# GROUP ASSESSMENT ITEM COVER SHEET

| Student Numbers: | Emails: | FIRST NAMES | FAMILY / LAST NAMES |
|---|---|---|---|
| 3425175 | c3425175@uon.edu.au | Tan Thuy Linh | Dao |
| 3493680 | c3493680@uon.edu.au | Yudie | Qiu |
| 3498592 | c3498592@uon.edu.au | Tianxiang | Bian |
| 3422184 | c3422184@uon.edu.au | Buu Quan | Luu |
| 3489082 | c3489082@uon.edu.au | Hanfei | Cao |

| Course Code | Course Title |
|---|---|
| FINN2003 | Financial Technologies and Innovations |
| *(Example)* FINN2003 | *(Example)* Financial Technologies and Innovations |

Campus of Study: City Campus (Nuspace) (eg Callaghan, Ourimbah, Port Macquarie)

Assessment Item Title: Predicting Corporate Default Rate Scenario Due Date/Time: 14 Sep

Tutorial Group (If applicable): Word Count (If applicable): 2,195

Lecturer/Tutor Name: Yvonne Fang

Extension Granted: ☐ Yes ☐ No Granted Until:

Please attach a copy of your extension approval

**NB: STUDENTS MAY EXPECT THAT THIS ASSIGNMENT WILL BE RETURNED WITHIN 3 WEEKS OF THE DUE DATE OF SUBMISSION**

**Please tick box if applicable**

☑ *Students within the Faculty of Business and Law, Faculty of Science, Faculty of Engineering and Built Environment and the School of Nursing and Midwifery:* We verify that we have completed the online Academic Integrity Module and adhered to its principles.

☑ *Students within the School of Education:* We understand that a minimum standard of correct referencing and academic literacy is required to pass all written assignments in the School of Education; and we have read and understood the School of Education *Course Outline Policy Supplement*, which includes important information related to assessment policies and procedures.

We declare that this assessment item is our own work unless otherwise acknowledged and is in accordance with the University's Student Academic Integrity Policy

We certify that this assessment item has not been submitted previously for academic credit in this or any other course. We certify that we have not given a copy or have shown a copy of this assessment item to another student enrolled in the course, other than members of this group.

We acknowledge that the assessor of this assignment may, for the purpose of assessing this assignment:

- Reproduce this assessment item and provide a copy to another member of the Faculty; and/or
- Communicate a copy of this assessment item to a plagiarism checking service (which may then retain a copy of the item on its database for the purpose of future plagiarism checking).
- Submit the assessment item to other forms of plagiarism checking.

We certify that any electronic version of this assessment item that we have submitted or will submit is identical to this paper version.

Turnitin ID:
(if applicable)

**DATE STAMP HERE**

Signature: Date: 14 Sep 2025

Signature: Date: 14 Sep 2025

Signature: Date: 10 Sep 2025

Signature: Date: 14 Sep 2025

Signature: Date: 14 Sep 2025

Insert this

## TABLE OF CONTENTS

# TABLE OF FIGURES

## 1. Introduction

This report aims to predict the probability of default risk which is affected by financial and firm-level variables. In particular, the key model implemented is Decision Tree, supported by comparative analysis with Logistic Regression, Random Forest, and Support Vector Classifier to compare their performance.

Decision tree, classification algorithm, produce basic if-then rules for the data to construct a tree-based method where any path starting at the root is illustrated by a series of data that separates until a Boolean result is attained at a leaf node, which are predicted outcomes (International Conference on Computational Science and Computational Intelligence, 2019). Specifically, this is composed of nodes and branches, nodes represent features, and branches correspond to variable ranges (Charbuty & Abdulazeez, 2021), dividing the firms progressively until one of the leaf nodes contains a classification of default or non-default. Random forests, as ensembles, reduce overfitting and capture nonlinear patterns but sacrifice transparency. Logistic regression, using the sigmoid function, offers stability and clear interpretability. Meanwhile, SVCs maximise class margins and, with kernels, model complex relationships.

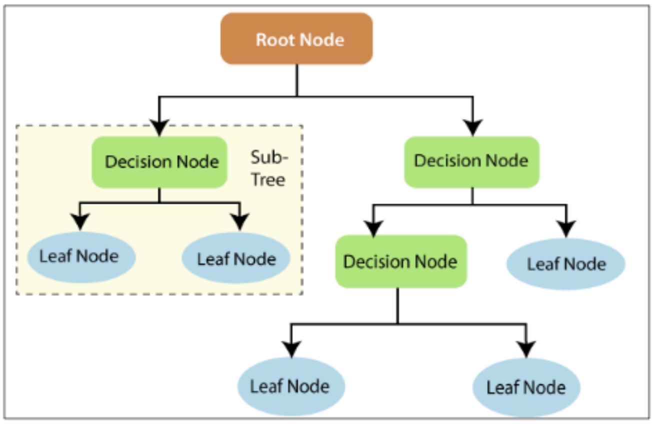


Figure 1: Decision Tree

The variables include debt-to-asset ratio, current ratio, return on assets (ROA), firm size, cash flow, ESG score, and industry, grouped into financial ratios and macroeconomic indicators as determinants of default risk. Regression-based prediction is used, with each variable influencing default likelihood before dataset generation and coefficient estimation. Data scaling is then applied to improve Decision Tree performance, which is then interpreted and compared with the baseline and other machine learning models.

## 2. Data Collection and Processing

The dataset was acquired randomly with 100,000 observations and set random seed every run ensuring consistent results, with 7 different variables on firm-level financial indicators to build the machine learning model for default prediction (Default = 0, No default = 1). Simulated data was chosen over real datasets due to limited availability, confidentiality, and methodological convenience (Chan-Lau,Hu, Mungo, Qu, Xin, & Zhong,2024). Unlike real data, it contains only continuous variables, avoiding issues such as missing values, noise, or inconsistent reporting (Charalambakis & Garrett, 2019).

In terms of variable construction, the dataset combines traditional financial ratios with emerging default risk factors. Firstly, the business industry is represented across five sectors and encoded numerically for machine learning analysis. Moreover, ESG ratings, ranging from 1–9 (AAA–C), reflect firms' environmental, social, and governance frameworks, with higher scores indicating weaker performance (Li, Zhang, & Zhao, 2022a). Besides that, the remaining six financial ratios (debt to asset ratio, current ratio, ROA, firm size, cash flow and interest rate) were randomly generated within reasonable ranges and rounded to four decimals to reduce complexity (Löffler & Maurer, 2011; Bajwa & Rashid, 2021a; Cathcart, Dufour, Rossi, & Varotto, 2020).

```python
# Set random seed for reproducibility (ensuring consistent results every run)
np.random.seed(42)
# Number of observations (companies)
n_samples = 100000
```
✓ 0.0s

```python
# Define industries
industries = ['Manufacturing & Construction','Distribution (Wholesale & Retail)','Food & Agriculture','Transportation','Services']
# Creating a DataFrame with random values simulating financial indicators
data = pd.DataFrame({
    'Debt_to_Asset Ratio': np.round(np.random.uniform(0, 1, n_samples), 4),        # Leverage
    'Current_Ratio': np.round(np.random.uniform(0.25, 3, n_samples), 4),          # Liquidity
    'ROA': np.round(np.random.uniform(-0.5, 0.3, n_samples), 4),                  # Profitability
    'Firm_Size': np.round(np.random.uniform(10, 20, n_samples), 4),               # Log firm size
    'Cash_Flow': np.round(np.random.uniform(-0.5, 1.0, n_samples), 4),            # Cash flow to debt ratio
    'ESG': np.random.randint(1,10, n_samples),                                    # ESG score
    'Industry': np.random.choice(industries, size=n_samples)                      # Industry categorical variable
})

# Map industries to numeric codes (1-9)
industry_mapping = {industry: i+1 for i, industry in enumerate(industries)}
data['Industry_Code'] = data['Industry'].map(industry_mapping)
data = data.drop(columns=['Industry'])                                            #Drop the 'Industry' column as it's not needed for EDA

print(industry_mapping)
data
```
✓ 0.1s

Figure 2: Data Collection and Selection

{'Manufacturing & Construction': 1, 'Distribution (Wholesale & Retail)': 2, 'Food & Agriculture': 3, 'Transportation': 4, 'Services': 5}

| | Debt_to_Asset Ratio | Current_Ratio | ROA | Firm_Size | Cash_Flow | ESG | Industry_Code |
|---|---|---|---|---|---|---|---|
| 0 | 0.4927 | 1.1155 | 0.1372 | 17.2332 | 0.8337 | 8 | 3 |
| 1 | 0.3592 | 2.4175 | -0.2028 | 17.2754 | 0.3010 | 2 | 4 |
| 2 | 0.5527 | 1.2748 | -0.1718 | 15.8680 | 0.4487 | 2 | 4 |
| 3 | 0.2386 | 1.2350 | -0.2459 | 17.2717 | 0.8836 | 6 | 3 |
| 4 | 0.6354 | 2.9967 | -0.1493 | 20.7824 | 0.2693 | 5 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 995 | 0.4150 | 1.3033 | -0.3566 | 20.9257 | 0.2434 | 7 | 2 |
| 996 | 0.4543 | 1.1809 | -0.3213 | 13.0064 | 0.1791 | 9 | 1 |
| 997 | 0.8300 | 0.5626 | -0.2752 | 10.9206 | 0.7720 | 7 | 4 |
| 998 | 0.1774 | 2.6761 | -0.4918 | 9.1009 | -0.0579 | 3 | 1 |
| 999 | 0.2203 | 1.9424 | 0.2336 | 12.4268 | -0.4451 | 6 | 1 |

1000 rows × 7 columns

Figure 3: Result of Data Selection and Preparation

## 3. Variable Selection and Economic Justification

Variables included in the current research were selected due to their theoretical and empirical significance in predicting the default and the previous literatures. The Debt-to-Asset Ratio is one of the critical variables as it portrays a firm's leverage relative to the total assets of the firm. Higher leverage signals greater financial commitments in comparison with the assets, which are intricately linked with a high likelihood of default (Löffler and Maurer, 2011a). Following established practice, the ratio is constrained between 0 and 1.

The Current Ratio was incorporated to reflect liquidity, as the ability to meet short-term liabilities is critical for survival. A ratio below 1 implies technical insolvency, whereas higher liquidity ratios reduce default risk (Grammenos, Nomikos, & Papapostolou, 2008). Nevertheless, excess liquidity may signal a lack of profitable investment opportunities, which could negatively impact long-term performance (Zhang et al., 2020a). Thus, the relationship is expected to be negative but nonlinear, with a practical range between 0.25 and 3.

Profitability is captured through Return on Assets (ROA), reflecting how efficiently assets generate net income. A higher ROA indicates stronger financial health, leading to a lower default probability. The higher a company’s return on assets compared to its industry, the better (Grammenos et al., 2008b). The variable is bounded between -0.5 and 0.3 to capture both loss-making and profitable firms, following prior studies (Löffler & Maurer, 2011b). In addition, ROA is also a key indicator for investors when making investment decisions (Tserng et al., 2014).

Firm Size, measured as the logarithm of total assets, which is ln (total assets), was selected as larger firms have lower default probability (DP) due to business diversification and broader financing channels (Zhang et al., 2020b). Large-scale industries exhibit higher external financial dependence than small-scale industries. In addition, they also face fewer credit constraints, whereas smaller firms are more vulnerable. Following Westgaard and Van der Wijst (2001a), firm size was restricted between 10 and 21 to exclude inactive firms with potential accounting errors.

Table 3
Calibrated average asset correlations at the 99.9% percentile for portfolios based on size categories

| World portfolios | | US portfolios | | Japanese portfolios | | European portfolios | |
|---|---|---|---|---|---|---|---|
| Asset size categories | $\bar{\rho}_A$ | Asset size categories | $\bar{\rho}_A$ | Asset size categories | $\bar{\rho}_A$ | Asset size categories | $\bar{\rho}_A$ |
| ($0, $20m] | 0.1000 | ($0, $20m] | 0.1000 | ($0, $100m] | 0.2000 | ($0, $25m] | 0.1125 |
| [$20m, $100m] | 0.1000 | [$20m, $100m] | 0.1500 | [$100m, $200m] | 0.2000 | [$25m, $75m] | 0.1125 |
| [$100m, $300m] | 0.1125 | [$100m, $300m] | 0.1750 | [$200m, $400m] | 0.2500 | [$75m, $200m] | 0.1250 |
| [$300m, $1b] | 0.1375 | [$300m, $1b] | 0.2250 | [$400m, $1,b] | 0.3000 | [$200m, $1b] | 0.1500 |
| ⩾ $1b | 0.2000 | ⩾ $1b | 0.3000 | ⩾ $1b | 0.4500 | ⩾ $1b | 0.2250 |

Figure 4: The total assets indicating firm size (Lopez, 2004)

Another important determinant is Cash Flow to Debt, which directly measures a firm’s ability to meet obligations (Westgaard & Van der Wijst, 2001b). Higher ratio reduces default risk, while persistent negative operating cash flow signals financial distress (Bhimani, Gulamhussen, & da Rocha Lopes, 2014). Negative values may result from losses, high working capital needs. For instance, a ratio of -0.5 means OCF covers only half of debt, while 1.0 implies the firm could repay its entire debt within one year's cash flow.

Beyond financial indicators, ESG ratings were included to capture the impact of sustainability practices. Firms with stronger ESG performance tend to increase shareholder value, have better credit scores, and stronger political connections (Li, Zhang, & Zhao, 2022b). The ESG investments are associated with cash flow and affect short-term repayment ability. Strong ESG performance has been found to enhance resilience and access to finance, reduce default risk.

Finally, Industry categories were considered to account for sector-specific variations in credit risk, as systematic factors such as business cycles, regulation, and capital intensity affect default rates (Memmel, Gündüz, & Raupach, 2015a). Industry affiliation also affects recovery values in default (Benzschawel, Haroon & Wu, 2011). Short-term interest rates were initially considered but excluded due to their ambiguous relationship with default risk, as high rates can simultaneously increase both borrowing costs and signal robust economic conditions (Zhang et al., 2020c).

## 4. Coefficient Interpretation

The regression analysis provides important insights into the relationship between firm-specific, financial, and non-financial factors and the probability of default. Starting with Debt-to-Asset Ratio (Leverage), the estimated coefficient is +0.5, indicating that higher leverage significantly increases default risk. This finding is consistent with Löffler and Maurer (2011c), who identify leverage as a central driver of default probability, with coefficients ranging from 0.47 to 0.6. The positive sign is economically sensible, as firms with higher debt levels face greater repayment pressure and weaker solvency positions.

The Current Ratio (Liquidity) is associated with a coefficient of -0.32, suggesting that greater liquidity reduces the likelihood of default. This result aligns with Nadarajah et al. (2021), who find that firms with stronger short-term assets relative to liabilities face lower insolvency risk with coefficient ranging from -0.2 to -0.5. However, Zhang et al. (2020d) note that excess cash may signal poor investment opportunities or higher idiosyncratic risk. Thus, while the negative sign supports financial theory, the magnitude reflects the moderate influence of liquidity compared to leverage.

The Return on Assets (ROA) show that higher profitability lowers default risk. This result is in line with the studies of Tserng et al. (2014b) and Qiu and Ronen (2025), which demonstrate that profitability remains a key determinant of financial stability. As ROA is a fundamental performance indicator, the strong negative relationship makes economic sense, highlighting how higher efficiency in asset utilisation reduces default probability; therefore, the coefficient of -0.4 has been set.

For Firm Size, the coefficient is –0.15 ($p < 0.01$), supporting prior research (Zhang et al., 2020e; Cathcart et al., 2020b) that larger firms are less vulnerable to default due to diversification, stronger market positions, and easier access to external finance. As there are marginal effects on the probabilities of default, which are different between large and small firms, range from 0.14 to 0.17. The magnitude suggests that while firm size matters, its effect is smaller compared to profitability and leverage.

Next, the Cash Flow to Debt ratio also exhibits a negative coefficient (–0.4), underscoring that stronger cash generation reduces default probability. This is consistent with Bhimani et al. (2014b), who note that firms with higher operating cash flows relative to debt enjoy greater repayment capacity, specifically, an increase in one unit in the cash to debt decreases the probability of default by approximately 30%. The large magnitude confirms that cash flow is one of the primary determinants of solvency, so the coefficient could not be low.

The ESG score has a positive coefficient of +0.02 ($p < 0.01$), reflecting its scaling where higher values indicate weaker performance. This aligns with Li, Zhang, and Zhao (2022c), who show that poor ESG standing increases long-term risk exposure, particularly ESG rating increases five times, leading to decreases three times, which means that one unit increase in ESG score raises default risk by 0.02. While the effect is modest compared to financial ratios, it is economically meaningful, as ESG factors play a secondary but significant role in creditworthiness.

Finally, Industry type was included as a categorical variable, though its effect is modest compared to firm-level financials. Manufacturing and Construction (-0.15) lower default risk due to their role in essential infrastructure and government support. Distribution (-0.1) also shows reduced risk, as strong logistics and supply chains provide resilience despite competition. Transportation (-0.04) modestly lowers risk, reflecting its importance in trade and connectivity. Food and Agriculture (-0.02) slightly reduce risk due to their necessity in all economies, though returns can be volatile. Services (-0.01) have the smallest effect, contributing modestly to stability.

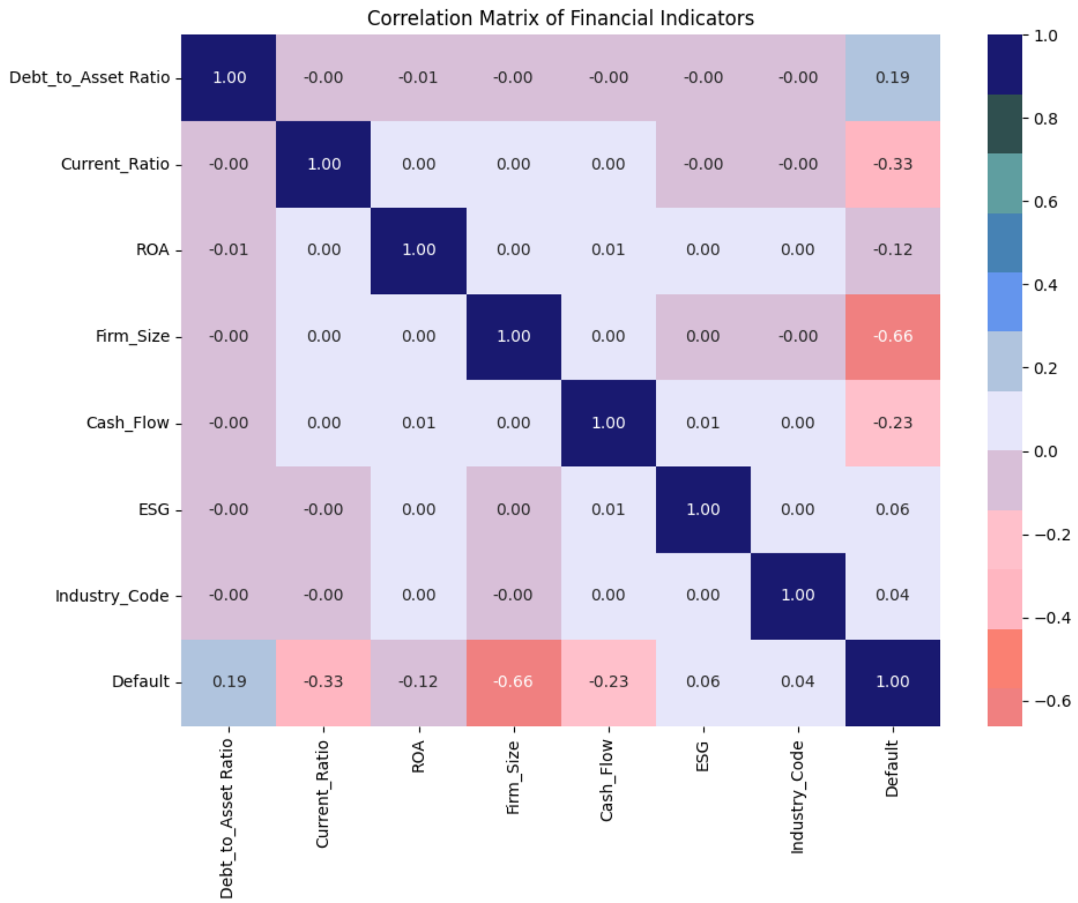


Figure 5: Correlation matrix between variables and 'Default'

## 5. Model Performance and Comparison

The Decision Tree classifier, using max_depth and random_state, achieved 88% accuracy for predicting default risk, which is considered a strong result for predicting default risk. For Class 0 (“No default”), precision was 88%, meaning that 88% of the outcomes were correct, while the recall was 87%, indicating that the model successfully identified 87% of all actual cases. For Class 1, the precision was 87%, while the recall was higher at 88%, showing that the model captured nearly all true “Default” cases. The difference between the two classes can be partly explained by the slight imbalance in the test data, with 10025 “No default” cases and 9975 “Default” cases, and by the natural trade-off between precision and recall. The model tends to predict “No default” only when it is highly confident, resulting in higher precision but lower recall for Class 0, while it predicts “Default” more often to avoid missing true defaults, which increases recall but reduces precision for Class 1. In terms of averages, the macro and weighted averages were both 88%, indicating stable performance across classes.

```python
clf = DecisionTreeClassifier(max_depth=5, random_state=42) # Create a Decision Tree Classifier (max_depth=4 restricts tree complexity)
clf.fit(x_train, y_train) # Train the model on the training data
y_pred = clf.predict(x_test) # Predict default status using the trained model on test data
accuracy_clf = accuracy_score(y_test, y_pred) # Evaluate accuracy of the model
print(f'model accuracy: {accuracy_clf:.2f}\n')
print('report:') # Detailed classification metrics (precision, recall, F1-score)
print(classification_report(y_test, y_pred))
```

```
model accuracy: 0.88

report:
              precision    recall  f1-score   support

           0       0.88      0.87      0.88     10025
           1       0.87      0.88      0.88      9975

    accuracy                           0.88     20000
   macro avg       0.88      0.88      0.88     20000
weighted avg       0.88      0.88      0.88     20000
```

Figure 6: Model performance of Decision Tree (before data scaling)

After conducting grid search, the most effective parameters were tuned to enhance accuracy. Optimal parameters were criterion = 'entropy', which improves class separation by splitting nodes based on information gain; max_depth = 15, limiting tree depth to prevent overfitting; min_samples_leaf = 10, ensuring each leaf has at least ten samples; min_samples_split = 2, allowing splits with at least two samples; and max_features = None, considering all features when splitting. After tuning, model accuracy rose to 91%. Class 0 (No Default) had 91% precision, 92% recall, and 91% F1-score. For Class 1 (Default), precision and recall also increased to 0.92 and 0.91 correspondingly. Overall, it indicates balanced performance across two classes.

```python
# Define parameter grid
param_grid = {
    'max_depth': [3, 5, 7, 10, 15, 20, None],
    'min_samples_split': [2, 5, 10, 20, 50],
    'min_samples_leaf': [1, 2, 4, 10, 20],
    'criterion': ['gini', 'entropy', 'log_loss'],
    'max_features': [None, 'sqrt', 'log2']}

# Create base model
dt = DecisionTreeClassifier(random_state=42)

# Grid search with cross-validation
grid_search = GridSearchCV(
    estimator=dt,
    param_grid=param_grid,
    cv=5,               # 5-fold cross-validation
    scoring='accuracy', # Optimizing for accuracy
    n_jobs=-1
)

# Fit on training data
grid_search.fit(x_train, y_train)

# Get the best model
best_dt = grid_search.best_estimator_
print("Best parameters:", grid_search.best_params_)

# Evaluate on test set
y_pred = best_dt.predict(x_test)
accuracy_clf = accuracy_score(y_test, y_pred)
print(f"Optimized Decision Tree Accuracy: {accuracy_clf:.2f}\n")
print("Classification Report:")
print(classification_report(y_test, y_pred))
```

```
Best parameters: {'criterion': 'entropy', 'max_depth': 15, 'max_features': None, 'min_samples_leaf': 10, 'min_samples_split': 2}
Optimized Decision Tree Accuracy: 0.91

Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.92      0.91     10025
           1       0.92      0.91      0.91      9975

    accuracy                           0.91     20000
   macro avg       0.91      0.91      0.91     20000
weighted avg       0.91      0.91      0.91     20000
```

Figure 7: Model Performance of Decision Tree (after data scaling)

Compared to the result of base line model, there are several mentioned modifications from the ranges while generating the analysis dataset, the addition of variables, adjustment of coefficients in financial indicators to estimate the probability of default risk, and changes of parameters for the training model. The current result indicates the superior accuracy performance, which is 91%, compared to 88% of the original result.

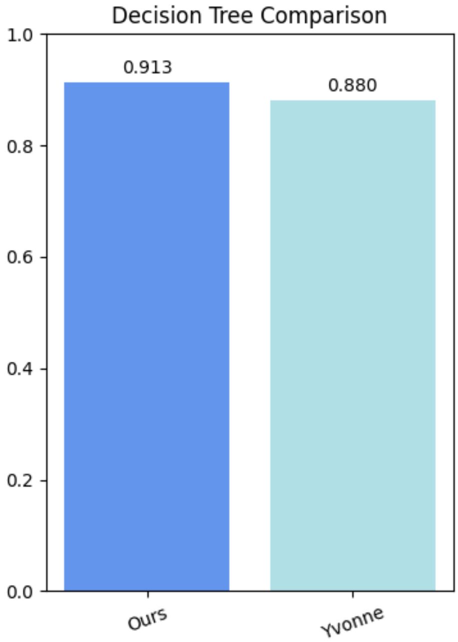


Figure 8: Current Model compared to Original Base Line Model

While the Decision Tree performed well, Logistic Regression (92%), Random Forest (91.5%), and SVC (92%) showed slightly better accuracy. Logistic Regression offers stability, generalisability, and useful probability outputs. Random Forest, by combining multiple Decision Trees, reduces overfitting and handles both linear and non-linear patterns, while SVC with the RBF kernel captures complex boundaries in higher dimensions. Overall, these models outperform the Decision Tree in stability, robustness, and flexibility.

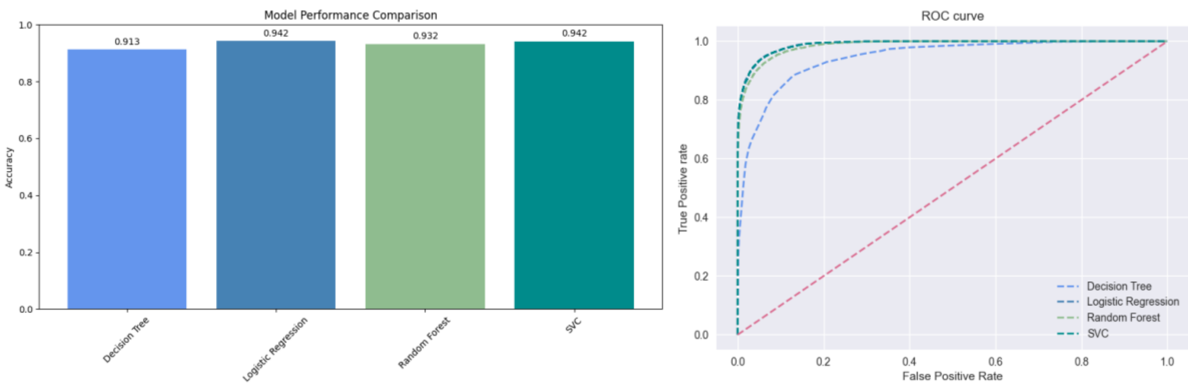


Figure 9: Comparison between Decision Tree and other ML models ($R^2$ & ROC curve)

## 6. Limitations and Recommendations

When variables have high correlations, the model may overfit. It then has good results on the training set but poor results on new data. This problem is worse because the model does not use methods like cross-validation and pruning (Löffler & Maurer, 2011d). Also, the use of simulated data lowers robustness. This is because values can be outside real ranges and there is no volatility. As a result, the model cannot show the differences found in the real world (Bajwa & Rashid, 2021b). Another problem is the small use of macroeconomic factors. Empirical studies show that recession and lower GDP growth raise the chance of default. They do this by changing revenues, refinancing costs, and cash flow buffers (Memmel, Gündüz, & Raupach, 2015b; Zhang, Ouyang, Liu, & Xu, 2020f).

To solve these issues, first, adding macroeconomic indicators like GDP growth can help show systemic risk (Cathcart, Dufour, Rossi, & Varotto, 2020c). Second, using ensemble learning methods like gradient boosting can lower overfitting and improve stability (Chan-Lau et al., 2024b). Third, using ESG ratings and real-time transaction data can give quick signals of solvency, though data availability is still a problem (Li, Zhang, & Zhao, 2022c).

## References

Bajwa, A. M. A., & Rashid, M. (2021a). Financial ratios: A tool for computing probability of corporate default. *Audit and Accounting Review, 1*(2), 71–90.

Bajwa, A. M. A., & Rashid, M. (2021b). Financial ratios: A tool for computing probability of corporate default. *Audit and Accounting Review, 1*(2), 71–90.

Benzschawel, T., Haroon, A., & Wu, T. (2011). A model for recovery value in default. *Journal of Fixed Income*, *21*(2), 15

Bhimani, A., Gulamhussen, M. A., & da Rocha Lopes, S. (2014a). Owner liability and financial reporting information as predictors of firm default in bank loans. *Review of Accounting Studies*, *19*(2), 769-804. https://doi.org/10.1007/s11142-013-9269-0

Bhimani, A., Gulamhussen, M. A., & da Rocha Lopes, S. (2014b). Owner liability and financial reporting information as predictors of firm default in bank loans. *Review of Accounting Studies*, *19*(2), 769-804. https://doi.org/10.1007/s11142-013-9269-0

Cathcart, L., Dufour, A., Rossi, L., & Varotto, S. (2020a). The differential impact of leverage on the default risk of small and large firms. *Journal of Corporate Finance, 60*, 101541. https://doi.org/10.1016/j.jcorpfin.2019.101541

Cathcart, L., Dufour, A., Rossi, L., & Varotto, S. (2020b). The differential impact of leverage on the default risk of small and large firms. *Journal of Corporate Finance*, *60*, 101541. https://doi.org/10.1016/j.jcorpfin.2019.101541

Cathcart, L., Dufour, A., Rossi, L., & Varotto, S. (2020c). The differential impact of leverage on the default risk of small and large firms. *Journal of Corporate Finance*, *60*, 101541. https://doi.org/10.1016/j.jcorpfin.2019.101541

Chan-Lau, J. A., Hu, R., Mungo, L., Qu, R., Xin, W., & Zhong, C. (2024a). Monitoring privately-held firms' default risk in real time: A signal-knowledge transfer learning model (IMF Working Paper No. WP/24/115). International Monetary Fund.

Chan-Lau, J. A., Hu, R., Mungo, L., Qu, R., Xin, W., & Zhong, C. (2024b). Monitoring privately-held firms' default risk in real time: A signal-knowledge transfer learning model (IMF Working Paper No. WP/24/115). International Monetary Fund.

Charalambakis, E., & Garrett, I. (2019). On corporate financial distress prediction: What can we learn from private firms in a developing economy? *Review of Quantitative Finance and Accounting, 52*(1), 59–87. https://doi.org/10.1007/s11156-018-0728-6

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of applied science and technology trends*, *2*(01), 20-28.

Grammenos, C. T., Nomikos, N. K., & Papapostolou, N. C. (2008a). Estimating the probability of default for shipping high yield bond issues. *Transportation Research Part E: Logistics and Transportation Review*, *44*(6), 1123-1138. https://doi.org/10.1016/j.tre.2007.10.005

Grammenos, C. T., Nomikos, N. K., & Papapostolou, N. C. (2008b). Estimating the probability of default for shipping hi-h yield bond issues. *Transportation Research Part E: Logistics and Transportation Review*, *44*(6), 1123-1138. https://doi.org/10.1016/j.tre.2007.10.005

Li, H., Zhang, X., & Zhao, Y. (2022a). ESG and firm’s default risk. *Finance Research Letters, 47*, 102713. https://doi.org/10.1016/j.frl.2022.102713

Li, H., Zhang, X., & Zhao, Y. (2022b). ESG and firm's default risk. *Finance Research Letters*, *47*, 102713. https://doi.org/10.1016/j.frl.2022.102713

Li, H., Zhang, X., & Zhao, Y. (2022c). ESG and firm's default risk. *Finance Research Letters*, *47*, 102713. https://doi.org/10.1016/j.frl.2022.102713

Löffler, G., & Maurer, A. (2011). *Credit risk modeling using Excel and VBA* (2nd ed.). Chichester: Wiley.

Löffler, G., & Maurer, A. (2011a). Incorporating the dynamics of leverage into default prediction. *Journal of banking & finance*, *35*(12), 3351-3361. https://doi.org/10.1016/j.jbankfin.2011.05.015

Löffler, G., & Maurer, A. (2011b). Incorporating the dynamics of leverage into default prediction. *Journal of banking & finance*, *35*(12), 3351-3361. https://doi.org/10.1016/j.jbankfin.2011.05.015

Löffler, G., & Maurer, A. (2011c). Incorporating the dynamics of leverage into default prediction. *Journal of banking & finance*, *35*(12), 3351-3361. https://doi.org/10.1016/j.jbankfin.2011.05.015

Löffler, G., & Maurer, A. (2011d). Incorporating the dynamics of leverage into default prediction. *Journal of banking & finance*, *35*(12), 3351-3361. https://doi.org/10.1016/j.jbankfin.2011.05.015

Lopez, J. A. (2004). The empirical relationship between average asset correlation, firm probability of default, and asset size. *Journal of Financial Intermediation*, *13*(2), 265-283. https://doi.org/10.1016/S1042-9573(03)00045-7

Memmel, C., Gündüz, Y., & Raupach, P. (2015a). The common drivers of default risk. *Journal of Financial Stability*, *16*, 232-247. https://doi.org/10.1016/j.jfs.2014.03.002

Memmel, C., Gündüz, Y., & Raupach, P. (2015b). The common drivers of default risk. *Journal of Financial Stability*, *16*, 232-247. https://doi.org/10.1016/j.jfs.2014.03.002

Nadarajah, S., Duong, H. N., Ali, S., Liu, B., & Huang, A. (2021). Stock liquidity and default risk around the world. *Journal of Financial Markets*, *55*, 100597. https://doi.org/10.1016/j.finmar.2020.100597

Qiu, L., & Ronen, J. (2025). Anticipatory effects of accounting standards: the lease exposure draft: L. Qiu, J. Ronen. *Review of Accounting Studies*, *30*(2), 1543-1591. https://doi.org/10.1007/s11142-024-09861-5

Tserng, H. P., Chen, P. C., Huang, W. H., Lei, M. C., & Tran, Q. H. (2014a). Prediction of default probability for construction firms using the logit model. *Journal of civil engineering and management*, *20*(2), 247-255. https://doi.org/10.3846/13923730.2013.801886

Tserng, H. P., Chen, P. C., Huang, W. H., Lei, M. C., & Tran, Q. H. (2014b). Prediction of default probability for construction firms using the logit model. *Journal of civil engineering and management*, *20*(2), 247-255. https://doi.org/10.3846/13923730.2013.801886

Westgaard, S., & Van der Wijst, N. (2001a). Default probabilities in a corporate bank portfolio: A logistic model approach. *European journal of operational research*, *135*(2), 338-349. https://doi.org/10.1016/S0377-2217(01)00045-5

Westgaard, S., & Van der Wijst, N. (2001b). Default probabilities in a corporate bank portfolio: A logistic model approach. *European journal of operational research*, *135*(2), 338-349. https://doi.org/10.1016/S0377-2217(01)00045-5

Zhang, X., Ouyang, R., Liu, D., & Xu, L. (2020a). Determinants of corporate default risk in China: The role of financial constraints. *Economic Modelling*, *92*, 87-98. https://doi.org/10.1016/j.econmod.2020.07.005

Zhang, X., Ouyang, R., Liu, D., & Xu, L. (2020b). Determinants of corporate default risk in China: The role of financial constraints. *Economic Modelling*, *92*, 87-98. https://doi.org/10.1016/j.econmod.2020.07.005

Zhang, X., Ouyang, R., Liu, D., & Xu, L. (2020c). Determinants of corporate default risk in China: The role of financial constraints. *Economic Modelling*, *92*, 87-98. https://doi.org/10.1016/j.econmod.2020.07.005

Zhang, X., Ouyang, R., Liu, D., & Xu, L. (2020d). Determinants of corporate default risk in China: The role of financial constraints. *Economic Modelling*, *92*, 87-98. https://doi.org/10.1016/j.econmod.2020.07.005

Zhang, X., Ouyang, R., Liu, D., & Xu, L. (2020e). Determinants of corporate default risk in China: The role of financial constraints. *Economic Modelling*, *92*, 87-98. https://doi.org/10.1016/j.econmod.2020.07.005

Zhang, X., Ouyang, R., Liu, D., & Xu, L. (2020f). Determinants of corporate default risk in China: The role of financial constraints. *Economic Modelling*, *92*, 87-98. https://doi.org/10.1016/j.econmod.2020.07.005