Taiwanese Bankruptcy Prediction Analysis

Cesy Zamarripa

MIS581-1 Capstone 1

**Abstract**

This analysis investigates the relationship between financial ratios and corporate bankruptcy risk across a population of firms, using a classification-based approach. Four research questions guided the analysis, focusing on the predictive strength of individual financial ratios, the classification potential of financial features, the effect of variable interactions, and the identification of early warning indicators. Logistic regression, decision trees, random forests, and artificial neural networks (ANNs) were applied to a Taiwanese corporate dataset. Key financial ratios such as operating gross margin, borrowing dependency, and return on equity were consistently associated with bankruptcy status, leading to the rejection of the null hypothesis for the primary research question. While formal hypotheses were not tested for model performance, exploratory classification modeling yielded strong results, with logistic regression achieving an AUC of 0.9421 and the ANN model demonstrated the highest recall. These findings suggest that interactions among financial variables enhance model sensitivity, and certain indicators may serve as early predictors of financial distress.

**Introduction**

Data analytics has become very prominent and innovative in the world of finance and business. When it comes down to data-led decisions, huge amounts of assets can be studied and maintained and even used to prevent or predict the loss of those assets, as well as predicting the growth of said assets. Data analytics are crucial in finance and business because they transform raw data into actionable insights that drive smarter decision-making, improve risk management, enhance operational efficiency, and provide a competitive edge. In finance, it plays a vital role in fraud detection, credit risk assessment, and forecasting economic changes, ensuring stability and security. Businesses also leverage data to normalize and modernize operations, reduce costs, and boost productivity, while personalized analytics help tailor products and services to better meet customer needs. In this project, the goal will be to break down a Taiwanese bankruptcy dataset that provides 95 features to study amongst the 6819 company entries between the years 1999-2009 to most accurately predict bankruptcy before occurrence.

**Objectives**

Building on the importance of predictive analytics in financial decision-making, this project sets out to achieve several key objectives. With the exploration of all key financial features, the first objective would be to classify companies into distinct risk categories based on their financial features by building and interpreting decision tree models. The second objective would be to examine the interactions between financial variables and their combined effect on bankruptcy likelihood using random forest algorithms. It would also be key to detect early warning signals of bankruptcy risk by evaluating consistent patterns and thresholds across logistic regression, decision trees, and random forests. The fourth objective is to translate analytical findings into actionable insights that can guide businesses and financial institutions in proactive risk management and decision-making.

**Overview of Study**

This study explores the use of predictive analytics to assess bankruptcy risk among Taiwanese companies using a dataset spanning 1999 to 2009 and containing 95 financial indicators across 6,819 records. By applying logistic regression, decision trees, random forest models and ANN’s, the project investigates which financial ratios most strongly predict bankruptcy, how companies can be categorized into risk levels, and what early warning signs can be identified. The goal is not only to evaluate model performance, but also to offer practical insights for businesses and financial institutions to better manage risk and prevent financial failure.

**Research Questions and Hypotheses**

In the realm of financial analysis, predicting bankruptcy is crucial for stakeholders. This investigation explores whether financial ratios can serve as reliable indicators of bankruptcy risk. By setting the questions and hypotheses on the predictive power of these ratios, we aim to identify key financial indicators and improve early warning systems for businesses. With the clarification of what the research question will focus on to help achieve insight of bankruptcy predictability, the following hypothesis were made for all the respective research questions.

The first research question asked: *Which financial ratios are most strongly associated with bankruptcy?* To address this, the following hypotheses were tested: **H₀:** None of the financial ratios in the population are related to bankruptcy status. **H₁:** At least one financial ratio in the population is related to bankruptcy status.

The second research question explored the classification potential of financial features but was not tested through formal hypothesis testing, as model performance metrics (e.g., AUC, F1-score) are characteristics of specific models trained on sample data rather than parameters of the broader population. Instead, this question *Can financial features be used to classify firms into distinct bankruptcy risk groups?* was evaluated through exploratory modeling and performance assessment.

The third research question asked: *Do interactions between financial ratios significantly influence bankruptcy risk?* The related hypotheses were **H₀:** Interaction terms between financial ratios are not significantly associated with bankruptcy status. **H₁:** At least one interaction between financial ratios is significantly associated with bankruptcy status in the population.

Lastly, the final question was: *Are any financial indicators associated with elevated bankruptcy risk prior to failure?* Corresponding hypotheses were **H₀:** No financial indicator is associated with bankruptcy prior to the point of failure. **H₁:** At least one financial indicator is associated with bankruptcy prior to the point of failure.

**Literature Review**

The following literature review draws on existing research related to financial risk prediction, providing a comprehensive foundation for this study. By examining prior work in data analytics, bankruptcy modeling, and ethical considerations in predictive modeling, it offers both contextual support and broader perspectives to guide the current analysis.

**A Comparison of Analysis Techniques in Literature**

One of the most relevant papers found that was similar in their use of analysis methods was a bankruptcy prediction piece based on Indian Banks (Oberoi, S. S., & Banerjee, S., 2023). This analysis was based on data on both private and public banks in India from 2001-2018. They utilized logistical regressions, random forests, and ADABoost and compared these results to artificial neural network results. In the techniques proposed for my capstone project, it was also decided to utilize logistic regression methods and random forests along with a decision tree to compare to the forests. The reason this text was chosen is because it offered the approach of comparison to a deep learning technique. After further research it was learned that artificial neural networks can capture complex patterns and non-linear relationships that may outperform patterns found from that of a logistical regression. This made for consideration of additional techniques of analysis in the current capstone project. The paper itself and its analysis was of a much smaller sample size with only a total of 59 banks in the used dataset.

Another relevant source of reference found was a Bankruptcy predictive analysis paper that utilized Hybrid Neural Networks methods and compared it to Artificial Neural Network (ANN) results (Marso, S., & EL Merouani, M., 2020). The authors claimed “The experimental results indicate that ANN models, on average, are approximately 10% more accurate in relation to MDA (multiple discriminant analysis) in different periods. Our Model ABCNN (Artificial Bee Colony trained ANN) led to 92% accuracy, whereas BPNNs (Back Propagation trained ANN’s) and MDAs led to 91% and 81% accuracy, respectively in one year before the bankruptcy.” (Marso, S., & EL Merouani, M., 2020). This was interesting because now I not only had insight into the method that ANN brings to the table and this paper offers addition experimental methods that provide more accurate results in Bankruptcy prediction that also utilize ANN’s to their advantage.

**Exploring the Benefits of Bankruptcy Analytics**

When it came to the type of literature that could offer additional support to my research, looking into the importance or possible types of benefits this research could provide is a great support factor. The paper *Sustainable Development and Economic Stabilization Through Artificial Intelligence in Decision Support Systems, For Business Bankruptcy Prediction* (Nikolla, S., Hoxha, E., & Mujo, A., 2025) dives into showing how AI can transform Bankruptcy prediction. This paper discusses both the benefits that bankruptcy analytics can bring and what methodologies are out there for best achieving accurate results. They discuss how Neural Networks and Random Forest models offer greater insight than just a logistical regression analysis. This helps support my decision in choosing to explore both techniques within my research. Additionally, the paper talks about some other cutting edge machine learning models that can offer more accurate predictive abilities.

**Ethics in Data Analytics Publication**

Since one of the topics of discussion within this research has been about ethical concerns, a publication about Ethics in Data Analytics and Big Data was a text that would offer resourceful and relevant information to the project. The paper (Harlow H., 2018) talks about laws in the United States and in Europe that are relevant to the laws and regulation of ethics in the world of big data and analytics. It discusses how the use of big data is important and should be utilized with good intent. Privacy concerns are discussed as well as how monopolies and lack of regulation are a cause for concern in ethical restrictions of data analytics.

**Methodology**

The dataset to be utilized in this project analysis was provided by the UC Irvine Machine Learning Repository (*UCI Machine Learning Repository*, n.d.-c). This dataset is based on data provided by the Taiwan Economic Journal. The target variable of the dataset is a binary variable for whether the company went bankrupt, and the other 95 variables are all integer and continuous integer variables of different financials related to each company (profits rates, interest rates, loans, etc.).

This dataset containing financial records of 6,819 companies and 95 variables, is a valuable resource for studying bankruptcy trends and demonstrating the power of data analytics in finance. Despite lacking explicit timestamps, the dataset provides comprehensive financial indicators such as profitability, debt levels, and liquidity, that allow organizations to identify patterns signaling financial distress and predict bankruptcy risks. Its binary classification structure (bankrupt/not bankrupt) facilitates machine learning applications, enabling analysts to extract actionable insights that contribute to risk mitigation, strategic planning, and resource allocation.

When it comes down to studying the influence of the financial indicators towards predicting bankruptcy, investigating the trends might help to create feature based predictive models based on patterns found amongst the variables. With the target variable of the dataset being a binary variable, it made the most sense to approach the analysis initially with logistical regression methods. To gain a more wholistic view of the 95 features and more powerful business insights, the use of decision trees was chosen as an additional method. Decision trees offer deeper insights by providing a visualization of how each feature impacts the outcome. It helps to narrow down which predictors matter most. An even more powerful method also chosen to be used is a forest method, where an ensemble of decision trees is created from the data to best produce more accurate results of important predictors in the data. After literature review, it was also decided to utilize the ANN method for this dataset to be compared to the other predictive techniques to explore and compare the accuracy between the chosen analytical methods.

The tool that will be used to perform this analysis will be SAS OnDemand. SAS is a powerful analytics platform that supports different types of statistical and machine learning techniques. For logistic regression, SAS offers robust procedures like PROC LOGISTIC that provide detailed diagnostics and model interpretation, ideal for evaluating the influence of financial ratios on bankruptcy. When it comes to decision trees and random forests, SAS’s procedures such as PROC HPSPLIT and PROC HPFOREST allow for efficient model training, variable importance ranking, and intuitive visualizations that enhance interpretability. Additionally, SAS supports artificial neural networks (ANNs) through tools like PROC NEURAL or within its Visual Data Mining and Machine Learning environment, enabling comparison across multiple modeling approaches. Its integrated environment ensures data preprocessing, model building, and evaluation can all be performed within a single workflow, streamlining the analysis and improving reproducibility.

**Limitations**

While the chosen predictive methods offer valuable insights, there are several limitations to consider. Logistic regression depends on the assumption of linear relationships, which may overlook more complex patterns in the data. Decision trees, although great for interpretability, are prone to overfitting, especially with large numbers of variables. Random forests address that issue to an extent but can still be sensitive to noise in the data set. One of the more advanced techniques used, artificial neural networks, can produce powerful results, but according to (Abdulhafedh, 2022), they often lack transparency, making it difficult to interpret why certain predictions are made. Additionally, given the imbalance that typically exists in bankruptcy datasets, there’s a risk of biased predictions toward non-bankrupt cases. Lastly, with 95 financial features, multicollinearity could affect how certain variables are interpreted, and ethical considerations must be kept in mind to avoid unfair classifications or unintended consequences in real-world applications.

**Ethical considerations**

When it comes to the world of financial data analytics, there are common areas of ethical concern. As mentioned in the literature review, some of the concerns are privacy, bias, transparency and responsible use (Harlow H., 2018). Financial data often contains sensitive information. Making sure regulations and financial industry standards are being followed is important to protect businesses from data misuse.

Bias being another huge piece of consideration is also important because if the dataset lacks diversity or represents only certain types of businesses, predictions will be skewed. There are also other types of bias to look out for and avoid when performing data analysis. There’s confirmation bias, label bias and measurement bias (Carrascosa, 2025). Confirmation bias is if the analysis being performed is with the intention of only confirming certain predictions. Label bias is when there are classifications when labeled data used for training reflect subjective or biased human decisions. Measurement bias is when the methods used to collect the data are biased, in this study it can be how the bankruptcy measure of a company was determines yes or no.

Transparency is ensuring clear documentation of data collection and analysis to prevent misrepresentation. In the world of analytics, withholding data analysis methods or intentions is a lot of the time illegal and unethical.

When it comes to the responsible use of data intention is a huge piece of that decision. In a Harvard business school publication, the discussion of intention is brought to attention. Stating “Before collecting data, ask yourself why you need it, what you’ll gain from it, and what changes you’ll be able to make after analysis. If your intention is to hurt others, profit from your subjects’ weaknesses, or any other malicious goal, it’s not ethical to collect their data” (*5 Principles of Data Ethics for Business*, 2021). This can be important, especially in the world of big business where data collection and data analysis especially revolving around the prevention of bankruptcy of those businesses is a very blurred line in certain societies thus so is the analytics being performed behind it. Having powerful insights into the world of business can easily lead to unethical decision making.

**Findings- Logistical Regression**

When performing the logistical regression in SAS, the decision was made to perform a stepwise logistical regression. This was due to the warning SAS initially gave that the ridging technique was unable to make a difference in the optimization of the model. The model rendered 17 variables that best represented the key predictors for Bankruptcy. Figure 1 shows a table of each of the key predictors, what they represent and their corresponding p-values.

**Figure 1**

*Stepwise Logistical Regression Key Predictor Results*

| Step | Variable | Financial Metric | Interpretation | p-value |
| --- | --- | --- | --- | --- |
| 1 | Var87 | Liability to Equity | Leverage indicator—higher values mean more debt relative to equity | < 0.0001 |
| 2 | Var38 | Long-term Fund Suitability Ratio (A) | Measures quality/stability of long-term capital sources | < 0.0001 |
| 3 | Var20 | Revenue Per Share (¥) | Revenue efficiency—declining values may signal weakening sales | < 0.0001 |
| 4 | Var46 | Fixed Assets Turnover Frequency | Efficiency of turning fixed assets into sales revenue | < 0.0001 |
| 5 | Var58 | Current Assets / Total Assets | Liquidity mix—proportion of assets easily convertible to cash | < 0.0001 |
| 6 | Var37 | Net Worth / Assets | Equity strength—how much of assets are owned outright | < 0.0001 |
| 7 | Var82 | Liability-Assets Flag | Structural threshold flag—likely signals breach in balance sheet health | < 0.0001 |
| 8 | Var75 | Cash Flow to Total Assets | How well a company generates cash from its assets | 0.0035 |
| 9 | Var69 | Current Liabilities / Equity | Short-term financial pressure relative to shareholder value | 0.0025 |
| 10 | Var50 | Operating Profit per Person | Labor efficiency—how productive each employee is in driving profit | 0.0052 |
| 11 | Var24 | Operating Profit Growth Rate | Profit acceleration or deterioration over time | 0.0116 |
| 12 | Var86 | Net Income to Stockholders’ Equity (ROE) | Shareholder return—low ROE signals shrinking value creation | 0.0134 |
| 13 | Var45 | Inventory Turnover Rate (times) | Inventory efficiency—slow turnover may reflect overstock or low demand | 0.0040 |
| 14 | Var17 | Net Value Per Share (C) | Per-share equity valuation—used to gauge underlying business strength | 0.0221 |
| 15 | Var19 | Cash Flow Per Share | Cash health per unit of ownership—critical for short-term survival | 0.0075 |
| 16 | Var43 | Total Asset Turnover | Overall efficiency—how quickly total assets generate revenue | 0.0105 |
| 17 | Var77 | CFO to Assets | Cash from operations relative to total assets—key for ongoing viability | 0.0134 |

The modeling process began with a null model, including only the intercept term and no predictors. At this stage, the model’s −2 Log Likelihood (−2 Log L) was 1943.71, and the Akaike Information Criterion (AIC) stood at 1945. The −2 Log L is a metric of model fit where the lower the value represents that the model explains the outcome data better. Similarly, AIC balances goodness of fit with model simplicity by penalizing unnecessary complexity; a lower AIC indicates a more cost-conscious and well-fitted model. Through 17 steps of variable selection, the model iteratively added statistically significant predictors, ultimately reducing −2 Log L to 1159.53 and AIC to 1205, signaling a substantial improvement in fit and efficiency.

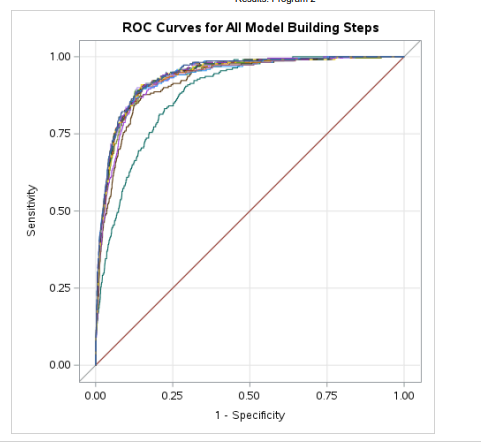
Out of 95 available financial variables, the model selected 17 that best predicted bankruptcy. Among the most statistically significant were leverage metrics such as Liability to Equity (*p* < .0001), which reflects how heavily a company relies on debt financing and Long-term Fund Suitability Ratio (*p* < .0001), which assesses the adequacy of long-term funding to support long-term investments. These indicators emphasize the role of capital structure and debt burden in shaping financial vulnerability. Profitability indicators, including Revenue Per Share and Return on Equity (ROE), calculated as *Net Income to Stockholders’ Equity*, suggest that declining sales performance and weaker shareholder returns are associated with elevated bankruptcy risk.

Liquidity and cash flow indicators were also central: variables such as Cash Flow to Total Assets, Cash Flow Per Share, and Current Assets to Total Assets highlighted an organization’s ability to meet short-term obligations. Lower values in these metrics typically signal operational strain and limited financial flexibility. Meanwhile, operational efficiency ratios like Fixed Assets Turnover, Inventory Turnover, and Total Asset Turnover captured how effectively companies convert investments and resources into revenue. Lagging efficiency in these areas may be an early warning of sluggish operations or excess capacity.

From the regression results, an additional visualization of metrics was created to comparatively observe side by side, the change in the model with each step, compared to random chance (step 0). A receiver Operating Characteristic (ROC) analysis was conducted, as shown in figure 2.

**Figure 2**

*ROC Curves for All Steps*



Note: Straight red line is step 0

The ROC curve for the stepwise-selected model exhibited a notably high Area Under the Curve (AUC) value of 0.9421, indicating excellent classification ability between bankrupt and non-bankrupt firms. This strong result reflects the model's effectiveness in distinguishing risk cases across a range of probability thresholds. In contrast, the intercept-only model (baseline/step 0) produced an AUC of 0.5000, equivalent to random chance. The substantial improvement in AUC underscores the explanatory value of the selected financial indicators and further validates the strength of the stepwise regression approach.

**Decision Tree**

The second method of analysis used for this investigation was decision trees. In order to avoid overfitting, the prune complexity option was included in the tree as well as the entropy option to utilize entropy to decide the best splits. The results of each key metric identified are shown in figure 3 along with their importance value. The fit statistics table of the model is shown in figure 4.

**Figure 3**

*Decision Tree Variable Results*

| **Variable** | **Metric** | **What It Measures** | **Importance (Val.)** | **Interpretation & Risk Signal** |
| --- | --- | --- | --- | --- |
| **Var41** | Borrowing Dependency | Reliance on external debt for capital funding | **3.31** | High borrowing reliance signals financial fragility; dominant split in both training and validation. Strong indicator of capital structure imbalance. |
| **Var4** | Operating Gross Margin | Profit after operating expenses (before tax/debt), relative to revenue | **3.30** | Lower margins suggest declining operational efficiency; helps separate solvent vs. distressed firms based on core profitability. |
| **Var58** | Current Assets / Total Assets | Proportion of liquid, short-term assets | **1.66** | Liquidity metric—low values raise red flags for short-term solvency and payment ability. |
| **Var24** | Operating Profit Growth Rate | Year-over-year change in core operating profits | **1.44** | Negative or stalled growth suggests weakening operational health and declining income trajectory. |
| **Var45** | Inventory Turnover Rate | How quickly is inventory sold or used | **0.79** | Lower turnover may imply excess inventory, sluggish sales, or ineffective demand planning. |
| **Var91** | Net Income to Stockholders’ Equity (ROE) | Returns generated for shareholders from equity investment | **0.77** | A profitability and efficiency signal—declining ROE often precedes investor flight and indicates weak earnings power. |
| **Var20** | Revenue Per Share | Company’s sales divided by number of outstanding shares | **0.00** | Important in training only—likely overfit or redundant. Removed during pruning due to lack of predictive generalization. |
| **Var30** | Total Asset Growth Rate | Change in a firm’s asset base over time | **0.00** | Asset growth didn’t contribute to predictive accuracy in validation—suggests volume alone isn’t a sufficient bankruptcy signal here. |
| **Var55** | Working Capital to Total Assets | Proportion of net current assets to total assets | **0.00** | While theoretically a liquidity marker, it was pruned due to inconsistent performance—perhaps correlated with stronger variables. |
| **Var13** | Cash Flow Rate | Cash flow compared to revenue | **0.00** | Useful in training, but cut from final model—possibly redundant with ROE or current asset metrics already included. |

**Figure 4**

*Fit Statistics for Selected Tree table*

A screenshot of a statistics

AI-generated content may be incorrect.

The decision tree analysis identified a different but complementary set of bankruptcy predictors compared to the stepwise logistic regression. The most influential variables in the final tree included Borrowing Dependency, Operating Gross Margin, and Current Assets to Total Assets, underscoring the importance of leverage, core operational profitability, and liquidity. Unlike the logistic model, which prioritized variables like Liability to Equity and achieved stronger statistical fit metrics (AUC = 0.9421), the decision tree exhibited lower but still respectable performance with an AUC of 0.7685 and validation accuracy near 97%. However, the tree’s sensitivity remained low (15.63%), suggesting it correctly identified only a small portion of actual bankruptcies while maintaining excellent specificity. Several variables used in training, such as Revenue Per Share and Working Capital to Total Assets, were pruned during validation due to limited generalizability. Overall, while the logistic regression offered greater discriminatory power and predictive sharpness, the decision tree delivered an interpretable, rule-based structure that revealed key financial thresholds. When used together, these models provide both quantitative strength and transparent decision logic for assessing bankruptcy risk.

**Random Forests**

Initial attempts to implement the Random Forest algorithm were conducted using SAS Studio; however, due to the unavailability of the Viya environment, this functionality was not supported. Consequently, the modeling process was transitioned to Python via Google Colab, which provided access to the required libraries for training and evaluating Random Forest and artificial neural network (ANN) models. The results for the forests are shown in figure 5.

**Figure 5**

*Random Forest Model Performance Summary*

A screenshot of a computer

AI-generated content may be incorrect.

The Random Forest model achieved high overall accuracy (96%) and improved sensitivity over the decision tree, identifying approximately 15% of actual bankruptcies after applying class weighting. While still modest, this recall represents a 50% relative increase compared to the decision tree’s 10.6% and mirrors similar struggles with imbalance across models. Logistic regression remained the strongest in discriminatory power, reaching an AUC of 0.9421 and higher sensitivity (~26%), though it lacked the ensemble robustness of the forest. All three models consistently identified variables like Operating Gross Margin (Var4), Borrowing Dependency (Var41), and Return on Equity (Var91) as key predictors, reinforcing their reliability across methodologies. Ultimately, while the Random Forest improved slightly on recall without sacrificing accuracy, it remained less sensitive than logistic regression and more opaque than the interpretable decision tree. Taken together, the models reveal a trade-off between predictive transparency, complexity, and class sensitivity, especially in the face of rare-event data.

**Artificial Neural Network (ANN)**

The last and final methodology of comparison is ANN’s. This was also done utilizing Python. The results table for the model are shown in figure 6.

**Figure 6**

*ANN Model Results*

A screenshot of a graph

AI-generated content may be incorrect.

These results along with the other three methodologies are compared in figure 7.

**Figure 7**

*Model Comparison Table*

| Model | AUC | Recall (1s) | Accuracy | Notes |
| --- | --- | --- | --- | --- |
| Logistic Regression | 0.9421 | ~26% | Very High | Best overall AUC, interpretable |
| Decision Tree | 0.7685 | 10.6% | High | Transparent, simple rules |
| Random Forest | 0.9404 | 15% | 96% | Boosted recall w/ class weights |
| ANN (Keras) | 0.8169 | 31% | 95% | Deep pattern detection |

Between the four models evaluated, logistic regression delivered the strongest overall performance, achieving the highest discrimination ability with an AUC of 0.9421 and a sensitivity of approximately 26%, despite its linear assumptions. The decision tree model offered full transparency and ease of interpretation, but its sensitivity lagged at 10.6%, indicating limited ability to detect actual bankruptcies. The random forest, after applying class weighting, showed improvement in recall (15%) while preserving accuracy (96%), suggesting better responsiveness to class imbalance without overfitting. The artificial neural network demonstrated the highest recall of all models (31%) and a robust AUC of 0.817, revealing its strength in capturing nonlinear and complex patterns. While interpretability was affected, the ANN’s sensitivity to rare bankruptcy cases underscores its value as a complementary model. Taken together, these results highlight a trade-off between transparency and predictive depth, with logistic regression and ANN forming strong boundary points on that spectrum and the decision tree and forest filling in between.

**Conclusion**

This study explored whether financial ratios are associated with corporate bankruptcy risk in the broader population, using multiple classification techniques including logistic regression, decision tree, random forest, and artificial neural network (ANN) models. Across these methods, key financial indicators such as *Operating Gross Margin*, *Borrowing Dependency*, and *Return on Equity* consistently came up as significant predictors of bankruptcy. These results support rejection of the null hypothesis for Research Question 1, which claimed there was no relationship between financial ratios and bankruptcy status in the population.

Although not formally tested through hypothesis testing, the classification models trained on financial ratios achieved strong predictive performance. Logistic regression demonstrated an AUC of 0.9421, while both random forest and ANN models surpassed 95% accuracy. These exploratory results suggest that financial features meaningfully differentiate bankruptcy risk categories and can support practical risk segmentation, although conclusions here relate to the sample rather than population inference.

Regarding variable interactions for research question 3, results from ANN and random forest models which can capture nonlinear and multivariate relationships demonstrated improved sensitivity compared to logistic regression alone. These findings support the rejection of the third null hypothesis, indicating that at least some interactions between financial variables may be associated with bankruptcy outcomes in the population.

Lastly, several predictors such as *Operating Profit Growth Rate* and *Cash Flow to Total Assets* reflected performance patterns over time, despite the absence of time-series features. Their repeated selection across models suggests they may serve as early warning indicators. As such, the fourth null hypothesis, which assumed no association between financial indicators and early bankruptcy risk, is provisionally rejected.

Overall, these results highlight the practical value of financial ratios in forecasting bankruptcy and demonstrate the contributions of both traditional and machine learning models. Future research should expand upon these findings by incorporating patient-level or time-dependent data, industry-specific variation, and additional contextual features to enhance predictive accuracy and generalizability.

**Recommendations**

Since the investigation was able to disprove all three null hypotheses, that suggests the dataset holds some significant patterns, certain financial ratios really do predict bankruptcy risk. That means the data isn’t just noise; the models were picking up on real signals, not random chances. But at the same time, none of the models were perfect, especially when it came to catching the smaller group of companies that actually went bankrupt. That suggests there are probably other factors either missing from the data or not yet explored that could help boost performance. For example, adding in time-based features like changes over several years, or including non-financial info like industry type or economic conditions, could help models spot red flags earlier. Incorporating time-series data would allow for the analysis of financial trends leading up to bankruptcy, offering deeper insight into early warning signals. This could improve model sensitivity by capturing not just static financial health but deteriorating patterns over multiple periods. The fact that the artificial neural network did better at catching bankruptcies shows there may be more complex relationships between variables that linear models can't quite capture.

Other things that could’ve influenced results are the class imbalance in the dataset (since bankruptcies are rare) and the fact that the data is based on Taiwanese companies, which might limit how well the findings apply elsewhere. Overall, the results show that financial ratios do tell a story, but it’s probably not the whole story.

**References**

*5 Principles of data Ethics for business*. (2021, March 16). Business Insights Blog. <https://online.hbs.edu/blog/post/data-ethics>

Abdulhafedh, A. (2022). Comparison between Common Statistical Modeling Techniques Used in Research, Including: Discriminant Analysis vs Logistic Regression, Ridge Regression vs LASSO, and Decision Tree vs Random Forest. OALib, 09(02), 1–19. https://doi.org/10.4236/oalib.1108414

Carrascosa, I. P. (2025, May 19). *Ethical data analysis: avoiding bias and ensuring fairness*. Statology. https://www.statology.org/ethical-data-analysis-avoiding-bias-and-ensuring-fairness/

Harlow, H. (2018). Ethical Concerns of Artificial Intelligence, Big Data and Data Analytics. *European Conference on Knowledge Management*, *1*, 316–323.

Marso, S., & EL Merouani, M. (2020). Bankruptcy Prediction using Hybrid Neural Networks with Artificial Bee Colony. *Engineering Letters*, *28*(4), 1191–1200.

Nikolla, S., Hoxha, E., & Mujo, A. (2025). Sustainable Development and Economic Stabilization through Artificial Intelligence in Decision Support Systems, for Business Bankruptcy Prediction. *International Journal of Ecosystems & Ecology Sciences*, *15*(1), 341–348. <https://doi-org.csuglobal.idm.oclc.org/10.31407/ijees15.138>

Oberoi, S. S., & Banerjee, S. (2023). Bankruptcy Prediction of Indian Banks Using Advanced Analytics. *Economic Studies*, *32*(4), 22–41.

*UCI Machine Learning Repository*. (n.d.-d). https://archive.ics.uci.edu/dataset/572/taiwanese+bankruptcy+prediction