Machine Learning and XAI for Predicting Corporate Bankruptcy

Nourhan Deif, Radwa Badran, Rahma Mourad, Rihana Nasr

Department of Computer Science, Zewail City of Science and Technology

Giza, Egypt

nourhan.sayed@zewailcity.edu.eg

radwa.badran@zewailcity.edu.eg

Rahma.sayed@zewailcity.edu.eg

Rihana.Nasr@zewailcity.edu.eg

Abstract

In today's fast-moving financial landscape, catching the first signs of corporate distress can spare investors and managers from costly surprises. We analyzed ten years of data from 6,819 Taiwanese firms—each described by 96 financial ratios—to train and compare six models: Logistic Regression, Support Vector Machines (SVM), Random Forest, AdaBoost (with Random Forest base learners), XGBoost, and a fully connected Neural Network. To make sure those rare bankruptcy cases weren't drowned out, we applied SMOTE for class balancing. We then used Explainable AI methods SHAP, LIME, Partial Dependence and ICE plots, plus surrogate decision trees to pinpoint exactly which financial metrics steer each model's decision. Our best models hit over 96% accuracy, and, crucially, they explain their reasoning in human-readable form. This blend of strong performance and clear explanations gives stakeholders the confidence to understand and act on risk signals before it's too late.

Keywords— bankruptcy prediction; machine learning; neural networks; explainable AI; SMOTE; financial ratios; Random Forest; AdaBoost; XGBoost; SHAP.

I. Introduction

Predicting corporate bankruptcy is a critical task with significant value for investors, banks, and businesses. Accurate predictions of bankruptcy enable companies to take early corrective actions to prevent failure, while banks can make more informed lending decisions to mitigate financial risks. The ability to predict whether a company will go bankrupt based on financial indicators can thus have far-reaching consequences for economic stability and decision-making processes.

In this study, we aim to develop a machine learning model that classifies companies as either bankrupt or not bankrupt. The model will be built using a dataset containing financial indicators of various companies, which will allow us to leverage machine learning algorithms to make reliable predictions.

The dataset used for this project consists of 6,819 samples (representing individual companies) and 96 financial features. The data spans from 1999 to 2009 and is sourced from the Taiwan Economic Journal, available on Kaggle. The target variable in this dataset is binary classification, where companies are labeled as either bankrupt (1) or not bankrupt (0).

Our approach will involve training multiple machine learning models, including Random Forest, Support Vector Machines (SVM), and XGBoost, to evaluate their performance in predicting bankruptcy. In addition to the models, we will apply Explainable AI (XAI) techniques such as LIME, SHAP, PDP, ICE, and surrogate models to provide transparency and interpretability to our predictions.

This research aims to explore the effectiveness of machine learning models in predicting corporate bankruptcy and assess how XAI methods can enhance the understanding of model decisions. The findings from this study will help businesses and financial institutions make better-informed decisions, potentially reducing financial risks and losses.

By the end of this study, we aim to answer several research questions, including: How accurate are the selected machine learning models in predicting bankruptcy? How can XAI techniques improve the interpretability of these models, and what insights can be drawn from the importance of features in bankruptcy prediction?

II. Related Work

Predicting bankruptcy means using financial data and information to estimate whether a company is likely to fail in the future. This is done by analyzing factors such as profits, debts, cash flows, and other financial ratios. Companies, investors, and banks use bankruptcy prediction to avoid risks and make better financial decisions.

In 2019, Yi Qu et al.[1] Reviewed how machine learning and deep learning techniques are being used to predict bankruptcy depending on analyzing financial statements, review several representative machine learning techniques used in bankruptcy prediction, such as Multivariate Discriminant Analysis (MDA), Logistic Regression (LR), Ensemble method and the well-known Neural Networks (NN), Support Vector Machines (SVM). The researchers analyzed the advantages and limitations of these models in bankruptcy prediction.

In 2022 Shetty et al.[2].Research has explored Bankruptcy prediction from statistical models like Z-score(Altman, 1968) and logit models(Ohlson,1980) to advanced machine learning techniques including extreme gradient boosting(XGBoost), support vector machine(SVM) and using deep Neural networks using Belgian Small and Medium Enterprises (SME) data achieving 82–83% accuracy using only three easily obtainable financial ratios(Bredart's,2014).

In 2021, Kim, Cho, and Ryu[3] utilized the evolution of traditional statistical models, such as logistic regression on bankruptcy prediction such including logistic regression, random forests, and artificial neural networks(ANNs), However, these models often fail to consider the sequential nature of financial distress. Deep learning methods, particularly RNNs and LSTMs, have demonstrated superior performance in analyzing time-series financial data, leading to more accurate bankruptcy predictions.

Alanis et al. [4] tried and rated how well different machine learning models can guess when a company will go bankrupt in 2022. The team tested data from 2,585 publicly traded U.S. companies covering 1990 to 2019. The models they used included gradient-boosted trees, random forests, and penalized logistic regression. Key variables such as excess stock returns, idiosyncratic risk, and company size played the most important roles in determining outcomes. Interestingly, the addition of textual data from corporate filings did not substantially improve the prediction accuracy (Alanis et al., 2022).

Arno et al.[5] released the ECL dataset in 2024. It takes textual and numerical data from 10-K files and connects it with binary bankruptcy labels. Using this dataset, they experimented with both conventional and neural models and found that combining different data sources increased prediction accuracy. Additionally, they investigated the role of large language models in this area, showing that while GPT-based models could produce useful text summaries, their zero-shot bankruptcy predictions were not as reliable (Arno et al., 2024).

A different study, titled "Performance Comparison of Multiple Discriminant Analysis and Logit Models in Bankruptcy Prediction" [6] accessible through ProQuest, analyzed the predictive performance of Multiple Discriminant Analysis (MDA) and Logistic Regression (Logit) models. The researchers found that Logit models outperformed MDA after analyzing financial data from multiple companies, particularly when handling nonlinear financial relationships. The study highlighted the reliability and accuracy of Logit models and suggested using them to make more accurate forecasts of bankruptcy (ProQuest, n.d.).

In [7], this research builds from a foundation of bankruptcy prediction studies, transitioning from traditional statistical models like Altman's Z-score and Ohlson's logistic regression to modern machine learning methods. It confesses the success of decision trees, neural networks, and ensemble methods, particularly random forests, as shown by Barboza et al. This study differentiates itself by poretelling the firm default, not just bankruptcy, and using a novel, granular credit dataset. It aims to demonstrate improved prediction accuracy using machine learning on this unique dataset, combining loan and balance sheet data.

In [8], the study explains the existing research in bankruptcy prediction, exploring multiple machine learning methods including logistic regression, neural networks, and deep learning. Prior work has sought to extract decision rules from qualitative data and address issues like overfitting in neural network models. Also, this study produces some methodologies to influence modern tools like Scikit-learn and explore deep architectures, while acknowledging the limitations of traditional techniques such as logistic regression. The foundation is built through previous datasets and algorithms, trying to improve prediction accuracy through advanced computational techniques.

In [9], this research reviews the existing bankruptcy prediction models, contrasting older methods like logistic regression, genetic algorithms, and decision trees, which have limitations in accuracy and stability. It then explores new machine learning models, such as support vector machines, neural networks with dropout, and autoencoders, which are less explored in this financial context. The research aims to expound on the improved accuracy of these newer models in predicting bankruptcy and to support comparative analysis. It will also analyze the shortcomings of the newer models.

Narvekar and Guha (2021b) [10] examined bankruptcy prediction using machine learning techniques, algorithms such as Random Forest, SVM, and XGBoost were applied to financial data, and scaling and imputing missing values were applied, and due to an imbalance in the balance of categories, oversampling techniques such as SMOTE and ADASYN were used, and the data was divided into 70% for training and 30% for testing, and the accuracy of the models was evaluated through accuracy and AUC metrics, and XGBoost achieved the best performance and its accuracy was about 99%, and the results confirm that learning models The machine is better than the traditional statistical methods.

Nguyen et al. (2023) [11] evaluate the effectiveness of machine learning models in predicting corporate bankruptcies, including Random Forest, LightGBM, XGBoost, and NGBoost, with a prediction horizon of one to five years. These models have achieved high accuracy, but they are characterized by a black-box, which limits their interpretability, and to address this problem, Shapley Additive Explanations (SHAP) to XGBoost was used to illustrate how each financial ratio affects the prediction of bankruptcy. This study analyzes financial ratios and their impact on forecasting in depth and how the results affect credit risk.

Michalkova and Ponisciakova (2025b) [12] examine the ability of five models to predict bankruptcy, and these models are Kliestik et al., Poznanski, modified Zmijewski, Jakubik-Teply, and Virag-Hajdu across different stages of the corporate life cycle. The performance of the models was evaluated using accuracy, balanced accuracy, F1-score, F2-score, and MCC. The results showed that the Kliestik et al. model was the most reliable, in contrast to models such as Poznanski and Virag-Hajdu showed unstable results and the modified Zmijewski model was the best for companies in transition.

III. Methodology

This section explains the steps followed in preparing and analyzing the dataset for bankruptcy prediction. The methodology includes data description, cleaning, exploratory analysis, and correlation study to support model building and explainability.

III.1 Dataset Description

The dataset used for bankruptcy prediction consists of 6,819 instances and 96 features, each representing various financial ratios and indicators relevant to a company's performance. These features encompass key financial dimensions such as profitability, liquidity, debt management, and asset efficiency, enabling a comprehensive analysis of a firm's financial health. Importantly, the dataset is complete with no missing values, ensuring robust input for machine learning models without requiring imputation. The target variable, labeled 'Bankrupt?', is binary: a value of 1 indicates a bankrupt company, while 0 denotes a solvent one. This structure allows for clear classification tasks and supports the application of supervised learning techniques aimed at early bankruptcy prediction

III.2 Data Preprocessing

The dataset underwent essential preprocessing steps to ensure data quality and suitability for modeling. No duplicate entries were found, so no removal was needed. All numerical features were standardized using StandardScaler, which transforms the data to have a mean of 0 and a standard deviation of 1, ensuring consistency in scale across features. This step is crucial for models sensitive to feature magnitude. Finally, the dataset was split into 80% training and 20% testing sets to evaluate model performance on unseen data effectively.

III.5 Model Selection and Training

SVM

A Support Vector Machine (SVM) with an RBF kernel and balanced class weights was selected due to its strong performance in previous financial prediction studies [2]. The model was trained on the preprocessed data and evaluated on the test set.

Logistic Regression Model

Logistic Regression was also implemented as a baseline due to its simplicity and wide adoption in binary classification problems. It was trained on the same preprocessed dataset both before and after feature selection.

XGBoost (Extreme Gradient Boosting)

XGBoost is a widely used, powerful, and scalable implementation of gradient boosting that builds decision trees sequentially to minimize prediction errors, making it highly effective for structured data and competition-winning models. It is particularly known for its efficiency and scalability in large datasets, making it a strong candidate for bankruptcy prediction tasks.

Random Forest Model

Random Forest is a powerful ensemble classification model that builds multiple decision trees and aggregates their predictions to make a final decision. It is well-suited for handling high-dimensional datasets, reduces overfitting, and performs well on imbalanced classification tasks like bankruptcy prediction.

III.6 Feature Selection

To enhance interpretability and potentially improve performance, Recursive Feature Elimination (RFE) was used for both models. RFE recursively removes the least important features based on the trained model until the top 30 features remain.

SVM

The selected features for SVM include key profitability metrics (e.g., ROA), turnover ratios, net value per share, and cash flow indicators. The model was retrained using only these top 30 features.

Logistic Regression

The selected features for Logistic Regression include indicators such as Operating Profit Rate, Cash Flow Rate, Revenue Per Share, Net Value Per Share, and Debt Ratios. The model was also retrained using these 30 selected features.

III. 7 Model Evaluation Techniques

Several evaluation techniques were applied to assess both models:

- **Confusion Matrix**: Shows the correct and incorrect predictions.
- **Classification Report**: Provides precision, recall, and F1-score.
- **ROC Curve and AUC**: Both models achieved an AUC close to 1, indicating strong class separability.
- **Learning Curve**: Evaluates model learning behavior over different training sizes.
- **Validation Curve**: Explores the effect of the regularization parameter C on model performance in the SVM model.

III.8 Model Explainability

To interpret the model's predictions:**Permutation Importance** was applied to rank features based on the impact of their random shuffling on model accuracy.

- **SHAP (SHapley Additive Explanations)**

SHAP values were computed to show how much each feature contributes to predictions. Summary plots provided a global view of feature impact and directionality for both models.

- **LIME**

  LIME was used to provide local explanations for individual predictions. It approximated the complex model locally with a simple interpretable model.

- **PDP and ICE**

  Partial Dependence Plots (PDP) and Individual Conditional Expectation (ICE) plots were generated for the top features to visualize their marginal and individual effects.

- **Surrogate Decision Tree**

  A decision tree surrogate model was trained to mimic the Logistic Regression model's behavior. It improved interpretability by simplifying complex decision boundaries into hierarchical rules.

IV. Results

IV.1 Logistic Regression Model

Logistic Regression Model (All Features):

- The model achieved an accuracy of 96.63% on the test set.

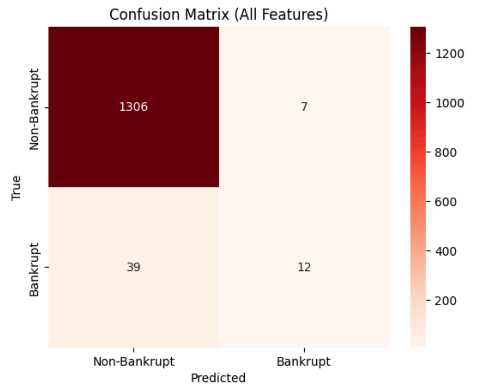
- **Classification Report:**

Class	Precision	Recall	F1-Score
0 (Non-Bankrupt)	0.97	0.99	0.98
1 (Bankrupt)	0.63	0.24	0.34

Table. Classification Report forLogistic Regression Model (All Features)

Confusion Matrix:

- True Positives (TP): 1306 bankrupt companies correctly identified.
- True Negatives (TN): 12 non-bankrupt companies correctly identified.
- False Positives (FP): 7 non-bankrupt companies incorrectly identified as bankrupt.
- False Negatives (FN): 39 bankrupt companies incorrectly identified as non-bankrupt.

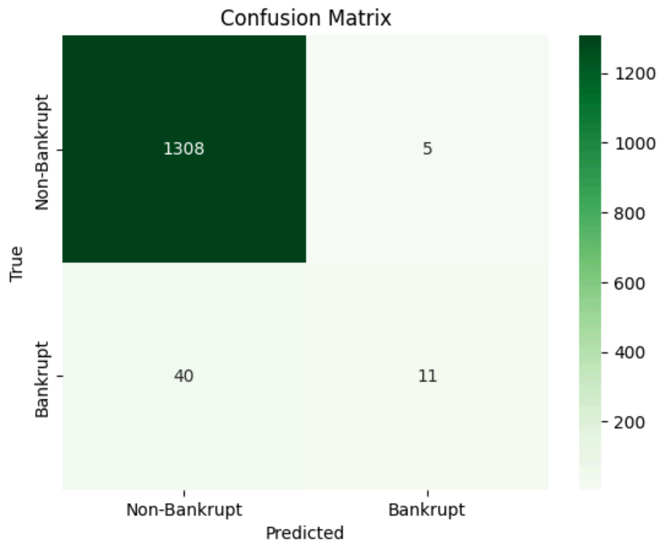


Fig[1]. The Confusion matrix for Logistic Regression Model (All Features).

Logistic Regression with Feature Selection (RFE):

- After applying Recursive Feature Elimination (RFE), which selected 30 features, the model achieved an accuracy of 96.70% on the test set.

- **Confusion Matrix (with selected features):**
- **True Positives (TP)**: 1308 bankrupt companies correctly identified.
- **True Negatives (TN)**: 11 non-bankrupt companies correctly identified.
- **False Positives (FP)**: 5 non-bankrupt companies incorrectly flagged as bankrupt.
- **False Negatives (FN)**: 40 bankrupt companies incorrectly identified as non-bankrupt.



- that the model is overfitting. This behavior highlights the model's strong performance on training data but a reduced ability to generalize to new, unseen data. To improve model generalization, techniques such as regularization may be necessary.

Interpretability Techniques:

1. **LIME**: It showed that even with some negative signs like high debt, good profits and cash flow made the model predict the company is not bankrupt with 99% certainty.
2. **Partial Dependence Plot (PDP)**:

- **Operating Profit Rate**: As the operating profit increases, the chance of a company being **non-bankrupt** increases.
- **Non-industry income**: Higher non-industry income suggests a higher risk of bankruptcy.

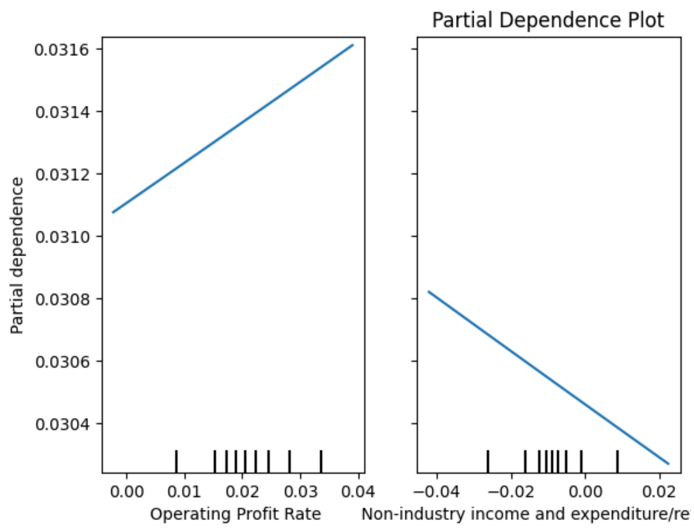
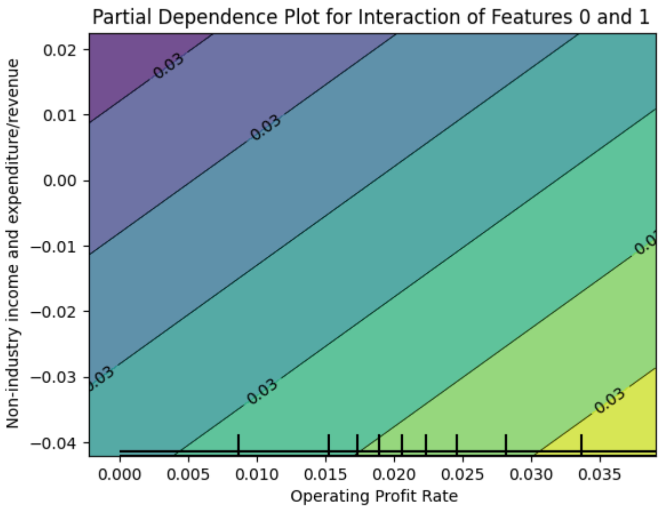


Fig. Partial Dependence Plot for Operating Profit Rate and Non-industry income.



3. **Individual Conditional Expectation (ICE)**:

- Showed that changes in features like **Operating Profit Rate and** Non-industry income had different effects on predictions for individual data points.

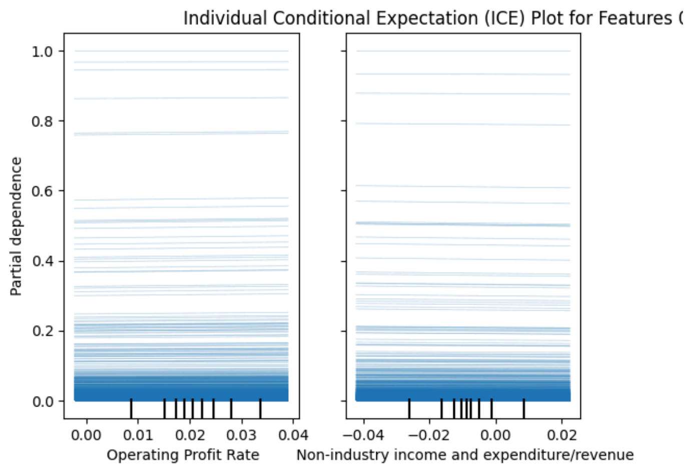


Fig. ICE Plot for Operating Profit Rate and Non-industry income.

4. SHAP:

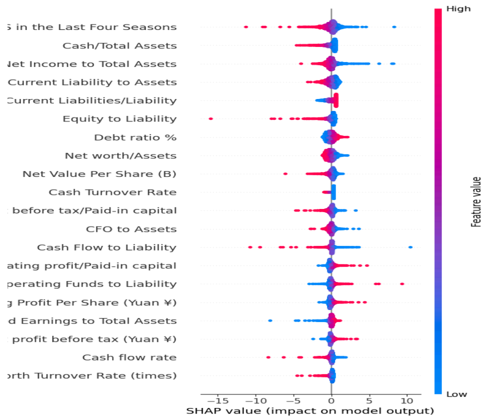


Fig. The SHAP summary plot showed that .Persistent EPS in the Last Four Seasons, Cash/Total Assets, and Net Income to Total Assets were the most important features contributing to the predictions

5. **Permutation Importance**

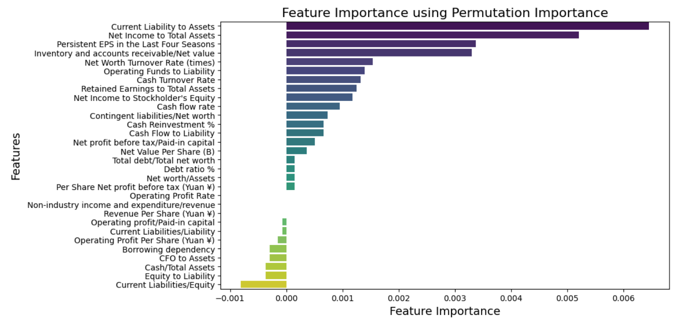


Fig . **Permutation Importance** Identified features such as **Current Liability to Assets** and **Net Income to Total Assets** as crucial for the model's decision-making.

6. **Surrogate Decision Tree**:

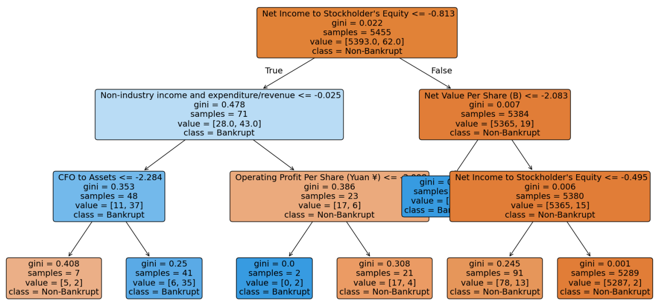


Fig. Trained a decision tree to approximate the logistic regression model, revealing the importance of features like **Net Income to Stockholders Equity** and **Non-industry income and expenditure**.

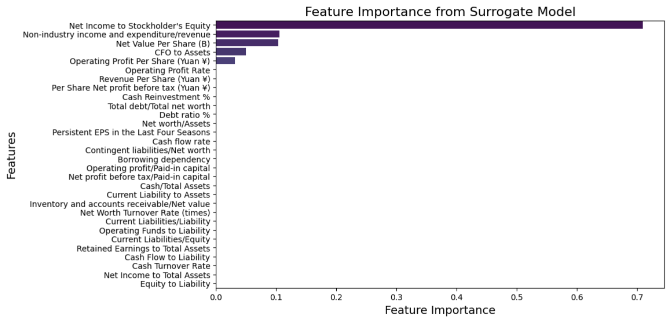


Fig. Feature Importance from Surrogate Model

- **ROC Curve and AUC:**

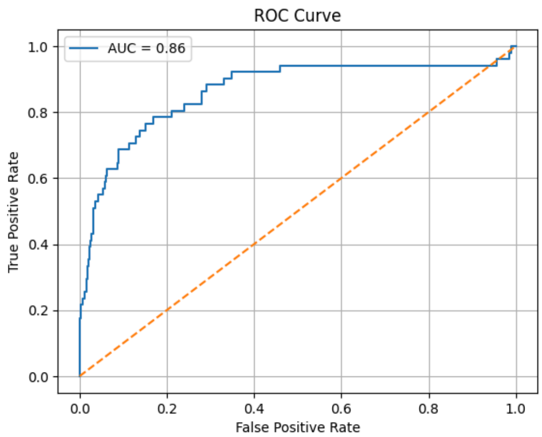


Fig. AUC close to 1: The model is good at telling the difference between positive and negative cases.

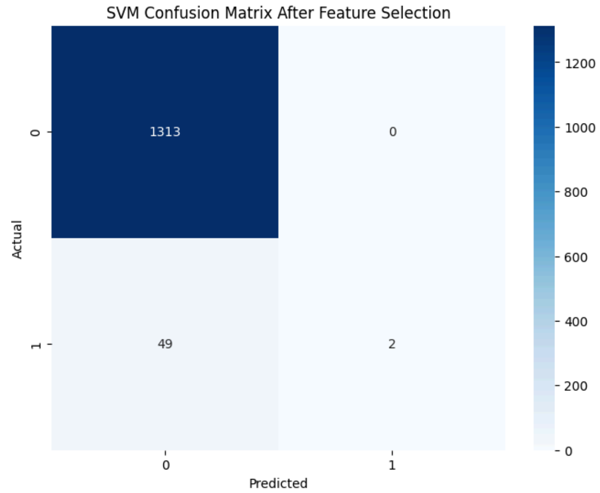


Fig. The Confusion matrix is used to compare the predicted label with the true label.

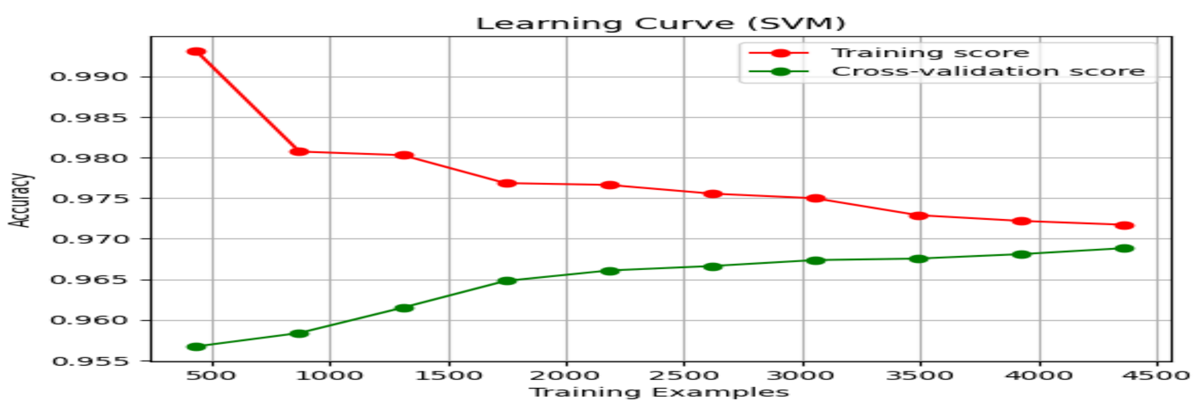
Classification Report:

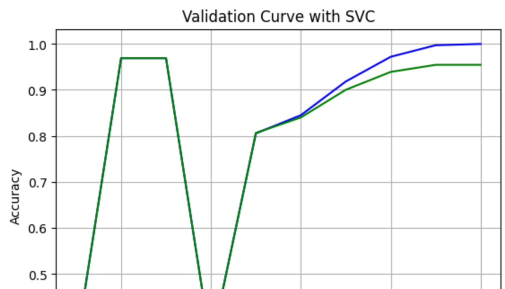
Table. Classification Report for SVM with Top 30 features

IV.2 SVM Model

Class	Precision	Recall	F1-Score
0 (Non-Bankrupt)	0.96	1.00	0.98
1 (Bankrupt)	1.00	0.04	0.08
Accuracy			0.96

Learning Curve: A learning curve shows how well a model is



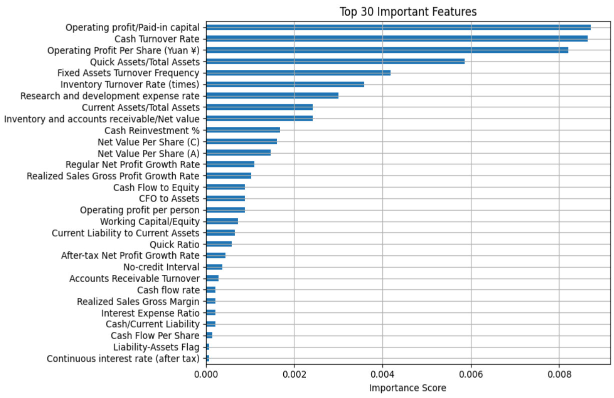


- **Validation Curve:**

III.8 Model Explainability

To interpret the model's predictions:

- **Permutation Importance** was applied to rank features based on the impact of their



random shuffling on model accuracy.

Fig, **Permutation importance** tells us how much each feature is important in the model.

- **SHAP (SHapley Additive Explanations)** was used to understand the contribution of each feature to individual predictions.

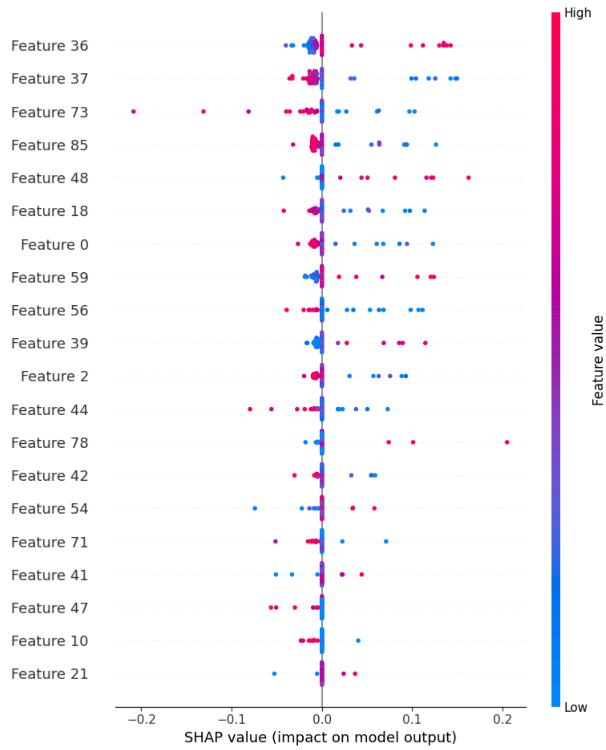


Fig. SHAP plot to provide a global feature importance overview

IV.4 XGBoost Model

XGBoost Model (All Features):

- The model achieved an accuracy of 96% on the test set.

- **Classification Report:**

Table. Classification Report for XGBoost Model

Confusion Matrix:

- True Positives (TP): 1302 bankrupt companies correctly identified.
- True Negatives (TN): 13 non-bankrupt companies correctly identified.

- False Positives (FP): 11 non-bankrupt companies incorrectly identified as bankrupt.

- False Negatives (FN): 38 bankrupt companies incorrectly identified as non-bankrupt.

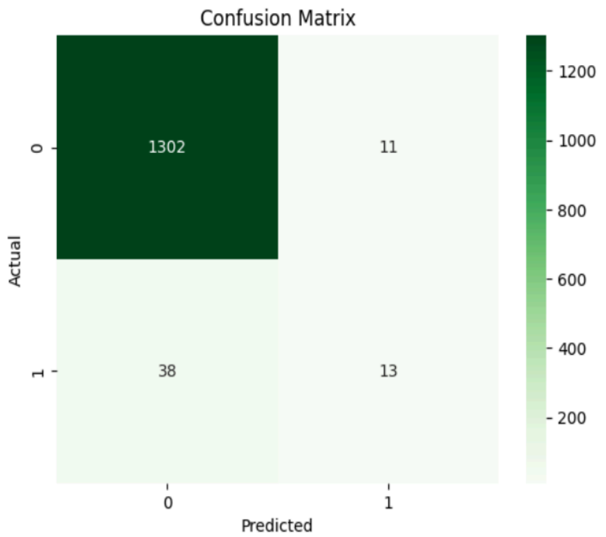


Fig. The Confusion matrix for the XGBoost Model (All Features).

XGBoost with Feature Selection (RFE):

- After applying Recursive Feature Elimination (RFE), which selected 10 features, the model achieved an accuracy of 96.2% on the test set.

Confusion Matrix (with selected features):

- **True Positives (TP)**: 1308 bankrupt companies correctly identified.

- **True Negatives (TN)**: 11 non-bankrupt companies correctly identified.

- **False Positives (FP)**: 5 non-bankrupt companies incorrectly flagged as bankrupt.

- **False Negatives (FN)**: 40 bankrupt companies incorrectly identified as non-bankrupt.

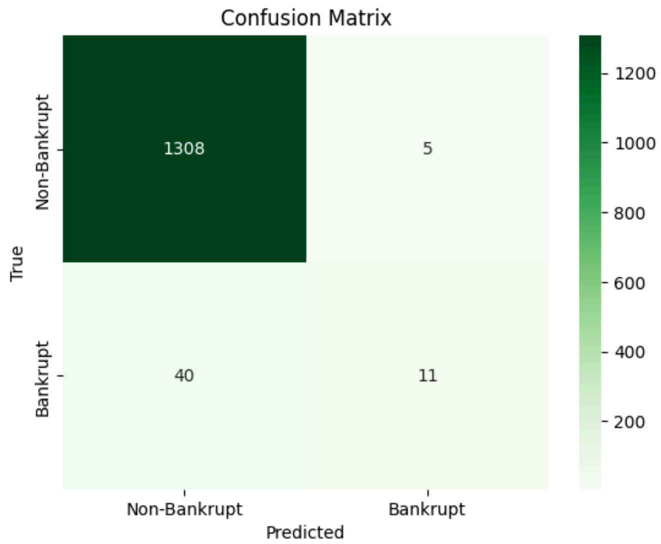


Fig. The Confusion matrix for XGBoost with Feature Selection (RFE).

ROC Curve and AUC:

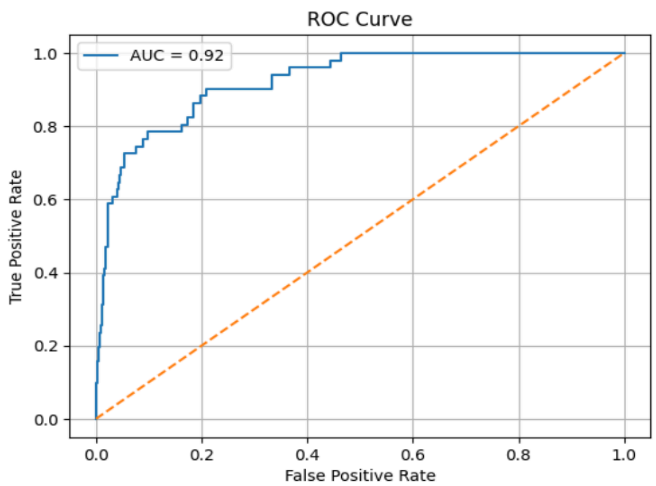
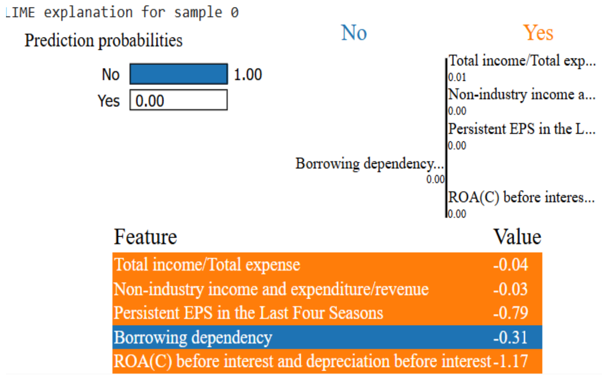


Fig. AUC almost 1. The model is good at telling the difference between positive and negative cases.

Interpretability Techniques:

1. **LIME**: It showed that even with some negative signs like persistent EPS, and ROA(C) made the model predicted the company is not bankrupt with 100% certainty.



2. **Partial Dependence Plot (PDP)**:
   - **Operating Profit Rate**: As the operating profit decreases, the company's chance of being **non-bankrupt** increases.
   - **ROA(B) before interest**: A Higher ROA(B) suggests a higher risk of bankruptcy.

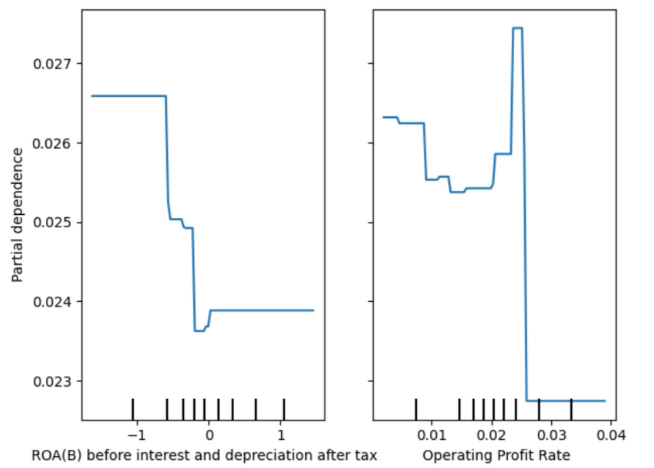


Fig. Partial Dependence Plot for ROA(B) before interest and Operating profit rate.

- 3. **Individual Conditional Expectation (ICE)**:

  Showed that changes in features like **Operating Profit Rate and ROA(B) before interest** had different effects on predictions for individual data points.

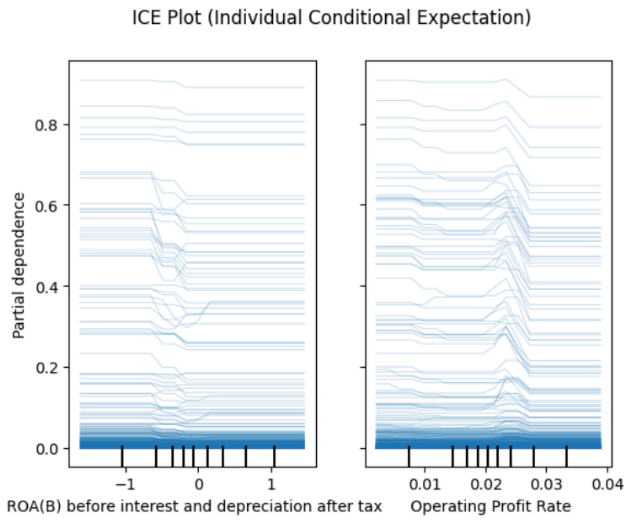


Fig. ICE Plot for Operating Profit Rate and ROA(B) before interest.

4. SHAP:

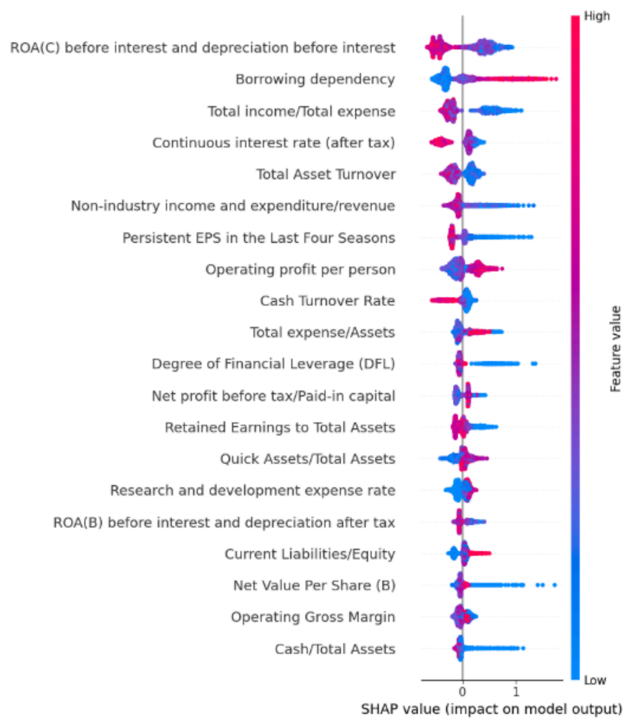


Fig. The SHAP summary plot showed that ROA(C) before interest, borrowing dependency, and total income/total expense were the most important features contributing to the predictions.

5. **Permutation Importance**

Fig. **Permutation Importance identifies** features such as **Persistent EPS** and **Working capital/Equity** as crucial for the model's decision-making.

IV.4 SVM Model

- The model achieved an accuracy of 97% on the test set.

Class	Precision	Recall	F1-Score
0 (Non-Bankrupt)	0.97	0.99	0.98
1 (Bankrupt)	0.50	0.24	0.32
Accuracy			96.0

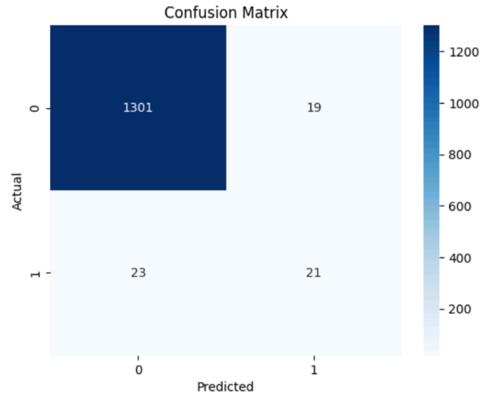


Fig. The Confusion matrix for the SVM Model (All Features).

ROC Curve and AUC:

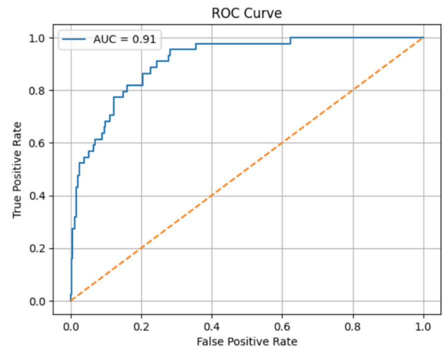


Fig. AUC almost 1. The model is good at telling the difference between positive and negative cases.

IV.5 Neural Network Model

Neural Network Model (Selected Features):
The model achieved an accuracy of **85%** on the test set.

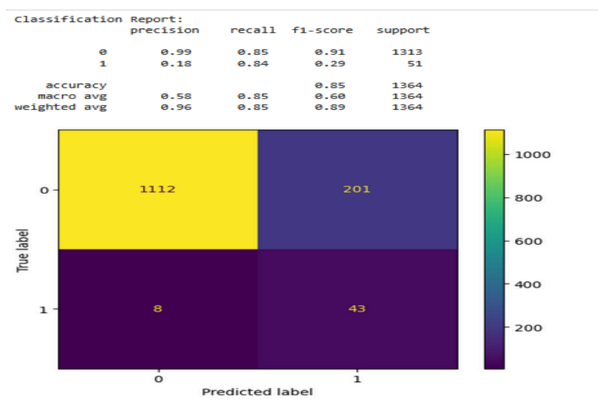


Table. Classification Report for Neural Network Model (Selected Features)

Interpretability Techniques:

1. Permutation Importance

```
 Permutation Importance (mean decrease in F1):
Net Income to Total Assets                                  0.347082
ROA(C) before interest and depreciation before interest     0.324569
Net Value Per Share (B)                                     0.320366
Persistent EPS in the Last Four Seasons                     0.274655
ROA(A) before interest and % after tax                      0.268201
After-tax Net Profit Growth Rate                            0.267647
ROA(B) before interest and depreciation after tax           0.246533
Regular Net Profit Growth Rate                              0.174684
Total debt/Total net worth                                  0.141290
Net Value Per Share (C)                                     0.140326
dtype: float64
```

Fig. **Permutation Importance identifies** features such as **Net income, ROA(C), and Net value per share** as crucial for the model's decision-making.

2. SHAP:

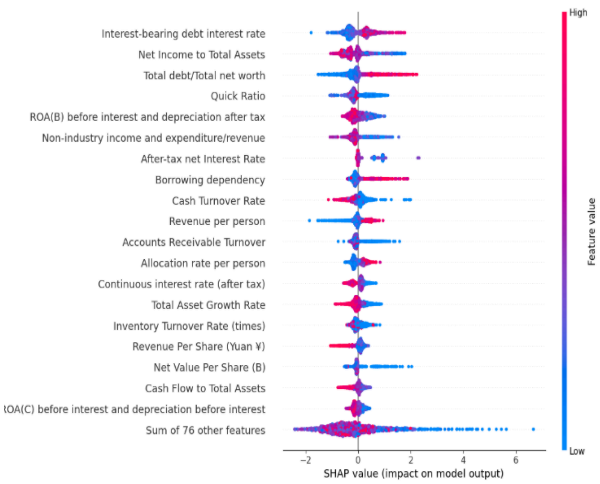


Fig. The SHAP summary plot showed that interest-bearing rate, net income, and total debt/total net worth were the most important features contributing to the predictions.

V. DISCUSSION

We experimented with a variety of machine learning models to classify companies as bankrupt or not based on financial indicators. The models included SVM, XGBoost, LightGBM, Logistic Regression, AdaBoost, Random Forest, Bagging, and Gradient Boosting. Most models showed strong performance, though some, like AdaBoost, exhibited signs of overfitting. After comparing results, **Random Forest** emerged as the most effective and reliable model. It achieved high accuracy while maintaining good generalization, making it the most suitable choice for the bankruptcy prediction task.

Research Gap

Despite strong performance on the Taiwan Economic Journal dataset (1999–2009), our models have yet to be validated on more recent data, other geographic markets, or in the face of dramatic economic events (e.g., the COVID-19 pandemic). Moreover, our focus on tabular financial ratios overlooks alternative data sources (textual filings, news sentiment) and advanced time-series or deep learning architectures that may capture evolving distress patterns. Future work should expand cross-market validation, integrate unstructured data, and explore real-time deployment to ensure robust, generalizable bankruptcy forecasting in dynamic global contexts.

Conclusion

Our study demonstrates that ensemble and boosting methods—particularly Random Forest and XGBoost—can predict corporate bankruptcy with high accuracy (>96%) when trained on a decade of financial ratios from 6,819 Taiwanese firms. By applying SMOTE to address severe class imbalance and integrating Explainable AI techniques (SHAP, LIME, PDP/ICE, surrogate trees), we turn traditionally opaque "black-box" models into transparent tools that reveal which financial indicators most drive risk. This combination of predictive power and interpretability equips investors, lenders, and corporate managers with actionable insights, enabling more informed, proactive decisions to mitigate financial distress.

V. Models Performance

Model name	Accuracy score	Precision	Recall	F1-score
RandomForest(balanced weight,using Smote)	0.9848	0.9738	0.99617	0.9848
SVM(SVC(rbf kerne,c=1000, gamma = scale)	0.97	0.98	0.99	0.98
LogisticRegression	0.966	0.97	0.99	0.98
XGBoost(objective=binary:logistic,,learning_rate=0.1))	0.9626	0.97	0.99	0.98
Lightgbm(LGBMClassifier(objective= binary,metric= auc,learning_rate=0.1))	0.963	0.97	0.98	0.97
GradientBoosting	0.9677	0.666	0.274	0.95
BootStrap Aggregating	0.969	0.777	0.274	0.96
SVM(SVC(kerne= linear))	0.884	0.99	0.89	0.94
XGboostClassifier	0.96	0.98	0.96	0.516
RandomForestClassifier	0.85	0.99	0.85	0.91
AdaBoost	0.964	1.0	0.058	0.11
LogisticRegression(max_iter=1000)	0.62	0.97	0.61	0.75

References

[1] Qu, Y., Liu, Y., & Lu, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. ResearchGate. https://www.researchgate.net/publication/338289416

[2] Shetty, S., Musa, M., & Brédart, X. (2022). Bankruptcy prediction using machine learning techniques. Journal of Risk and Financial Management, 15(1), 35. https://doi.org/10.3390/jrfm15010035

[3]Kim, H., Cho, H., & Ryu, D. (2021). Corporate Bankruptcy Prediction Using Machine Learning Methodologies with a Focus on Sequential Data. Computational Economics, 59(3), 1231–1249. https://doi.org/10.1007/s10614-021-10126-5

[4] Alanis, M., Wang, Y., & Tsai, C. (2022). Benchmarking machine learning models for bankruptcy prediction using U.S. publicly traded firms. arXiv:2212.12051. https://arxiv.org/abs/2212.12051

[5] Arno, D., Chen, Y., & Liu, Q. (2024). Exploring multi-modal bankruptcy prediction models with the ECL dataset. arXiv:2401.12652. https://arxiv.org/abs/2401.12652

[6] ProQuest. (n.d.). Performance comparison of multiple discriminant analysis and logit models in bankruptcy prediction. Retrieved from https://www.proquest.com

[7] Aliaj, T., Anagnostopoulos, A., & Piersanti, S. (2020). Firms Default Prediction with Machine Learning. In Lecture notes in computer science (pp. 47–59). https://doi.org/10.1007/978-3-030-37720-5_4

[8] Shi, Y., & Li, X. (2019, May 26). An overview of bankruptcy prediction models for corporate firms: A Systematic literature review. Shi | Intangible Capital. https://www.intangiblecapital.org/index.php/ic/article/view/1354/756

[9] Wang, N. (2017). Bankruptcy prediction using machine learning. Journal of Mathematical Finance, 07(04), 908–918. https://doi.org/10.4236/jmf.2017.74049

[10] Narvekar, A., & Guha, D. (2021). Bankruptcy prediction using machine learning and an application to the case of the COVID-19 recession. Data Science in Finance and Economics, 1(2), 180–195. https://doi.org/10.3934/dsfe.2021010

[11] Nguyen, H. H., Viviani, J., & Jabeur, S. B. (2023b). Bankruptcy prediction using machine learning and Shapley additive explanations. Review of Quantitative Finance and Accounting. https://doi.org/10.1007/s11156-023-01192-x

[12] Michalkova, L., & Ponisciakova, O. (2025c). Bankruptcy prediction, financial distress and Corporate life Cycle: case study of Central European Enterprises. Administrative Sciences, 15(2), 63. https://doi.org/10.3390/admsci15020063