Finance and Risk Management of Financial Institution

Project A: Credit Rating Analysis

Classification of credit rating

1. Summary

The credit rating of a company is an evaluation based on the assessment of the company's financial condition and credit risk. Credit ratings can impact the company's cost of financing and investor trust, providing an objective standard.

In this study, we first collected over 60 variables that may influence a company's credit rating through a review of literature. Subsequently, we utilized modern machine learning models such as XGBoost, Random Forest and Support Vector Machine, which have factor selection capabilities, along with traditional classification models like Naive Bayes and Logistic Regression. We also employed a sequential feature selector (stepwise regression) to further filter factors. By evaluating the performance of ten different models, we identified the most promising ones for predicting credit ratings. We then examined the features utilized by these models to infer the most predictive characteristics for company credit ratings.

2. Data and Sample

2.1 Full picture of the data

Table 1: Variables and its Types

Type	Variables
Liquidity	- Current Ratio - Quick Ratio - Cash Ratio - Days of Sales Outstanding - Cash Conversion
Profitability	- Gross Profit Margin - Operating Profit Margin - Pre-tax Profit Margin

	- Net Profit Margin - Effective Tax Rate - Return on assets - Return on Equity - Return on Capital Employed - Pre-tax Profit Margin - Pre-tax return on Net Operating Assets - After-tax Return on Average Common Equity - After-tax Return on Invested Capital - After-tax Return on Total Stockholders' Equity - Gross Profit/Total Assets
Cash Flow Indicator Ratios	- Operating Cash Flow Per Share - Free Cash Flow Per Share - Cash Per Share - Operating Cash Flow Sales Ratio - Free Cash Flow Operating Cash Flow Ratio
Efficiency	- Asset Turnover - Fixed Asset Turnover - Payables Turnover - Inventory Turnover - Sales/Invested Capital - Sales/Stockholders Equity
Valuation	- Book to Market Ratio - Shillers Cyclically Adjusted P/E Ratio - Enterprise Value Multiple - Price/Operating Earnings (Basic, Excl. EI) - Price/Operating Earnings (Diluted, Excl. EI) - Dividend Yield - Company Equity Multiplier - Enterprise Value Multiplier
Solvency	- Total Debt to Equity Ratio - Total Debt to Total Assets Ratio - Total Liability to Total Assets Ratio - Total Debt to Capital Ratio - After-tax Interest Coverage

	- Interest Coverage Ratio - Debt to Equity Ratio
Capitalization	- Common Equity/Invested Capital - Long-term Debt/Invested Capital - Total Debt/Invested Capital - Capitalization Ratio
Financial Soundness	- Cash Flow Margin - Inventory/Current Assets - Receivables/Current Assets - Free Cash Flow/Operating Cash Flow - Operating CF/Current Liabilities - Cash Flow/Total Debt - Cash Balance/Total Liabilities - Short-Term Debt/Total Debt - Profit Before Depreciation/Current Liabilities - Current Liabilities/Total Liabilities - Total Debt/EBITDA - Long-term Debt/Book Equity - Interest/Average Long-term Debt - Interest/Average Total Debt - Long-term Debt/Total Liabilities - Total Liabilities/Total Tangible Assets
Others	- Accruals/Average Assets - Research and Development/Sales - Advertising Expenses/Sales - Labor Expenses/Sales - EBIT Per Revenue

Table 2: Ratings and their respective number of samples

Rating	Number of Sample
BBB+	1189
BBB	1186
A	759
A+	714
A-	711
BBB-	504

AA-	331
BB+	214
AA	178
BB	173
AA+	61
BB-	46
AAA	26

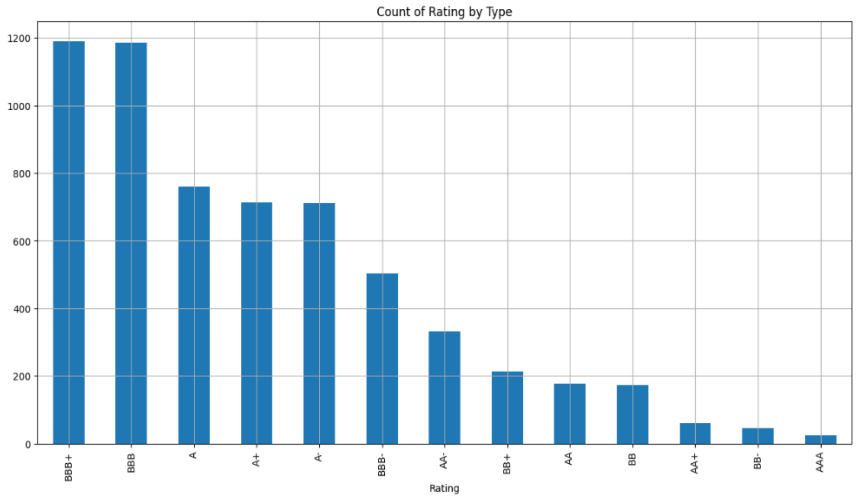

Figure 1: Ratings and their respective number of samples (bar chart)

2.2 Data cleaning.

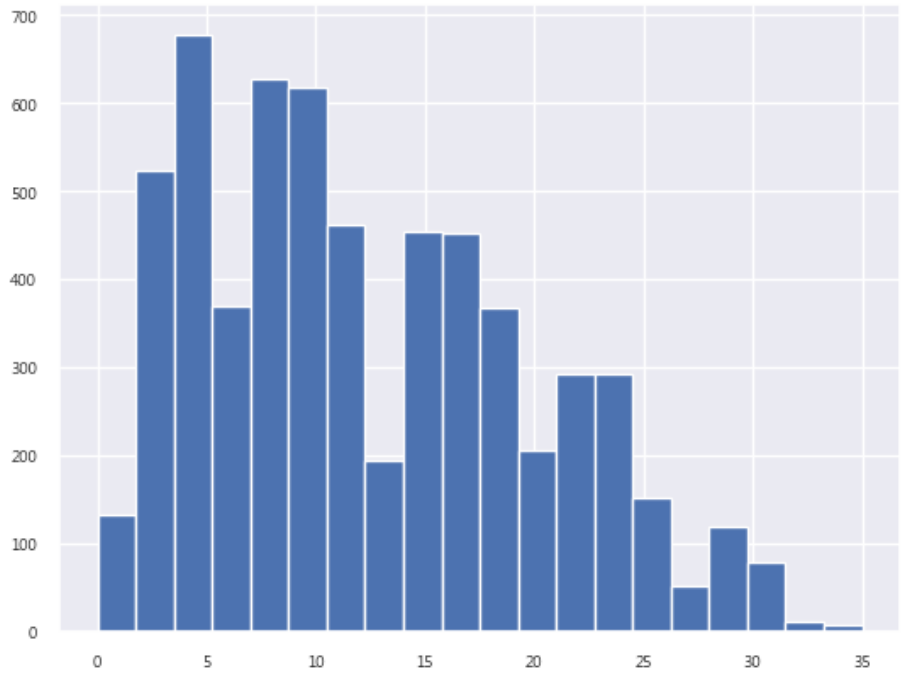

Figure 2: Feature number of outlier in each sample

The horizontal axis in the chart represents the number of outlier variables in a single sample, while the vertical axis represents the sample size. Therefore, based on the chart, it can be observed that the majority of samples contain outliers, yet these outliers may still carry meaningful information. Thus, we would only transform them rather than outright remove them.

Our approach involves using the 'MinMaxScaler' from scikit-learn to scale the numerical values in the data columns, followed by applying a logarithmic transformation using ‘NumPy’.

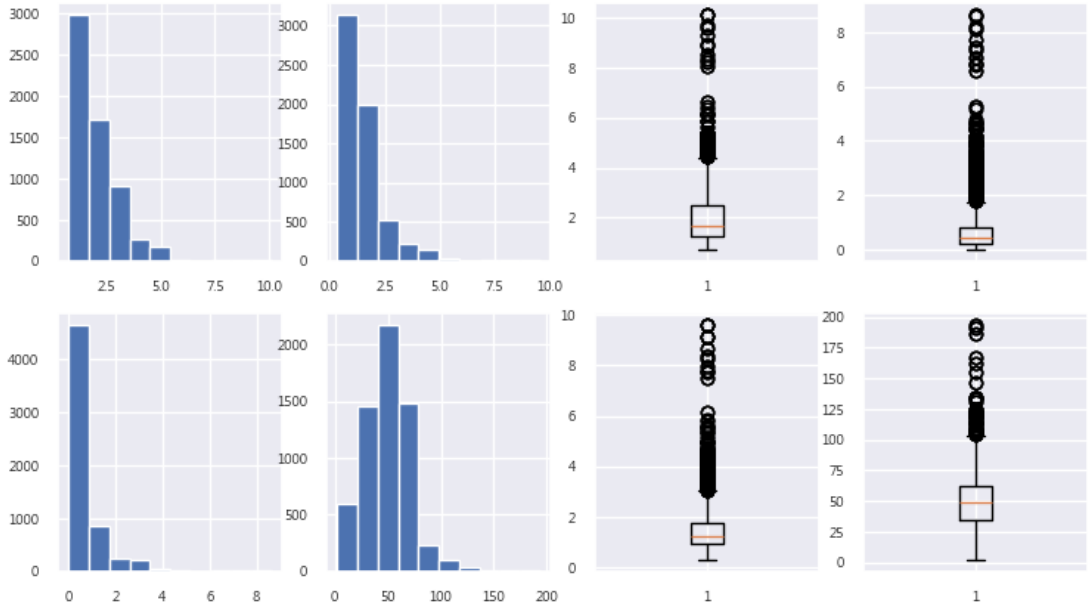

Figure 3: Sample distribution before transformation

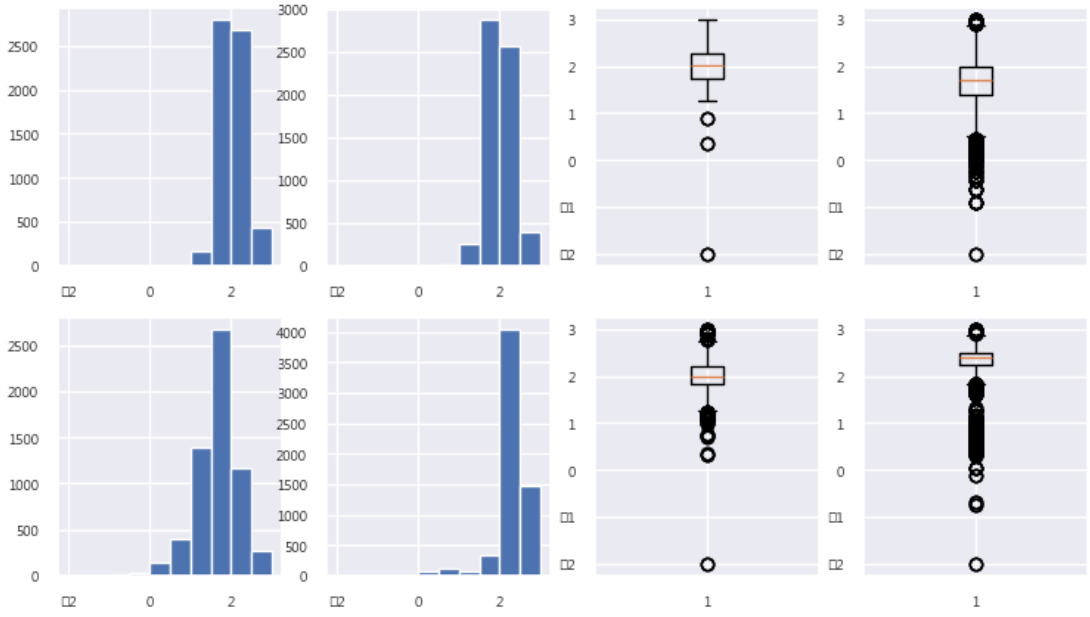

Figure 4: Sample distribution after transformation

We can observe that after we transform the data, extreme outliers still exist. Thus, we need to remove the remaining outlier after the transformation.

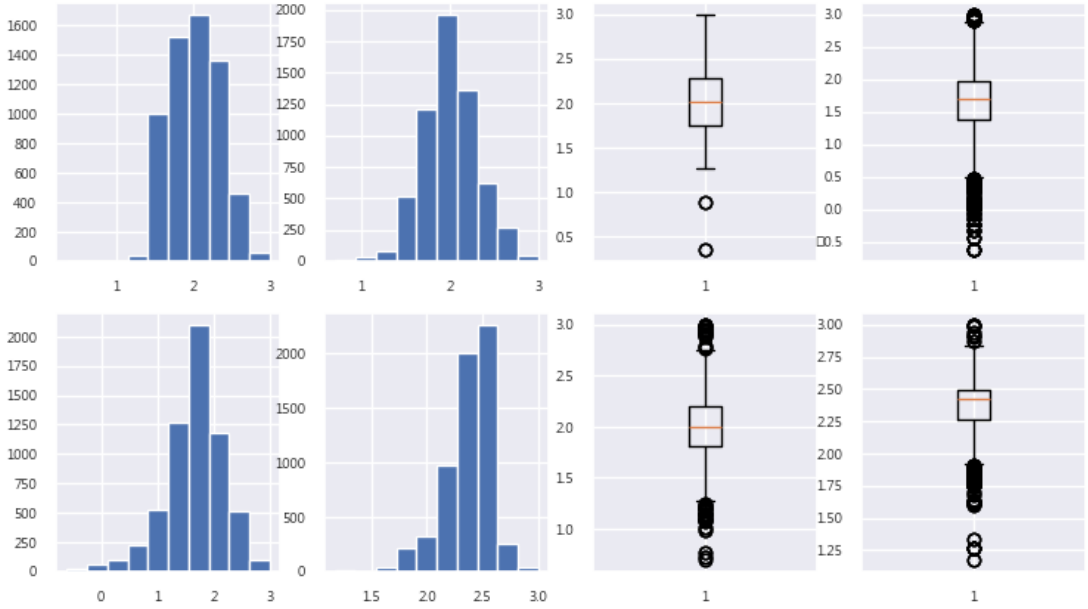

Figure 5: Sample distribution after deleting outliers

2.3. Examine whether the correlation between the sample characteristics and its grading exist.

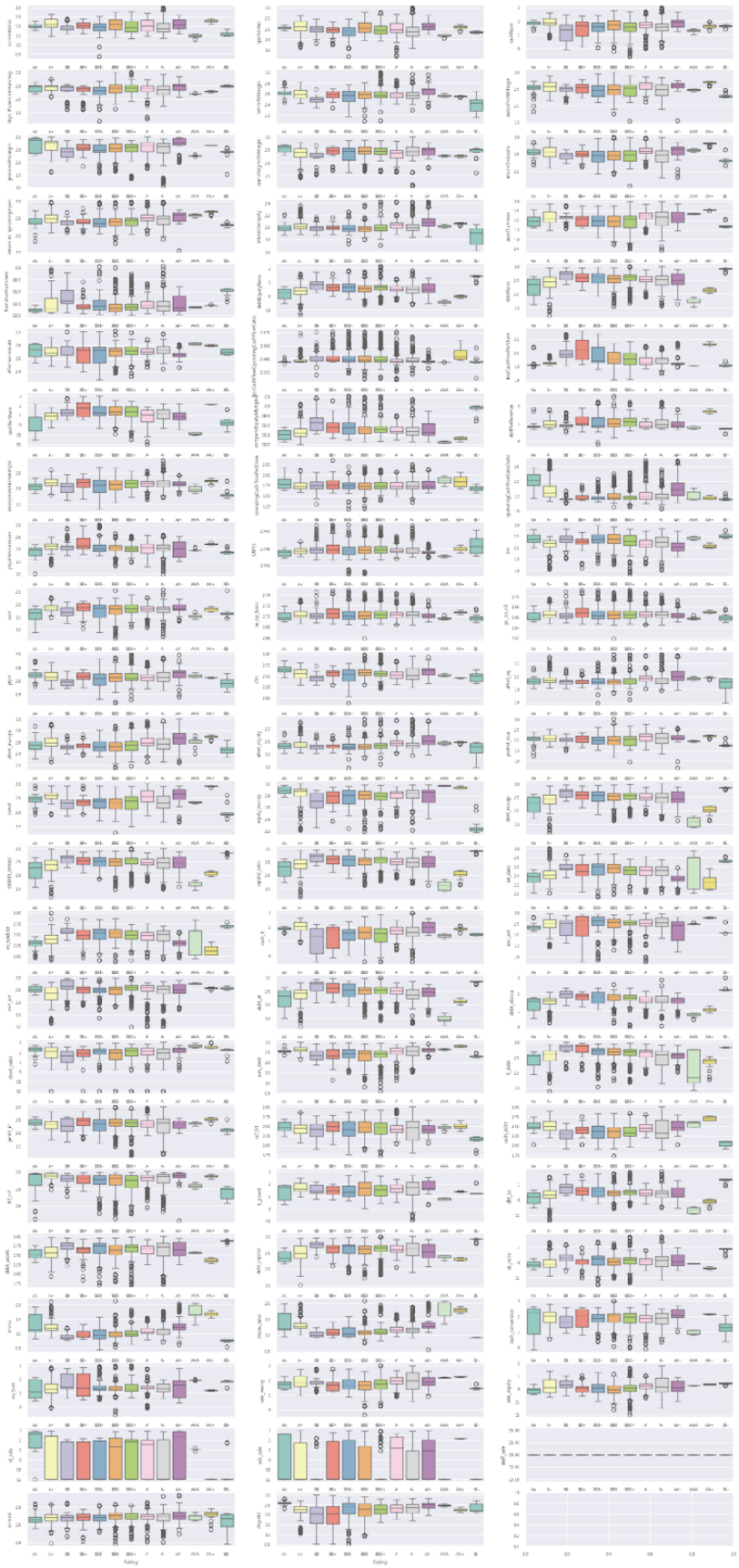

Figure 6: Correlation between the sample characteristics and its grading

According to the chart, we can observe that the mean of some features were correlated with rating.

3. Methodology

3.1 Introduction to Traditional Regression and Machine Learning Methods

We use numerous machine learning method to build the model:

- XGBoost

XGBoost is an ensemble learning method that utilizes a gradient boosting framework, where weak learners (usually decision trees) are trained sequentially, with each subsequent tree correcting errors made by the previous ones.

- Random Forest

Random Forest is an ensemble learning method that builds multiple decision trees during training and outputs the mode (classification) or mean prediction (regression) of the individual trees.

- Gradient Boosting

Gradient Boosting is an ensemble learning technique that builds a series of weak learners (often decision trees) sequentially. Each tree corrects errors made by the previous ones, leading to improved overall performance.

- K Nearest Neighbors

KNN is a simple and intuitive classification algorithm that assigns a class label to an input based on the majority class of its k nearest neighbors in the feature space.

- Linear Discriminant Analysis

LDA is a classification and dimensionality reduction technique that finds linear combinations of features to characterize or separate classes.

- Support Discriminant Analysis

SVM is a supervised learning algorithm that finds a hyperplane to separate classes in the feature space. It is effective in high-dimensional spaces and can handle non-linear relationships through kernel functions.

- Navie Bayes

Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem, assuming independence between features given the class label.

- Quadratic Discriminant Analysis

QDA is similar to LDA but does not assume equal covariance matrices for different classes, allowing for more flexibility in modeling.

- Logistic Regression

Despite its name, logistic regression is a classification algorithm used to model the probability of a binary outcome using the logistic function.

- Neural Network

Neural networks, particularly deep neural networks, are a class of algorithms inspired by the structure and function of the human brain. They consist of interconnected nodes organized into layers.

Models we use:

```
('XGBoost', XGB_model),
    ('Random Forest', RF_model),
    ('Gradient Boosting', GBT_model),
    ('K Nearest Neighbours', KNN_model),
    ('Linear Discriminant Analysis',
LDA_model),
    ('Support Vector Machine', SVC_model),
    ('Naive Bayes', GNB_model),
('Quadratic Discriminant Analysis',
QDA_model),
    ('Logistic Regression', LR_model),
    ('Neural Network', MLP_model)
```

Figure 7: The code of model

3.2 Model Performance Evaluation

- Confusion Matrix:

A Confusion Matrix is a table used in machine learning to evaluate the performance of a classification model. It presents the correspondence between the model's predicted results and the actual outcomes in matrix form. The Confusion Matrix is primarily used to measure the accuracy, precision, and other performance metrics of a model. The horizontal axis represents the predicted values, indicating the results predicted by the model, while the vertical axis represents the actual values observed in the data. The diagonal line represents the number of correctly predicted samples. A higher proportion of samples on the diagonal indicates more accurate predictions. The numbers in the matrix represent the count of correctly predicted instances.

- Feature Importance Ranking :

This table displays the explanatory power of different variables in the same model. The higher the ranking, the more important the factor is to the model.
We will first assess the predictive accuracy of 10 machine learning methods. The top three methods in terms of accuracy will be further discussed using the Confusion Matrix and Feature Importance Ranking Table to examine the effectiveness of machine learning

3.3 Ensemble Learning Methods

After examining the effectiveness of machine learning methods, we further enhance the models by combining them using two ensemble learning methods: Voting and Stacking. Through these approaches, we aim to increase the interpretability of the models. Additionally, we will explore whether incorporating all models into ensemble learning or using only the top three effective models yields superior predictive capabilities.

- **Voting**

Voting is an ensemble learning method commonly applied to classification problems. It combines the predictive results of multiple independent models, and the final prediction is determined by the majority vote of these models. There are two main implementations of voting.

- Hard Voting: In hard voting, each model casts a vote, and the final prediction is the class that receives the most votes. This method is prevalent in classification problems.
- Soft Voting: Soft voting considers the weighted probabilities of model predictions, and the final prediction is the weighted average of the predicted probabilities from all models. This method is more suitable for models that support probability predictions.

- **Stacking**

Stacking is another ensemble learning method that typically involves a combination of multiple hierarchical models. In stacking, the predictions of different base models serve as inputs, and a secondary model (meta-model) is used for the final prediction. The steps of stacking are as follows

- Stage One: Train data using multiple base models to obtain their predictions.
- Stage Two: Use the predictions of the base models as inputs and the true labels as outputs to train a secondary model (meta-model).
- Stage Three: Use the stacked model for predictions.

The advantage of stacking lies in its ability to capture the strengths of different base models, enhancing the overall performance of the model. However, its design and implementation are relatively complex, requiring more computational resources.

4. Empirical result

4.1 Model Performance Evaluation

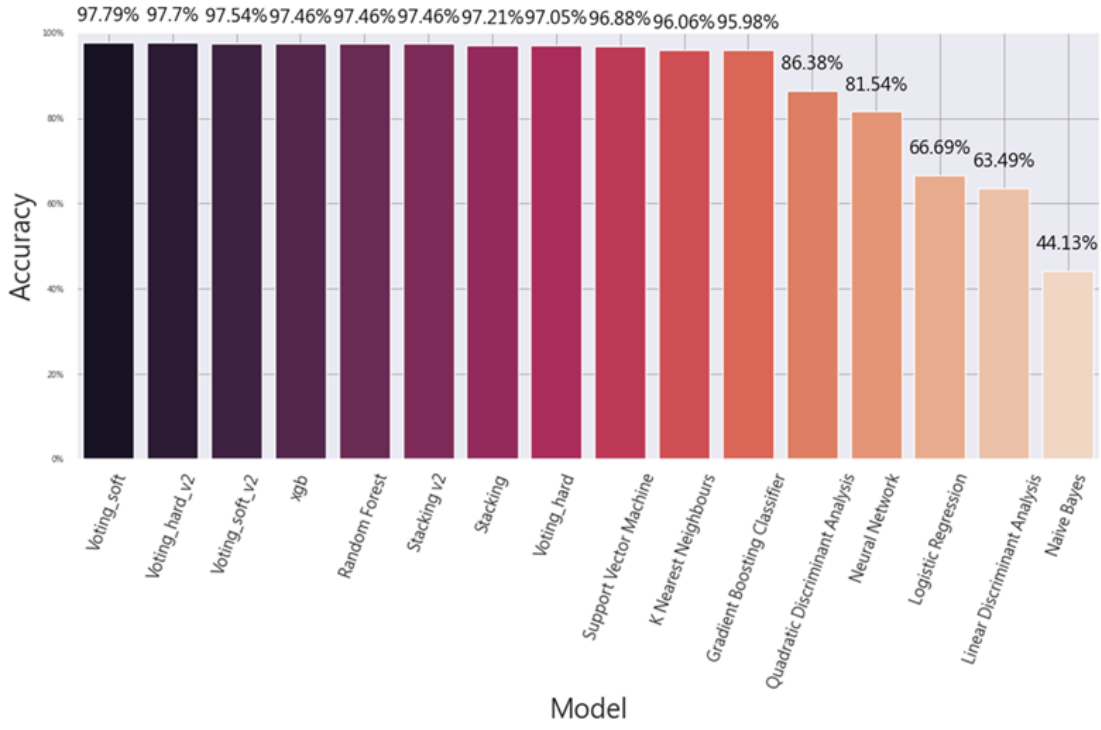

Figure 8: Accuracy of each model

We first assess the predictive accuracy of all machine learning methods. According to the chart, Random Forest, XGB, and Support Vector Machine exhibit the highest accuracy in predicting company credit ratings. Therefore, we will delve into the discussion of these three methods in the following sections. We will use confusion matrices and feature importance ranking tables to examine the effectiveness of machine learning.

- Random Forest

The model accuracy of Random Forest reaches an impressive 97.46%. The most influential factors in the model are Free Cash Flow Per Share, Cash Per Share, It_ppent (Long-Term Asset Turnover), Intcov_ratio (Interest Coverage Ratio), and EBIT Per Revenue (Earnings Before Interest and Taxes to Revenue Ratio).

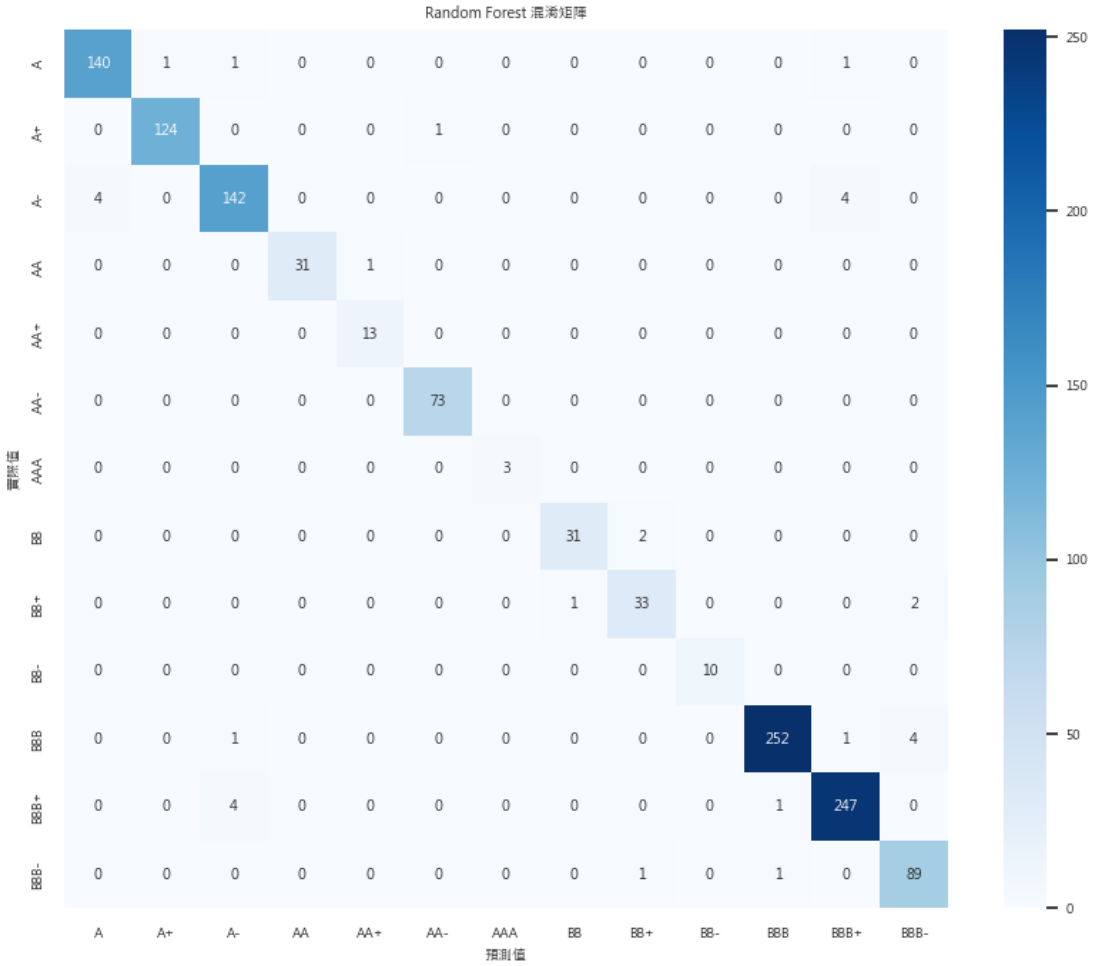

Figure 9: Random Forest Confusion Matrix (RF Accuracy 97.46%)

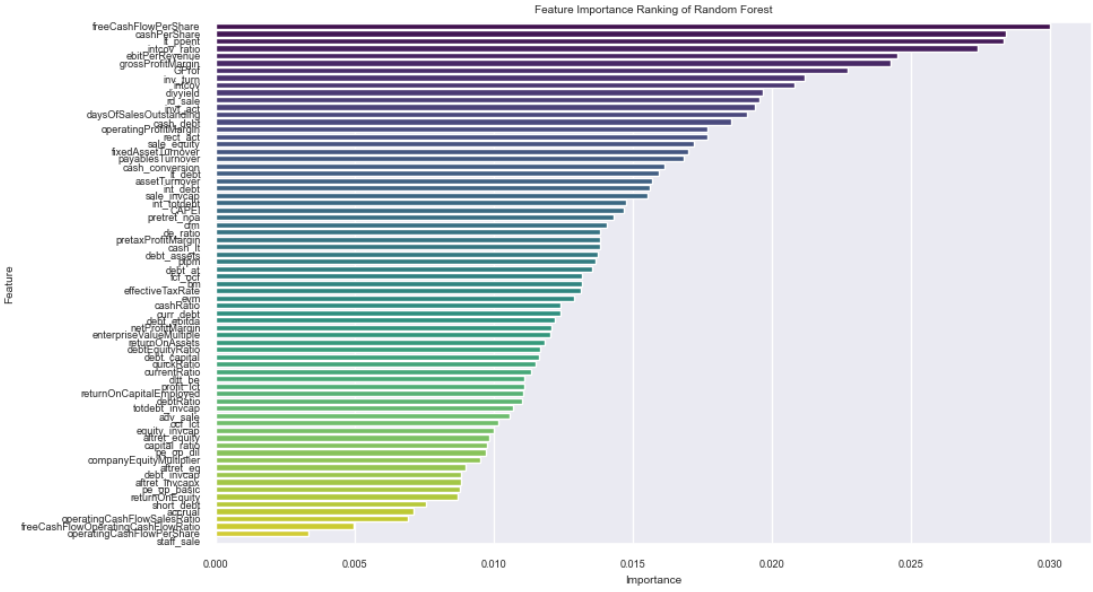

Figure 10: Random Forest Feature Importance Ranking

- XGB

The model accuracy of XGB is also 97.46%. The most influential factors in the model are Intcov_ratio (Interest Coverage Ratio), Current Ratio, EVM (Enterprise Value Multiple), rd_sales (Research and Development to Sales Ratio), and Fixed Assets Turnover.

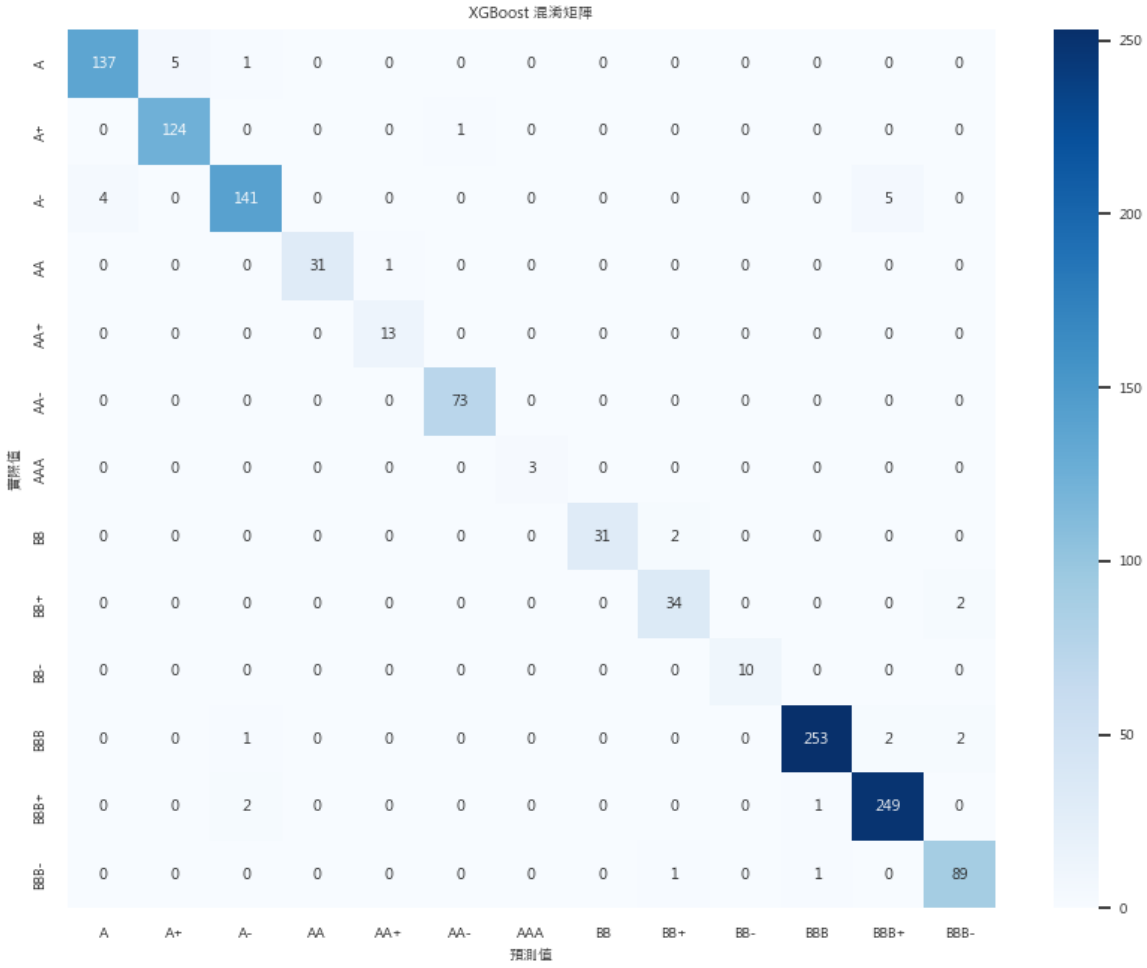

Figure 11: XGB Confusion Matrix (XGB Accuracy 97.46%)

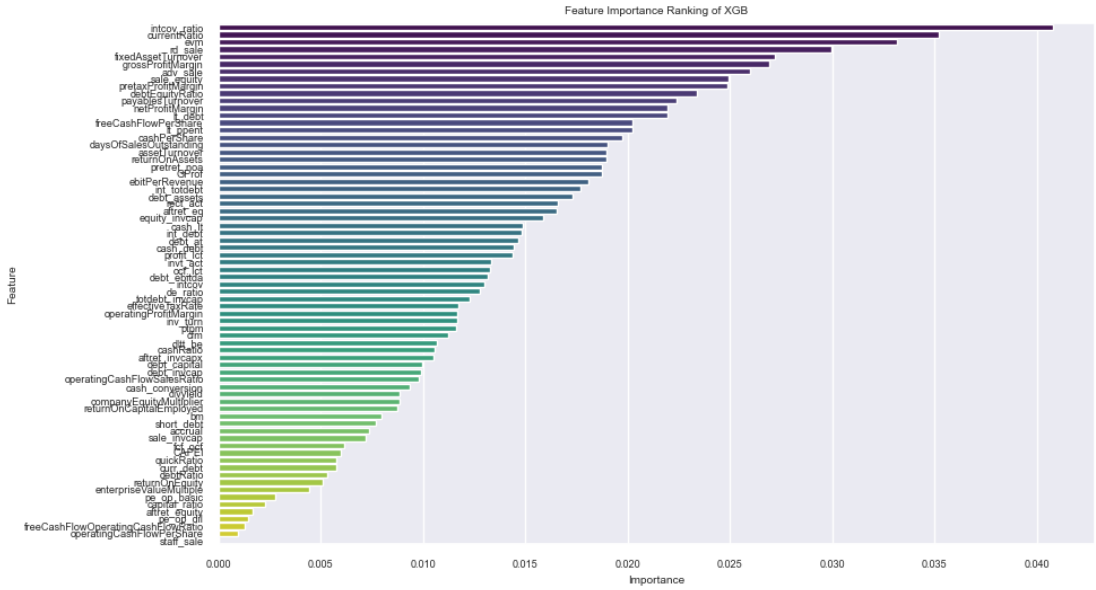

Figure 12: XGB Feature Importance Ranking

- Support Vector Machine

The accuracy of the SVM model is 96.88%. However, in Support Vector Machines (SVM), the decision boundary of the model is determined by support vectors (points in the training data closest to the decision boundary). The model's focus is usually between support vectors rather than on individual features. Therefore, SVM, in its original form, does not provide a direct assessment of feature importance.

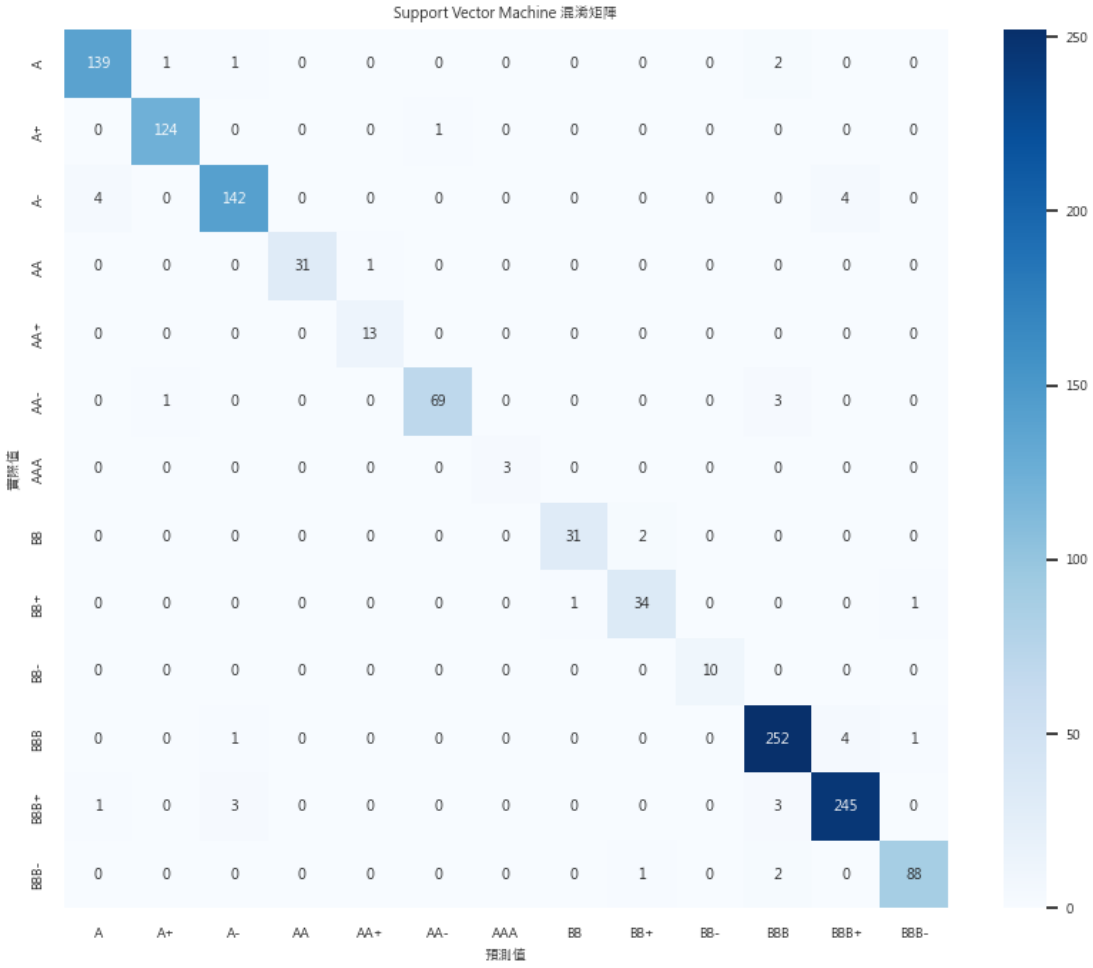

Figure 13: SVM Confusion Matrix (SVM Accuracy 96.88%)

4.2 Ensemble Learning Methods

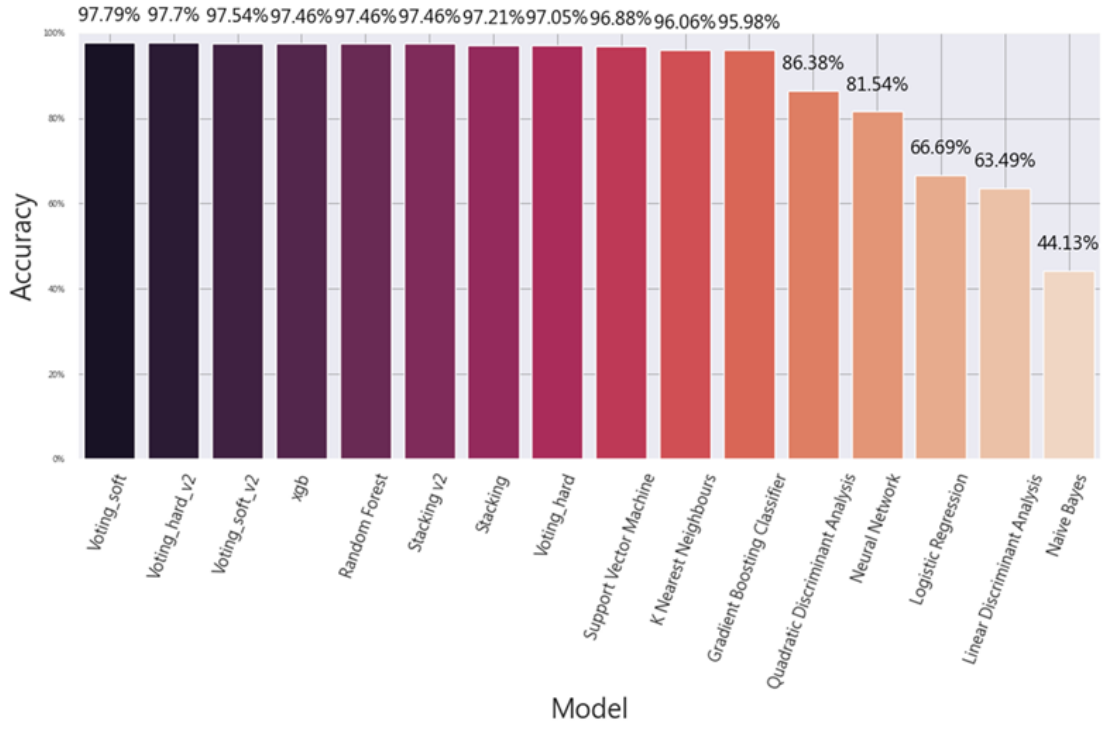

Figure 14: Accuracy of each model

From the above chart, we can observe a comparison of predictive accuracy among different models. Among the three ensemble learning methods (Voting_hard, Voting_soft, Stacking), they are further divided into forms that include only the top three effective models (V2) and forms that include all models, discussing their differences.

Overall, after adding the Voting and Stacking ensemble learning methods, Voting_soft, Voting_hard_v2, and Voting_soft_v2 are the top three models with the highest interpretability. In conclusion, we derive two findings from the above results:

(1) In Voting_soft, the interpretability of including all models is superior to the results of merging with only three models.
(2) In Voting_hard and Stacking, the interpretability of using only three models is superior to the results of merging with all models."

5. Conclusion

5.1 Model Performance Evaluation

(i) Among the ten machine learning methods, Random Forest, XGB, and SVM have the highest predictive abilities, with accuracies of 97.46%, 97.46%, and 96.88%, respectively.

(ii) In Voting_soft, the interpretability of including all models is superior to the results of merging with only the top three effective models

(iii) In Voting_hard and Stacking, the interpretability of using only the top three effective models is superior to the results of merging with all models.

5.2 Factor Importance

According to XGB and Random Forest, the top five features in terms of importance are as follows:

- **Interest Coverage Ratio**

The interest coverage ratio is a debt and profitability ratio used to determine how easily a company can pay interest on its outstanding debt. The interest coverage ratio is calculated by dividing a company's earnings before interest and tax(EBIT) by its interest expense during a given period.

- **Free Cash Flow Per Shares**

Free cash flow per share (FCF) is a measure of a company's financial flexibility that is determined by dividing free cash flow by the total number of shares outstanding.

- **Total Liability to Net Tangible Asset Ratio**

The ratio of total liabilities to net tangible assets measures the security provided to all creditors, not just long-term, by the firm's more readily realizable assets. The higher ratio indicates that the company has lost some ground with respect to covering all its debts with net tangible assets.

- **Gross Profit Margin**

Gross profit margin refers to a financial metric that analysts use to assess a company's financial health. Gross profit margin is the profit after subtracting the cost of goods sold. Put simply, a company's gross profit margin is the money it makes after accounting for the cost of doing business. This metric is commonly expressed as a percentage of sales and may also be known as the gross margin ratio.

- **Cash per Shares**

Cash per share is the broadest measure of available cash to a business divided by the number of equity shares outstanding. Cash per share tells us the percentage of a company's share price available to spend on strengthening the business, paying down debt, returning money to shareholders, and other positive campaigns.

6. Reference :

(1) 官顥（2022），運用機器學習模型於公司信用評等預測效力之研究：公司年報文本特徵之新證據，未出版碩士論文，國立陽明交通大學管理科學研究所，台灣新竹。

(2) 詹佩俞（2019），機器學習法應用於企業信用評等之預測，未出版碩士論文，國立中正大學會計與資訊科技研究所，台灣嘉義。

(3) Parisa G., Ionut F., Rupak C. (2020). A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees

(4) Pu Liu, Fazal J. Seyyed and Stanley D. Smith (1999). The Independent Impact of Credit Rating Changes ± The Case of Moody's Rating Refinement on Yield Premiums.

(5) Chen, K. H., & Shimerda, T. A. (1981). An empirical analysis of useful financial ratios. Financial Management, 10-17.