarce. Furthermore, comparative assessments of multiple machine learning models in this context are limited. Most prior studies on financial distress prediction have focused on large firms or single models. Few have examined mai-listed companies typically smaller and more vulnerable post-COVID using a comparative machine learning approach. This study addresses that gap using Z-score labeling and data from 2019-2023.

The objective of this research is to evaluate whether the Z-score model can effectively support the prediction of financial distress in mai-listed companies. In addition, the study develops predictive models using three machine learning algorithms: Logistic regression, Deep learning, and k-Nearest Neighbors (k-NN). These models aim to help stakeholders proactively identify financial risk and make informed decisions. Data from 2019 to 2023 are utilized.

This research contributes to the literature by (1) comparing the predictive performance of three machine learning models tailored to mai-listed companies; (2) integrating the Altman EM-Score with the The Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology for structured and replicable modeling; and (3) offering insights into post-COVID financial behavior specific to small-cap Thai firms.

Materials and Methods

Theoretical Background

This study on forecasting financial failures utilized two relevant theories of financial distress

Agency Theory, proposed by Jensen and Meckling in 1976, explains the relationship between a principal and an agent, with an emphasis on optimizing investment returns while mitigating the risk of agent misconduct. A significant challenge within this framework is information asymmetry, wherein agents may withhold vital financial information, thereby undermining the organization's credibility and potentially adversely affecting stock prices and decision-making processes (Wanganusorn *et al.*, 2022; Rahman *et al.*, 2019; Kumpha, 2022).

Signaling Theory, developed by Spence in 1973, explores the methods by which parties possess differing levels of information communicate. Executives frequently strive to cultivate investor trust through transparency, often employing dividend payments as indicators of a company's financial health. A reduction in dividends can negatively influence investor perceptions, suggesting potential declines in profitability; however, it may also signal necessary capital investments during periods of growth (Santiwat & Sukcharoensin, 2021).

Stakeholder Theory, articulated by Freeman in 1984, underscores the significance of stakeholder interests in organizational management and corporate ethics. By prioritizing the needs of stakeholders, companies can foster mutual benefits. Freeman classifies stakeholders into primary (customers, employees, shareholders) and secondary (government, communities, activist groups) categories, emphasizing the importance of nurturing positive relationships to facilitate knowledge exchange and enhance public perception (Poonpool & Plukphonngam, 2020).

Financial Failure

The Altman model has faced challenges due to varying coefficients of independent variables, leading to regular updates by Altman and others. These modifications account for differences in accounting practices and economic factors like currency fluctuations and industry risks across countries. To tackle these issues, One notable adaptation is the Emerging Market Scoring (EMS) Model, incorporating four financial variables: working capital to total assets ratio (X_1), return on assets ratio (X_2), earnings before interest and taxes to total assets ratio (X_3), and book value of equity to total liabilities ratio (X_4), with recalibrated coefficients (Kumpha, 2022) The form of the Z-EM score equation is:

$$Z = 3.25 + 6.56X1 + 3.26X2 + 6.72X3 + 1.05X4$$

Where:

X_1 = the capital-to-total assets ratio

X_2 = Return on Assets Ratio

X_3 = Earnings before interest and taxes to total assets ratio

X_4 = Book value of equity to total liabilities ratio

Several studies affirm the model's effectiveness. Sawatwutthiphong (2022) found significant correlations between stock price volatility, the debt-to-equity ratio (DE), and the current ratio (CR). Kumpha (2022) reported 95.50% accuracy in forecasting financial distress using the EMS model. Mongkolsareechai (2020) documented a 92.38% success rate, while Nandee (2016) observed a 67% prediction accuracy among SMEs. Other research also supports the model's utility. Elewa (2022) identified key ratios such as Net Working Capital to Total Assets (NWC/TA), Retained Earnings to Total Assets (RE/TA), EBIT to Total Assets (EBIT/TA), Book Value of Equity to Total Liabilities (BVE/TL), and Sales to Total Assets (S/TA) as having a significant impact on predictive performance. Durica *et al.* (2019) developed a decision tree model using 37 financial and economic indicators, achieving over 98% accuracy in classifying non-prosperous companies and 83% success in forecasting future distress. Their results underscore the potential of decision trees for financial failure prediction in both Polish and international contexts. This study investigates the applicability of Altman's EM-Score model in forecasting financial failure among mai-listed companies in Thailand. Additionally, it evaluates the predictive performance of machine learning techniques and compares their effectiveness to that of the EM-Score model, to determine whether these modern approaches offer superior accuracy.

Machine Learning

Machine learning is a key form of data analysis using automated models, situated at the intersection of artificial intelligence and computer science. It focuses on developing algorithms that allow computers to learn from data without explicit programming. Definitions from pioneers in the field emphasize its statistical and computational basis aimed at improving performance through learning. Currently, machine learning is applied in various industries, such as healthcare for enhancing diagnostic accuracy and in finance for predicting market trends and detecting fraud. These advancements have the potential to significantly improve quality of life (Samuel, 1959; Mitchell, 1997; Goodfellow *et al.*, 2016; Shen *et al.*, 2017).

Machine Learning is widely applied across sectors such as healthcare to enhance diagnostic accuracy and finance for forecasting market trends and detecting fraudulent activity. These advancements play a key role in supporting data-driven decision-making while enhancing overall quality of life.

Research Utilizing Machine Learning Techniques

This research employs three supervised machine learning techniques Logistic Regression, Deep Learning, and k-Nearest Neighbors (k-NN) to build predictive models for identifying financial distress among mai-listed companies. Each method is selected for its unique strengths in handling structured financial data and classification tasks.

Logistic Regression

Logistic regression is a Machine Learning method used for analyzing datasets containing a categorical dependent variable, particularly when that variable is binary (e.g., yes or no, success or failure). There are two types of logistic regression.

Binary Logistic Regression

This Machine Learning method is specifically utilized in classification tasks where the dependent variable is binary. The analysis aims to estimate the probability of a particular event occurring by examining the relationship between the binary dependent variable and various independent factors (Sukprasert *et al.*, 2024).

Multinomial Logistic Regression

A Machine Learning method called multinomial logistic regression is used to examine data including a categorical dependent variable with more than two categories. Stated differently, it is applied when the result may be separated into several groups. This study can make use of both continuous and categorical independent variables.

Deep Learning

Artificial neural networks with multiple hidden layers are referred to by Minaee *et al.* (2021) There are two primary processes in deep learning for action recognition.

Feature Extraction

High-level features of the data are extracted by stacking several layers, such as pooling and convolutional layers. A high-level representation of the processed data is the result.

Inference Model

A model, like a multilayer perception, is developed using the features that were retrieved to forecast the likelihood of the target class.

Deep learning contains more than two hidden layers, which is the main distinction between it and conventional neural networks. By employing filters of different sizes that carry out matrix multiplication on multi-dimensional arrays, convolutional layers automatically extract features from data. As illustrated in

DOI: 10.3844/jcssp.2025.2171.2180

Figure 1, for instance, a positive matching value in image processing denotes a similarity between the image and the filter, whereas a negative value denotes a difference.

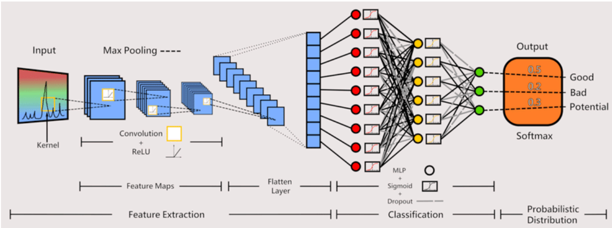

Fig. 1: Convolution Neural Network

Source: Kissi et al. (2021)

k-Nearest Neighbor

k-Nearest Neighbors is a method for determining a data point's closest neighbors. Calculating the distance between the data point to be classed and its k nearest neighbors is its main application. Two data points are separated by a distance function, and the distances are ranked from smallest to greatest using a combination function. The number of nearest neighbors considered when classifying depends on the value of k. This method works especially well with numerical data since it makes it easy to compute the distances between various attributes.

Model Development and Performance Evaluation

To evaluate the predictive capability of the financial forecasting models, this study employed the k-fold cross-validation technique. The complete dataset was partitioned into *k* equally sized subsets or "folds". In each iteration, one-fold was used as the validation set, while the remaining *k-1* folds served as the training set. This process was repeated *k* times, ensuring that each subset was used once for testing. The final model performance metrics, including accuracy, precision, recall, and F-measure were calculated as the average across all folds.

k-fold Cross Validation

k-fold cross validation is a technique used to assess the accuracy of a predictive model. The dataset is divided into k equal-sized subsets, or folds. In each iteration, one-fold is designated as the test set, and the remaining k-1 folds are used as the training set. This process is repeated K times, with each fold serving as the test set exactly once.After completing all iterations, the average and standard deviation of the performance metric (e.g., accuracy) are calculated.

This method ensures that every data point has an opportunity to be in both the training and test sets, which is a significant advantage over the traditional approach of splitting data into a single training set and a single test set. The latter method might not accurately represent the overall performance of the model, as the specific split of data can significantly influence the results.

k-fold cross-validation helps in identifying the optimal training set and provides a more robust estimate of the model's performance. It also enables comparisons between different modeling techniques. Figure 2 illustrates this process.

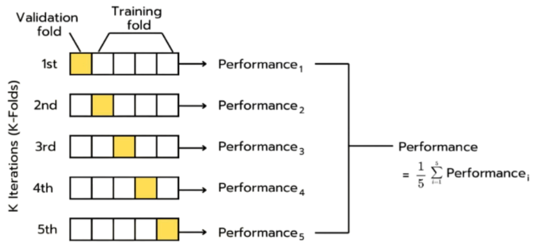

Fig. 2: k-Fold Cross Validation

Cross-Industry Standard Process for Data Mining Framework for Data Analysis

Data analysis in this study involves the application of various tools, techniques, and technologies to convert raw financial data into meaningful insights that support issue resolution, pattern recognition, and strategic decision-making. The Cross-Industry Standard Process for Data Mining (CRISP-DM) was adopted as the methodological framework guiding this research. CRISP-DM is a well-established and widely utilized data mining framework comprising six sequential phases:

1. Business Understanding: The initial step involves defining specific business objectives and understanding the associated opportunities or challenges. This phase is crucial as it ensures that subsequent stages of the data mining process align with organizational goals.
2. Data Understanding: Collect, explore, and describe the dataset to identify quality issues and gain initial insights.
3. Data Preparation: Cleanse, integrate, and transform the data into a suitable format for modeling.
4. Modeling: Apply machine learning algorithms to develop predictive models that capture patterns and trends.
5. Evaluation: Assess the model's performance by using predefined metrics to ensure it meets the intended business objectives.
6. Deployment: Implement the validated model in a real-world context, enabling data-driven decisions and practical application.

A graphical representation of these phrases is illustrated in Figure 3.

Sample Selection and Industry Distribution

The population for this study consists of 127 enterprises listed on the Market for Alternative

Investment (mai) in Thailand excluding companies in the financial sector. This exclusion is due to the distinct regulatory framework and operational characteristics of financial institutions governed by the Bank of Thailand. Table 1 presents the distribution of companies by industry.

Fig. 3: Cross Industry Standard Process for Data Mining (CRISP-DM)

Table 1: Distribution of mai-Listed Companies in Thailand by Industry

No.	Industry	Number of Companies
1	Agriculture and Food	7
2	Consumer Goods	10
3	Industrial Goods	31
4	Real Estate and Construction	18
5	Resources	10
6	Services	41
7	Technology	10
Total		127

Construction and Data Analysis Tools

The analytical design is based on eight financial ratios, identified from prior studies (Kumpha, 2022; Kosanlawit *et al.*, 2021): 1) Current Ratio to Total Assets): 1) Current Ratio to Total Assets 2) Return on Assets 3) EBIT Ratio 4) Net Profit Margin 5) Gross Profit Margin 6) Net Profit Ratio 7) Debt-to-Equity Ratio and 8) Quick Ratio. These serve as independent variables, while the dependent variable-financial distress-is derived using the Altman Z-Score Model (Apichaimongkol, 2023).

The dataset was structured using Microsoft Excel and subsequently analyzed using RapidMiner Studio 10.2, a data analytics software (Sukprasert, 2021). The research employs supervised machine learning classification techniques within the CRISP-DM framework, which includes the following six phases: Business Understanding, data understanding, data preparation, modeling, evaluation, and deployment (Sukprasert *et al.*, 2025).

Business Understanding

The increasing incidence of corporate failures among market for alternative investment listed companies particularly during and after the COVID-19 pandemic has led to business restructuring, layoffs, and bankruptcies. This study seeks to address such challenges by applying machine learning techniques to internal company data, to develop predictive models that support proactive financial risk management. The combined use of Altman's Z-Score and machine learning enables the construction of a forecasting model that provides valuable decision-making tools for investors, executives, and regulators.

Data Understanding

Financial data from market for alternative investment listed companies for the years 2019 to 2023 were obtained from the SEC and SETSMART databases. Eight financial ratios were selected and compiled into an Excel spreadsheet. The dataset consists of 635 rows and 9 columns, representing 5,715 data points. Table 2 presents a summary of the attributes used in the analysis.

Table 2: Description of Attributes Used in Predictive Modeling

No.	Attribute	Data Type	Description
1	Year	Date	Fiscal year: 2019-2023
2	Companies	Polynomial	Symbol of companies listed on the mai
3	Current Ratio	Real	Liquidity ratio
4	Return on Asset	Real	Return on assets
5	EBIT	Real	Earnings before interest and taxes
6	Return on Equity	Real	Return on equity
7	Gross Profit Margin	Real	Gross profit margin
8	Net Profit Margin	Real	Net profit margin
9	Debt to Equity Ratio	Real	Debt-to-equity ratio
10	Quick Ratio	Real	Quick ratio
11	Z-Score (Label)	Real	Altman's Z-score, a financial distress predictor

Data Preparation

This research utilized RapidMiner Studio 10.2 to import pre-arranged data from an Excel file. The data was inspected and prepared prior to model building. The process was divided into four main parts:

1. Data Selection: Firms with incomplete records, newly listed firms, and companies that migrated from mai to SET were excluded. Data were segmented into three periods: pre-COVID (2019), during COVID (2020-2021), and post-COVID (2022-2023).
2. Data Integration: Financial data across five years were combined into a single dataset with 635 rows and 9 financial indicators

DOI: 10.3844/jcssp.2025.2171.2180

3. Data Cleansing: Only firms with complete annual financial data were retained. Incomplete records and delisted entities were removed
4. Data Transformation: The Altman's EM Score formula $Z = 3.25 + 6.56X_1 + 3.26X_2 + 6.72X_3 + 1.05X_4$ was used to transform the dependent variable into a format suitable for analyzing and forecasting financial distress, as shown in Table 3 (Altman *et al.*, 2013).

Table 3: Input Variables for Financial Distress Prediction Using Altman's EM-Score Model

No	Co.	X_1 (CR)	X_2 (ROA)	X_3 (EBIT)	X_4 (Equity/Liability)
1	AU	4.4	25.54	5.53	6.43
2	MOONG	1.77	9.83	11.99	3.38
3	NDR	1.12	-0.98	-1.3	1.96
4	CRD	1.38	-1.85	-0.75	0.88
5	SONIC	2.46	8.3	6.16	2.40
6	VCOM	1.49	8.34	5.59	0.67

Modeling

Model Development using RapidMiner Studio 10.2, predictive models were developed to classify financial failure among mai-listed firms. Based on a review of relevant literature and the classification nature of the problem, three machine learning algorithms were selected: Logistic Regression, Deep Learning, and k-Nearest Neighbors (k-NN).

Logistic Regression is a statistical analysis technique, a type of machine learning, used to predict the probability of a binary outcome, such as yes or no, pass or fail, success or failure. The calculation formula is presented in equation (Chuchip, 2018; Goi *et al.*, 2019).

$$Prob(even) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1)}}$$

Where:

e = Base of the natural logarithm (Approximately 2.718)

< class="no_indent"p>β_0 = Intercept
β_1 = Coefficient

X_1 = Independent variable

Deep learning is a machine learning technique that utilizes artificial neural networks with multiple layers to process complex data, such as images, sound, and text. It excels at learning from large datasets with intricate structures, as illustrated in Figure 4.

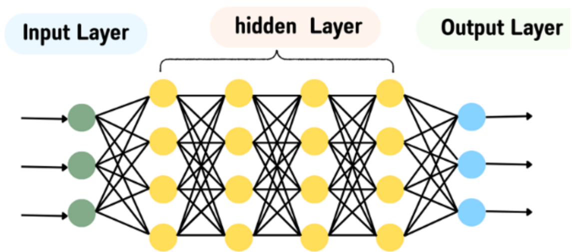

Fig. 4: Deep Learning Technique

k-Nearest Neighbors (k-NN) is a statistical learning algorithm employed for both classification and regression problems. It determines the class or value of a new data point by identifying its K nearest neighbors within the existing dataset, as depicted in Figure 5.

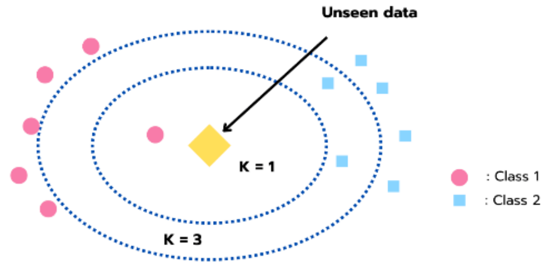

Fig. 5: k-Nearest Neighbor Technique

Feature Extraction

No complex feature extraction methods like Principal Component Analysis (PCA) were used. Instead, eight financial ratios were chosen based on their theoretical relevance and consistent application in bankruptcy prediction studies. This approach ensures transparency and maintains interpretability aligned with standard financial analysis practices.

Evaluation

The study used 10 folds Cross-validation to evaluate model performance. The data was divided into two parts: training and testing, with the latter partitioned into ten partitions for improved reliability. Table 4 shows a cross-tabulation that assesses expected accuracy across four performance measures.

Where:

- True Positive (TP): Correctly predicted financial distress cases.
- True Negative (TN): Correctly predicted non-distress cases
- False Positive (FP): Incorrectly predicted distress for a healthy company
- False Negative (FN): Failed to identify an actual distress case.

Table 4: Confusion Matrix

	Actual: Positive	Actual: Negative
Predicted: Positive	True Positive (TP)	False Positive (FP)
Predicted: Negative	False Negative (FN)	True Negative (TN)

Deployment

The results of this study can be used to create a model for anticipating financial difficulty among listed companies on the mai in Thailand. This model can assist executives and investors in forecasting future failures, mitigating risks in management and investment decisions, identifying opportunities to remedy prospective failures.

Results

The researchers employ a confusion matrix to assess the accuracy of financial failure predictions by comparing actual and anticipated statuses. A confusion matrix contains crucial data such as True Positive, True Negative, False Positive, False Negative, which aid in understanding accuracy, recall, precision, and F-measure. In addition, confusion matrices are utilized to display the proportion and percentage of correct guesses. The analysis results are evaluated based on these variables to summarize the reliability and efficacy of the prediction model.

Logistic Regression

The researchers created a model by choosing eight independent variables that were believed to influence the chance of a company's financial failure. The model's performance was assessed using a confusion matrix, the results are shown in Table 5.

Table 5: Confusion Matrix and Classification Performance of the Logistic Regression Model

	Actual: Positive	Actual: Negative	Total (Correct %)
Predicted: Positive	307	52	359 (85.52%)
Predicted: Negative	56	220	276 (79.71%)
Classification Performance			
Accuracy		83.00%	
Recall		80.86%	
Precision		80.39%	
F-measure		80.28%	

Table 5 presents the confusion matrix for the logistic regression model, demonstrating its predictive performance. The model correctly identified 307 instances of companies that did not experience financial failure (True Positives) and 220 instances of companies that did (True Negatives). It achieved an accuracy of 83.00%, with a recall of 80.86%, precision of 80.39%, and an F-measure of 80.28%.

Deep Learning

Researchers developed a model by considering and selecting eight independent variables that are likely to predict financial statement errors. The researchers created a model by considering and selecting a total of eight independent variables that are likely to predict financial statement failure. Model creation researchers used Deep Learning techniques to develop a forecasting model. Using the RapidMiner Studio 10.2 application, an equation was created with all eight selected independent variables, and the results are shown in Table 6.

Table 6 presents the confusion matrix for the deep learning model, demonstrating its predictive performance. The model correctly identified 348 instances of companies that did not experience financial failure (True Positives) and 261 instances of companies that did (True Negatives). It achieved an accuracy of 95.91%, with a recall of 95.22%, precision of 94.57%, and an F-measure of 94.70%.

Table 6: Confusion Matrix and Classification Performance of the Deep Learning Model

	Actual: Positive	Actual: Negative	Total (Correct %)
Predicted: Positive	348	11	359 (96.94%)
Predicted: Negative	15	261	276 (94.57%)
Classification Performance			
Accuracy		95.91%	
Recall		95.22%	
Precision		94.57%	
F-measure		94.70%	

k-Nearest Neighbor

The k-Nearest Neighbors technique was used by the researchers to create a prediction model. They used RapidMiner Studio 10.2 to create an equation with all eight specified independent variables. This analysis's conclusions will be reported in Table 7, which outlines the process of calculating the ideal value for the parameter 'k' in the k-Nearest Neighbors technique.

Table 7: Accuracy Comparison for Different k Values in the k-Nearest Neighbors Algorithm

Number of Neighbors (k)	Accuracy
1	96.5
3	95.9
5	96.4
7*	96.5
9	95.7
11	95.9
13	96.1
15	96.1

** Indicates the optimal value for the parameter k*

According to Table 7 the optimal k-Nearest Neighbors value is 7, which yields the best accuracy of 96.5%. The researchers then used the ideal value to create the predicted results shown in Table 8.

Table 8: Confusion Matrix and Classification Performance of the k-Nearest Neighbors Model

	Actual: Positive	Actual: Negative	Total (Correct %)
Predicted: Positive	351	10	361 (97.23%)
Predicted: Negative	12	262	274 (95.62%)
Classification Performance			
Accuracy		96.54%	
Recall		96.34%<	
Precision		95.70%	
F-measure		96.00%	

Table 8 presents the confusion matrix for the k-Nearest Neighbors model, demonstrating its predictive performance. The model correctly identified 351 instances of companies that did not experience financial failure (True Positives) and 262 instances of companies that did (True Negatives). It achieved an accuracy of 96.54%, with a recall of 96.34%, precision of 95.70%, and an F-measure of 96.00%.

Following the confusion matrix for each forecasting technique that is created, the model's performance will be compared using the proportion or percentage of correct predictions, As can be seen, the k-Nearest Neighbor forecasting technique is more accurate than the logistic regression and deep learning techniques. As can be seen, k-Nearest Neighbor predicting technique is more accurate than the logistic regression and deep learning techniques (Table 9).

Table 9: Comparative Classification Performance of Machine Learning Models Note Technique that provides the best prediction of failure

Classification Techniques	Classification Performance			
	Accuracy	Recall	Precision	F-measure
Logistic Regression	83.00%	80.86%	80.39%	80.28%
Deep Learning	95.59%	95.22%	94.76%	94.87%
k-Nearest Neighbors*	96.54%	96.34%	95.70%	96.00%

Table 9 presents the comparative of predictive performance. The logistic regression model achieved an accuracy of 83.00%, with a recall of 80.86%, precision of 80.39%, and an F-measure of 80.28%. The deep learning model achieved an accuracy of 95.91%, with a recall of 95.22%, precision of 94.57%, and an F-measure of 94.70%. The k-Nearest Neighbors (k-NN) model achieved an accuracy of 96.54%, recall of 96.34%, precision of 95.70%, and F-measure of 96.00%.

These results suggest that the k-Nearest Neighbors offer superior capability in forecasting financial distress among market for alternative investment listed companies compared to logistic regression and deep learning techniques.

Conclusion

This study examined the predictive performance of Altman's EM-Score model combined with machine learning techniques to forecast financial failure among companies listed on Thailand's Market for Alternative Investment (mai) from 2019 to 2023, using a dataset of 635 records, found that: creating a model using the Z-Score value based on the Altman's EM-Score model and the logistic regression technique resulted in an accuracy of 83.00%, recall of 80.86%, precision of 80.39%, and F-measure of 80.28%. The deep learning model got an accuracy of 95.59%, recall of 95.22%, precision of 94.76%, An F-measure of 94.87%.

Furthermore, the k-Nearest Neighbors model has an accuracy of 96.54%, recall of 96.34%, precision of 95.70%, and F-measure of 96.0. The k-Nearest Neighbors (k-NN) model demonstrates a strong capability for managing data complexity and nonlinearity, particularly within the context of financial data. This assertion is supported by research conducted by Uludağ & Gürsoy (2020) who analyzed financial conditions using both k-NN and Naive Bayes classification algorithms. Their findings indicated that the k-NN technique achieved a classification accuracy ranging from 84% to 88%, outperforming the Naive Bayes method, which recorded an accuracy between 75% and 86%. This aligns with prior research showing that the k-NN model attained a forecast accuracy of 89.80% for consumer goods stocks on the Thai stock exchange (Kumsin *et al.*, 2024). Additionally, Uludağ & Gürsoy (2021) found that the k-NN technique could effectively assess financial risk through specific financial ratios, achieving an accuracy of 88.42%.

This study builds predictive models for financial failure among mai-listed firms using machine learning and domain-specific financial ratios. k-NN achieved the highest accuracy and proved suitable for non-linear data. By integrating Z-score labeling with the CRISP-DM methodology and domain-specific financial ratios, this research delivers a robust and interpretable framework for predicting financial distress in mai-listed companies

Implications and Future Directions

Future studies should expand the dataset to include companies beyond the mai, adopt quarterly rather than annual data, and integrate governance-related variables such as shareholder structure and board composition. The inclusion of more advanced modeling techniques such as hybrid neural networks or ensemble learning could further improve predictive power and generalizability. These enhancements will support investors, managers, and regulators in mitigating financial risks and promoting sustainable investment strategies.

Acknowledgements

This research project was financially supported by Mahasarakham University

Conflict of Interest

The authors declare no conflicts of interest.

Author's Contributions

Pattanapong Kumsin: Designed the research methods, collected and analyzed the data, drafted the manuscript and prepared the figures and tables

Salakjit Ninlaphay: Provided overall supervision, advised on the research framework and methodology and reviewed and edited the manuscript

Anupong Sukprasert: Supported research execution assisted in data interpretation and contributed to reviewing and refining the manuscript.

References

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, *23*(4), 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

Altman, E. I., Danovi, A., & Falini, A. (2013). Z-Score Model's Application to Italian Companies Subject to Extraordinary Administration. *Journal of Applied Finance*, *21*(1), 1–10.

Altman, E. I., Hartzell, J., & Peck, M. (1998). *Emerging market corporate bonds — a scoring system*. *2*, 391–400. https://doi.org/10.1007/978-1-4615-6197-2_25

Apichaimongkol, S. (2023). Correlation Between Financial Failure and Financial Ratios of Listed Companies on the Stock Exchange of Thailand. *Rajapark Journal*, *17*(54), 48–58.

Bank of Thailand (2024). *Thailand's Economy in 2024 and Structural Challenges*.

Bradley, C., & Stumpner, P. (2021). The Impact of COVID-19 on Capital Markets, one Year in *McKinsey&Company*.

Chuchip, K. (2018). Logistic Regression. *Remote Sensing Technical Note*, *5*(2018), 1–10.

Dechanubeksa, N. (2024). Financial Ratios that Influence the Price of Securities of Companies Listed on the Stock Exchange of Thailand: The Medical Business Sector. *Rajabhat Maha Sarakham University Journal*, *18*(1), 126–137.

Durica, M., Frnda, J., & Svabova, L. (2019). Decision tree based model of business failure prediction for Polish companies. *Oeconomia Copernicana*, *10*(3), 453–469. https://doi.org/10.24136/oc.2019.022

Elewa, M. M. (2022). Using Altman Z-Score Models for Predicting Financial Distress for Companies – The Case of Egypt panel data analysis. *Alexandria Journal of Accounting Research*, *6*(1), 1–28.

Finspace. (2022). *Investment Opportunities in the mai Stock Market*. https://www.finspace.co/stock-exchange-of-thailand-mai

Goi, M. T., Palil, M. R., Zain, N. H. M., & Mamat, N. H. (2019). Predicting Private Saving Within Malay Community. *Asian Academy of Management Journal*, *24*(Supp. 1), 1–17. https://doi.org/10.21315/aamj2019.24.s1.1

Goodfellow, Ian, Bengio, Yoshua, & Courville, Aaron. (2016). *Deep learning*.

Kim, S. S., Juliana, R., & Handoko, L. (2023). The effects of economic policy uncertainty and large investment project on leverage. *Asian Academy of Management Journal*, *28*(2), 333–369. https://doi.org/10.21315/aamj2023.28.2.12

Kissi, J., McIsaac, T., Osinski, K., R Osinski, G., & Shieh, S. (2021). A Convolutional Network Approach to Generalised. *Preprints*. https://doi.org/10.20944/preprints202111.0078.v1

Kumpha, P. (2022). Accuracy of the Altman's EM-Score Model and the Prediction of Financial Distress of Listed Companies in the Stock Exchange of Thailand. *Journal of Federation of Accounting Professions*, *4*(12), 26–43.

Kumsin, P., Bhongchirawattana, U., & Sukprasert, A. (2024). Forecasting Stock Price of Consumer products group in the Stock Exchange of Thailand using Box-Jenkins method. *International Journal of Mathematics and Computer Science*, *19*(3), 967–976.

Matejić, T., Knežević, S., Arsić, V. B., Obradović, T., Milojević, S., Adamović, M., Mitrović, A., Milašinović, M., Simonović, D., Milošević, G., & špiler, M. (2022). Assessing the Impact of the COVID-19 Crisis on Hotel Industry Bankruptcy Risk through Novel Forecasting Models. *Sustainability*, *14*(8), 4680. https://doi.org/10.3390/su14084680

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2022). Deep Learning--based Text Classification. *ACM Computing Surveys*, *54*(3), 1–40. https://doi.org/10.1145/3439726

Mitchell, T. M. (1997). Artificial neural networks. *Machine Learning*, *45*, 81–127.

Mongkolsareechai, C. (2020). An Analysis of Bankruptcy of Companies Registered in The Stock Exchange of Thaiand : A Case Study of Automotive Industry. *Journal of MCU Social Science Review*, *6*(2), 55–68.

Nandee, P. (2016). *Predicting financial distress in determining preliminary condition for approving loans to SMEs using Altman's EM Score Model*. https://doi.org/10.14457/TU.the.2016.97

Panigrahi, A. (2019). Validity of Altman's "Z" Score Model in Predicting Financial Distress of Pharmaceutical Companies. *NMINS Journal of Economics and Public Policy*, *4*(1), 65–73.

Poonpool, N., & Plukphonngam, C. (2020). Social and Environmental Accounting to Tax Incentive. *Journal of Graduate School of Commerce-Burapha Review*, *15*(1), 1–15.

Rahman, I. K. A., Hussain, M. D., & Hossin, Md. S. (2019). Microfinance Governance: A Multitheoretical Approach for Ascertaining the Wider Stakeholder Influencing Forces. *Asian Academy of Management Journal*, *24*(Supp. 1), 203–216. https://doi.org/10.21315/aamj2019.24.s1.14

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Santiwat, N., & Sukcharoensin, S. (2021). The relationship between cash dividend changes and future earning changes. *Chulalongkorn Business Review*, *43*(3), 1–21. https://doi.org/10.14456/cbsr.2021.9

Sawatwutthiphong, N. (2022). Forecasting volatility of stock price by using a model that predicts. *Thai Digital Collection – Thammasat University Repository*. https://doi.org/10.14457/TU.the.2022.436

Shen, D., Wu, G., & Suk, H.-I. (2017). Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, *19*(1), 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442

Srijunpetch, S., Boonlert-U.-Thai, K., & Phakdee, A. (2019). *Financial Reporting and Analysis*.

Sukprasert, A., Dananjoyo, R., Supachaiwat, J., & Boonlua, S. (2024). Developing a Model to Predict the Value of Thailand's Gem and Jewelry Exports Using Multiple Linear Regression Analysis. *Mathematical Modelling of Engineering Problems*, *11*(12), 3411–3419. https://doi.org/10.18280/mmep.111219

Sukprasert, A., Supachaiwat, J., & Khruahong, S. (2025). Modelling for Forecasting the Value of Thailand's Gems and Jewelry exports to the United States using Machine Learning. *ICIC Express Letters, Part B: Applications*, *16*(6), 599–608. https://doi.org/10.24507/icicelb.16.06.599

Uludağ, O., & Gürsoy, A. (2020). On the Financial Situation Analysis with KNN and Naive Bayes Classification Algorithms. *Iğdır üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *10*(4), 2881–2888. https://doi.org/10.21597/jist.703004

Uludağ, O., & Gürsoy, A. (2021). Financial Risk Estimation with KNN Classification Algorithm on Determined Financial Ratios. *European Journal of Science and Technology*, *29*, 26–29. https://doi.org/10.31590/ejosat.1001663

Varnaprux, R., Wiwatjarernwong, C., & Bosakoranut, S. (2022). The Effect of Financial Ratios on Market value ratios of Thai listed companies in the Stock Exchange of Thailand: Technology Industry. *National and International Sripatum University Conference*.

Wanganusorn, B., Tulasombat, S., & Thechatakerng, P. (2022). Agent Theory: Challenges of Business Administration in Thailand. *Maejo Business Review*, *4*(1), 1–22. https://doi.org/10.14456/mjba.2022.1

World Bank (2023). Thai Economy to Accelerate in 2023 as Demand from Major Economies Picks Up, Tourists Return. *World Bank*. https://www.worldbank.org/th/news/press-release/2023/06/28/thai-economy-to-accelerate-in-2023-as-demand-from-major-economies-picks-up-tourists-return