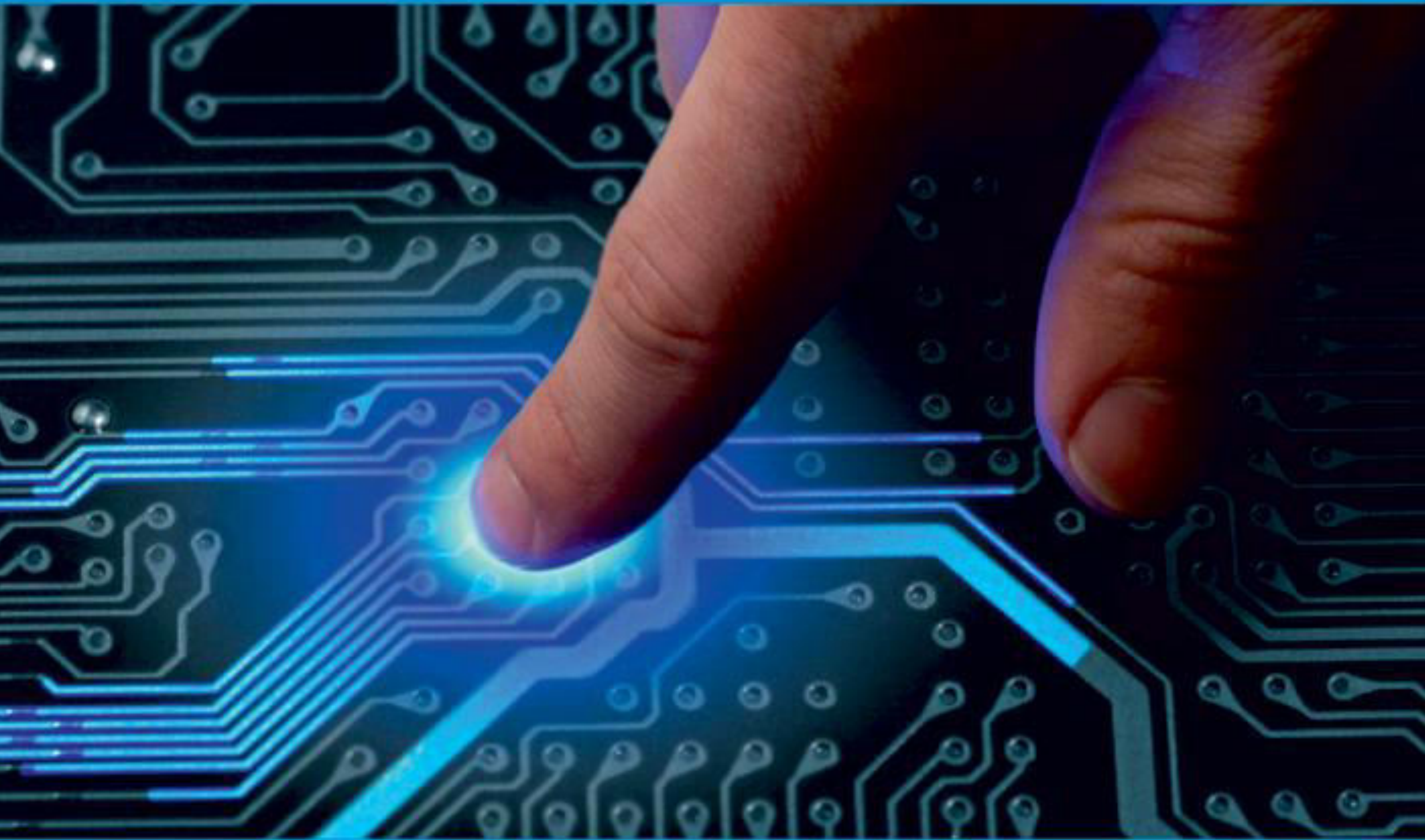
e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 11, November 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Bankruptcy Prediction using SVM: A Machine Learning Approach with Financial Ratios

Alias Itten, **Soumya K S**

Lecturer, Dept. of Computer Engineering, Govt. Polytechnic College, Kaduthuruthy, Kottayam, Kerala, India

Lecturer, Department of Computer Engineering, Govt. Polytechnic College, Kalamassery, Ernakulam, Kerala, India

ABSTRACT: This study presents a machine learning-based approach for predicting company bankruptcy using financial ratio data. Utilizing a Support Vector Machine (SVM) classifier, we analyze a comprehensive dataset comprising 95 financial attributes from Taiwanese firms, as provided on Kaggle. The objective is to accurately classify companies as either bankrupt or non-bankrupt based on their financial metrics. To improve prediction performance, the data is preprocessed with standardization and split into training and testing sets using stratified sampling to preserve class distribution. Given the inherent class imbalance, we employ class weighting in the SVM to enhance sensitivity to minority classes. The model is evaluated using classification metrics including accuracy, precision, recall, and F1-score, along with a confusion matrix visualization. Results indicate that the SVM model is effective in identifying financially distressed companies. This methodology demonstrates the potential of machine learning tools in financial risk assessment and early warning systems. Future enhancements may involve ensemble models and additional data sources to further boost predictive power.

KEYWORDS: Bankruptcy Prediction, Support Vector Machine, Financial Ratios, Machine Learning, Classification, Corporate Insolvency, Imbalanced Data, Risk Assessment.

I. INTRODUCTION

In today's rapidly changing economic landscape, financial distress and corporate bankruptcy remain significant concerns for investors, stakeholders, and regulatory authorities. Early identification of potentially bankrupt companies allows for timely intervention and more informed financial decision-making. Traditional bankruptcy prediction models, such as Altman's Z-score and logistic regression, rely on a limited set of financial ratios and assumptions that may not fully capture the complexity and dynamics of modern corporate finance. With the advancement of computational tools and the growing availability of financial data, machine learning techniques have emerged as effective alternatives for risk assessment and classification tasks in finance.

This study leverages a Support Vector Machine (SVM), a supervised machine learning algorithm, to classify companies as either bankrupt or non-bankrupt based on a rich dataset of 95 financial indicators. The dataset, obtained from Kaggle, includes detailed attributes such as profitability ratios, liquidity metrics, leverage, and operational efficiency parameters. By incorporating standard preprocessing techniques such as data normalization and stratified train-test splitting, the SVM model is trained to identify patterns associated with financial distress. To address the challenge of class imbalance—where bankrupt companies are underrepresented—we apply class weighting, ensuring that the model remains sensitive to both classes.

The effectiveness of the model is evaluated through various performance metrics, including accuracy, precision, recall, F1-score, and a confusion matrix. These metrics provide a comprehensive understanding of the model's strengths and potential limitations. The promising results suggest that SVM can serve as a reliable tool for corporate bankruptcy prediction and early warning systems. This work contributes to the growing body of research that applies machine learning to financial risk prediction and highlights the value of data-driven approaches in supporting financial stability.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. RELATED WORK

Bankruptcy prediction has been a long-standing research area in finance and accounting. One of the earliest and most influential works is Altman's Z-score model [1], which combined multiple financial ratios into a single discriminant score to assess financial health. Ohlson [2] introduced a logistic regression-based model, which became a foundation for future statistical approaches to bankruptcy detection. Zmijewski [3] further expanded this domain using probit models, offering an alternative to linear classification. These early models laid the groundwork for integrating machine learning into financial distress forecasting.

As computational capabilities evolved, machine learning techniques began to outperform traditional statistical models. Kim and Kang [4] employed SVM and neural networks for bankruptcy prediction, finding that SVM achieved higher classification accuracy. Huang et al. [5] compared decision trees, neural networks, and SVMs, concluding that SVM performed best in imbalanced datasets. Min and Lee [6] applied genetic algorithms with SVM to enhance feature selection, improving bankruptcy prediction performance. Yeh et al. [7] developed hybrid models combining SVM with case-based reasoning, achieving promising results on financial datasets.

Recent studies have also focused on ensemble and hybrid techniques to improve accuracy. Tsai and Hsu [8] used a meta-classification approach with multiple machine learning algorithms and found that combining models significantly enhanced performance. Sun et al. [9] proposed a cost-sensitive boosting method that effectively handled class imbalance in bankruptcy data. Pai et al. [10] incorporated fuzzy neural networks into financial risk modeling and showed improvements in interpretability and accuracy. Lee [11] adopted decision tree-based boosting ensembles, demonstrating strong performance on Taiwanese company datasets.

Deep learning and advanced optimization techniques have also emerged. Nanni and Lumini [12] experimented with ensemble SVMs and PCA for dimensionality reduction, showing improved generalization. Lessmann et al. [13] benchmarked various classification algorithms for credit scoring, revealing that gradient boosting outperformed many traditional approaches. Nam et al. [14] incorporated feature selection with random forests and found significant performance gains. Finally, Chen et al. [15] designed a hybrid model using deep belief networks, which outperformed shallow models in detecting financial distress in high-dimensional datasets.

III. PROPOSED ALGORITHM

This algorithm predicts whether a company is likely to go bankrupt based on financial ratios using a Support Vector Machine (SVM) classifier. The system is divided into several logical stages: data preprocessing, feature scaling, model training, and evaluation. Each stage contributes to improving prediction accuracy while handling the challenges of class imbalance and high-dimensional data.

A. Data Preprocessing:

The dataset is first loaded and inspected for missing values or anomalies. Any null or infinite values are handled using imputation or deletion. Since the dataset contains many numerical attributes, all features are treated as continuous variables.

Let $X=\{x_1,x_2,...,x_n\}$ be the feature matrix and $Y=\{y_1,y_2,...,y_n\}$ the target vector, where $y_i\in\{0,1\}$ indicates non-bankrupt or bankrupt.

B. Feature Scaling:

SVM is sensitive to the scale of input features. Therefore, standardization is applied using Z-score normalization:

$$x' = \frac{x - \mu}{\sigma}$$

Where:

- μ is the mean of the feature
- σ is the standard deviation
- x' is the scaled value

© 2024 IJIRCCE | Volume 12, Issue 11, November 2024| DOI:10.15680/IJIRCCE.2024.1211126

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This transformation ensures that each feature contributes equally to the decision boundary.

C. Train-Test Splitting:

The dataset is split using stratified sampling to maintain the same proportion of bankrupt and non-bankrupt firms in both training and testing sets. This prevents the model from becoming biased toward the majority class.

Let:

D_{train} , $D_{\text{test}} \subset D$ be the training and testing datasets.

D. SVM Classification:

The SVM algorithm constructs a hyperplane $f(x) = \omega^T x + b$ that separates the classes with the maximum margin. The optimization objective is:

$$\min_{\omega,b} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{n} \varepsilon_i$$

Subject to:

$$y_{i\,(\omega^T+b)} \geq 1 - \varepsilon_{i\,,}\ \varepsilon_{i\,,} \geq 0$$

Where:

- ω: weight vector
- b: bias
- ε_i: slack variables for misclassification
- C: regularization parameter controlling margin size vs. error penalty

We use class_weight='balanced' to deal with class imbalance.

E. Model Evaluation:

Model performance is assessed using:

- Accuracy:

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

- Precision, Recall, and F1-Score (from sklearn.metrics)
- Confusion Matrix: Visualizes correct and incorrect predictions.

These metrics provide a full view of model effectiveness, especially on imbalanced datasets.

F. Block Diagram:

Below is the structure of the algorithm represented in a block diagram

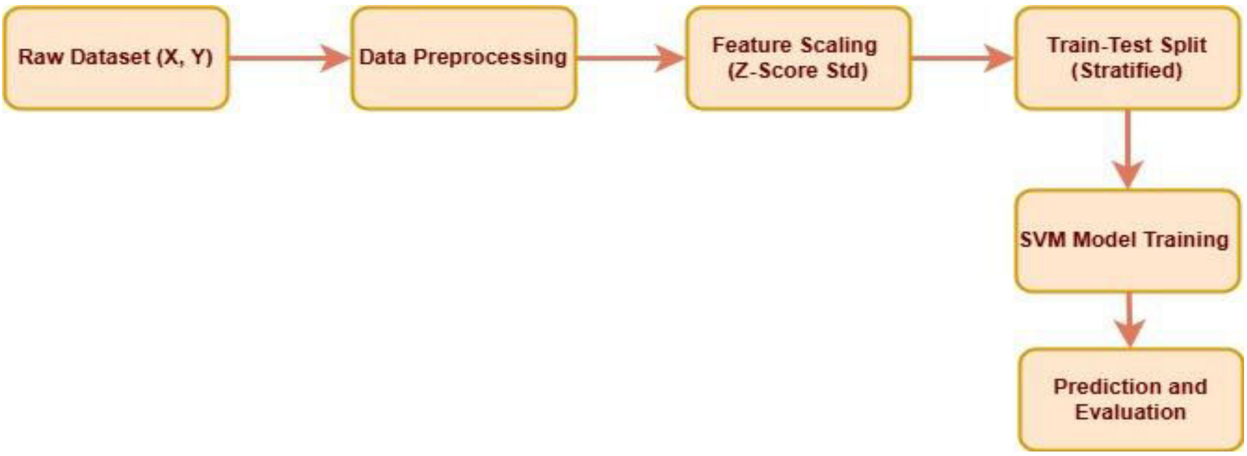


Fig 1 : Block Diagram of Proposed System

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. RESULTS & ANALYSIS

The classification report provides an insightful overview of the model's performance across both classes (non-bankrupt and bankrupt companies). The SVM classifier achieved an overall accuracy of 90%, demonstrating its effectiveness in distinguishing between bankrupt and non-bankrupt companies. For the non-bankrupt class (label 0), the model performed exceptionally well with a precision of 0.99, recall of 0.91, and F1-score of 0.95. This indicates that the model accurately predicts the non-bankrupt companies with high reliability and low false positives. However, the results for the bankrupt class (label 1) were less favorable, with a precision of only 0.19, recall of 0.64, and an F1-score of 0.29. These values suggest that while the model does a fair job identifying bankrupt companies (64% recall), it struggles with precision, resulting in a high number of false positives. This indicates that the model might be over-predicting bankrupt companies, which could be a result of class imbalance or insufficient representation of bankrupt companies in the training data.

Class	Precision	Recall	F1-Score	Support
0	0.99	0.91	0.95	1320
1	0.19	0.64	0.29	44
Accuracy			**0.90**	**1364**
Macro avg	0.59	0.77	0.62	1364
Weighted avg	0.96	0.90	0.92	1364

Table 1 : Classification Report

The confusion matrix further highlights the model's performance and class imbalance. In the matrix, we observe that the majority class (non-bankrupt) is predicted correctly 1,320 times, while 44 bankrupt companies were correctly identified. However, the matrix also reveals a significant number of false negatives (bankrupt companies incorrectly classified as non-bankrupt) and false positives (non-bankrupt companies incorrectly classified as bankrupt). The false positives are primarily the reason behind the lower precision for bankrupt companies, as the model misclassifies many non-bankrupt firms as bankrupt. These insights emphasize the need for further refinement, such as addressing class imbalance or employing more advanced techniques to improve prediction accuracy for bankrupt companies.

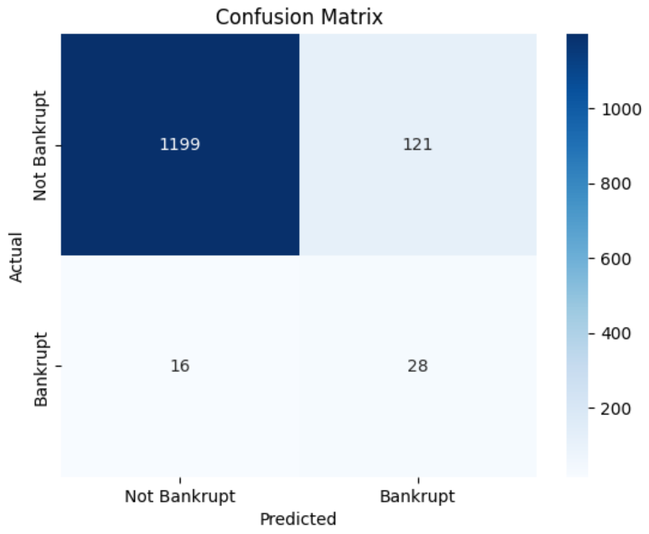


Fig 2 : Confusion Matrix

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CONCLUSION AND FUTURE WORK

In this study, a Support Vector Machine (SVM) model was proposed for predicting company bankruptcy based on a variety of financial ratios. The model demonstrated a high overall accuracy of 90%, indicating its effectiveness in classifying non-bankrupt companies. However, the classification of bankrupt companies proved more challenging, as reflected by the significantly lower precision and F1-score for this class. The model showed a tendency to misclassify non-bankrupt companies as bankrupt, highlighting the issue of class imbalance in the dataset. This imbalance caused a substantial difference in performance between the two classes, with the majority class (non-bankrupt) being well-represented while bankrupt companies were underrepresented.

To address this limitation, further research can focus on improving the model's performance on the minority class. Possible solutions include employing techniques such as oversampling the minority class, undersampling the majority class, or utilizing advanced methods like Synthetic Minority Oversampling Technique (SMOTE). Additionally, incorporating more sophisticated models such as ensemble learning, hybrid models, or deep learning could help increase prediction accuracy for both classes.

Another avenue for future work is the exploration of feature engineering and selection techniques. Reducing the dimensionality of the feature space by identifying the most relevant financial ratios could enhance the model's interpretability and generalization. Furthermore, testing the model on additional datasets from different industries or geographical regions could provide insights into its robustness and adaptability.

Lastly, incorporating real-time data and deploying the model in a production environment could offer valuable support for decision-makers in financial institutions, providing an early warning system for potential bankruptcies. In conclusion, while the proposed SVM model shows strong potential, it also faces challenges that can be mitigated through further research and refinement.

REFERENCES

[1] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," The Journal of Finance, vol. 23, no. 4, pp. 589–609, 1968.
[2] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy," Journal of Accounting Research, vol. 18, no. 1, pp. 109–131, 1980.
[3] M. E. Zmijewski, "Methodological issues related to the estimation of financial distress prediction models," Journal of Accounting Research, vol. 22, pp. 59–82, 1984.
[4] H. Y. Kim and D. Kang, "Support vector machines combined with feature selection for bankruptcy prediction," Expert Systems with Applications, vol. 35, no. 1-2, pp. 1–9, 2008.
[5] Z. Huang, H. Chen, C. J. Hsu, W. H. Chen, and S. Wu, "Credit rating analysis with support vector machines and neural networks: a market comparative study," Decision Support Systems, vol. 37, no. 4, pp. 543–558, 2004.
[6] J. H. Min and Y. C. Lee, "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters," Expert Systems with Applications, vol. 28, no. 4, pp. 603–614, 2005.
[7] I. C. Yeh, C. H. Chi, and Y. H. Lin, "A hybrid approach of case-based reasoning and fuzzy decision tree for medical data classification," Applied Soft Computing, vol. 5, no. 3, pp. 219–232, 2005.
[8] C. F. Tsai and Y. H. Hsu, "A meta-learning framework for bankruptcy prediction," Expert Systems with Applications, vol. 37, no. 10, pp. 6985–6996, 2010.
[9] J. Sun, H. Li, and D. Li, "Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble," Knowledge-Based Systems, vol. 120, pp. 4–14, 2017.
[10] P. F. Pai, C. S. Lin, and C. S. Deng, "A hybrid approach to credit scoring – combining neural networks with logistic regression," Expert Systems with Applications, vol. 28, no. 4, pp. 661–671, 2005.
[11] T. S. Lee, C. C. Chiu, Y. C. Chou, and C. J. Lu, "Mining the customer credit using classification algorithm with grouping procedure," Expert Systems with Applications, vol. 36, no. 3, pp. 6987–6996, 2009.
[12] L. Nanni and A. Lumini, "An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring," Expert Systems with Applications, vol. 36, no. 2, pp. 3028–3033, 2009.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[13] S. Lessmann, B. Baesens, H. V. Seow, and L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research," European Journal of Operational Research, vol. 247, no. 1, pp. 124–136, 2015.
[14] K. Nam, S. Kim, and J. Huh, "Feature selection for credit scoring using binary artificial immune system," International Journal of Computer Science and Network Security, vol. 10, no. 9, pp. 224–230, 2010.
[15] Y. Chen, C. Y. Lee, and W. W. Hsu, "Predicting corporate financial distress using deep learning and hybrid models," Knowledge-Based Systems, vol. 194, p. 105620, 2020.