## 

Bankruptcy Prediction of the

Companies in Taiwan Stock Exchange

from 1999 to 2009

**QTM2000**

Professor Pachamanova

December 6, 2021

**Written By:**

Altug Kalfazade

Efe Bora

Linh Ha

Ziya Aydin

TABLE OF CONTENTS

[**Executive Summary**](#_ayql39unkfaw) **2**

[**Research Problem and Data**](#_keyjanuiesu7) **3**

[**Significance of the Predictors**](#_k687c04u7y1d) **4**

[**Unsupervised Learning: Clusters**](#_1ez6eu5okrtl) **5**

[**Logistics Regression**](#_5f0ci0zfjqpg) **6**

[**K-Nearest Neighbors (KNN)**](#_luryy8ii2inb) **7**

[**Classification Trees**](#_7if64kbjvfoa) **9**

[**Optimal Model**](#_p2b9uvs6c9z) **10**

[**Findings and Recommendations**](#_gyobi5hguy5n) **10**

[**Data Findings**](#_4fad488i72oh) **10**

[**Business Findings**](#_22wapo873oa9) **11**

[**Appendix**](#_5lmngsr5l1k7) **12**

[**Works Cited**](#_m8y5exqjnlss) **28**

## 

## **Executive Summary**

We, DataBusters Consulting, have analyzed the financial ratios of 6,819 companies that our client, Dessislava A. Pachamanova, has invested in Taiwan Stock Exchange for the years 1999 to 2009. Analysis was conducted with the use of Logistic Regression, K-Nearest Neighbors (kNN) and Classification Trees. Due to confidentiality reasons, our team was not able to disclose the names of these 6,189 companies; therefore, we were not able to predict the bankruptcy of the firms outside of our data set. The objective of this research was to see if we can forecast bankruptcy based on the financial ratios. Although there are more than 300 financial ratios in the finance industry, DataBusters were able to provide an accurate bankruptcy prediction based on seven financial ratios which is applicable to both private and public firms.

Based on our algorithms and models, our main business findings are:

* Although there are more than 300 financial ratios in the finance industry, in order to predict the bankruptcy these 7 are the most significant ones:
  + Cash/Total Assets
  + Working Capital to Total Assets
  + Net Income to Total Assets
  + Working Capital/Equity
  + Retained Earnings to Total Assets
  + ROA(C) before interest and depreciation before interest
  + Borrowing Dependency
* Among these 7 financial ratios Borrowing Dependency is the most important financial ratio to predict Bankruptcy
  + Borrowing Dependency: It is the effective interest rate that a company pays on its debts, such as bonds and loans.So, the lower the Borrowing Dependency, it is better for the company.
* We should not invest in “High Risk” companies to avoid bankruptcy. All of the companies we identified as “High Risk” have bankrupted.

Lastly, we should be aware of the limitations of these recommendations since they were created based on Taiwanese Stock Exchange and the first 10 years of the 21st century. Significant and good financial ratios may differ a lot based on the industry, geography and the time of the companies.

## 

## **Research Problem and Data**

A bankruptcy these days might seem as something that can be negligible or easily avoidable with very minimal repercussions; nonetheless, while it has been easier to declare and move on forward in these situations, these are rarely this simple. Bankruptcy has a negative impact both on the enterprise itself, the local economy, and even the aggregate global economy for certain firms. Business practitioners, investors, governments, and academic researchers have long studied ways to identify the potential risk of business failure in order to reduce the economic loss caused by bankruptcy to the firm, their partners, shareholders, and anyone who came in contact with them. For certain firms that hold significant status and prowess over the global economy, a bankruptcy can be used as an analogy with a big chain and key links. The big chain is the global operations in a large-scale economy and the links among the chain are the firms that sustain it via their connection and collaboration. Now imagine a link going missing. This is when massive potential disasters could occur (Brent George Law).

We started our work by finding the dataset we wanted to work with. We found a company bankruptcy prediction dataset that had data over 10 years about Taiwanese companies. The target variable was a binary 0 or 1 that essentially answered the question of whether a company would go bankrupt or not, where 1 = bankrupt and 0 = not bankrupt. Once we decided to use this dataset, we realized that there were no missing data so preprocessing the data was simpler. After that, we realized that there were many predictor variables that could have been eliminated for the purpose of concision. Once the best set of predictors were selected, we utilized our analysis methods we learned in class, starting from Clustering and Logistics Regression to kNN and Decision (Classification) Trees to understand the quantitative values behind our problem. Finally, we looked at all of the methods and decided on an optimal one to be used in a scenario like this. All of our findings along with our business recommendations are mentioned below in their respective sections. Also, it is essential to note that we found our dataset from Kaggle.com and we realized that Kaggle.com also obtained it from UCI Machine Learning Repository. The dataset was deemed to have the best usability in the website, which was a score of 10.0 that was rated out of a max value of 10.0.

In our dataset, we have 6,819 companies, the names of which we could not obtain due to the confidentiality of the data. However, our research suggested that the majority of Taiwanese companies are manufacturing companies (World Atlas). So we can assume that most of the companies in our dataset are also manufacturing businesses. Our data covers the years from 1999 to 2009. Although there can be hundreds of financial ratios in the finance industry, our data set consists of 95 different financial ratios **(Exhibit 1)**.

As regards our goal, we, as DataBusters, have aimed to explain the relationship between financial ratios and the risk of companies going bankrupt. Therefore, our target variable is named as *Bankruptcy* and it is binary. Hence, the outcome of a model is either zero or one. As mentioned above, we have 95 financial ratios as predictors, and they are all numerical and scaled from 0 to 1. The 95 financial ratios consist of, solvency ratios, capital structure ratios, cash flow ratios, profitability ratios and turnover ratios.

As far as the overall outcome of the cases is concerned, among more than six thousand companies, only 3% of them are classified as bankrupt, and the rest 97% are classified as otherwise **(Exhibit 2)**. Therefore, it can be inferred that our data set is quite imbalanced. Additionally, our dataset brings some limitations as we do not have the means and the standard deviation of the financial ratios in the unscaled data and we do not have the names of companies. For this reason, we could not add a new observation and predict its outcome.

## **Significance of the Predictors**

We wanted to find the significant predictors in our data set since we had 95 financial ratios. First of all, we had a manual accounting and financial analysis which consisted of looking at each financial ratio and understanding the meaning of it.

While having the preliminary analysis for our data set we found out that 7 of the financial ratios were only applicable to public companies. As we wanted to develop a model that can be applied to any kind of company, we hence decided to remove these variables. Also, since shares outstanding in a company can be altered upon the decision of board members, we did not find measuring such values by share much significant for the performance of the company. These 7 ratios are:

* Net Value Per Share (B): Book Value Per Share(B)
* Net Value Per Share (A): Book Value Per Share(A)
* Net Value Per Share (C): Book Value Per Share(C)
* Revenue Per Share (Yuan ¥): Sales Per Share
* Operating Profit Per Share (Yuan ¥): Operating Income Per Share
* Per Share Net profit before tax (Yuan ¥): Pretax Income Per Share
* Cash Flow Per Share

Later, we realized that the Liability-Asset Flag and Net Income Flag were not providing significant insights. We found out that, among 6819 companies only 8 of them were classified as Class 1 for the variable of Liability-Assets Flag, meaning that the total liability exceeded total assets. In addition to that, all cases were classified as Class 1 for the variable of Net Income Flag. Given these facts, we found these variables to be uninformative and insignificant and decided not to include them in our model.

Lastly, because of the fact that we already had the variable of Liability to Equity, we considered the Equity to Liability variable to be unnecessary for our model. In conclusion with our manual financial and accounting analysis we were able to eliminate 10 predictors from our data set.

After going down to 85 predictors, we ran a stepwise regression. This stepwise regression gave us 25 significant predictors **(Exhibit 3)**.

Then we ran a one variable logistic regression with each 85 predictors and we looked at their P-values and AICs in order to compare their significance **(Exhibit 4)**. Then we ranked the predictors with the lowest 25 P-value and lowest 25 AIC. We compared the predictors that we obtained from the stepwise regression, predictors with the lowest 25 P-values and AICs. We found out that 7 predictors were in common among these 3 categories **(Exhibit 3)**.

At the end we concluded that these 7 common predictors are the most significant financial ratios to predict bankruptcy:

* **Cash/Total Assets:** Measures the portion of a company's assets held in cash (Inc.).
* **Working Capital to Total Assets:** The working capital to total assets ratio compares the net liquid assets to the total assets of the firm. Working Capital to Total Assets ratio determines the short-term company's solvency (App For Finance).
* **Net Income to Total Assets:** This ratio also means Return on Assets. Return on assets is a profitability ratio that provides how much profit a company is able to generate from its assets. In other words, it measures how efficient a company's management is in generating earnings from their economic resources or assets on their balance sheet (Investopedia).
* **Working Capital/Equity:** This ratio is the difference between current assets and current liabilities and divides it to equity to show the level of companies leverage (Investopedia).
* **Retained Earnings to Total Assets:** It is the ratio measuring the accumulated earnings over a company’s total asset.In other words, it shows management intention to overuse debt or new shares to invest in the company asset (AccountingInside).
* **ROA(C) before interest and depreciation before interest:** Pretax interest before depreciation recurring net profit divided by total assets. It shows how well the management team used its assets to create an interest free net income (Hindawi).
* **Borrowing dependency, Cost of Interest-bearing Debt:** It is the effective interest rate that a company pays on its debts, such as bonds and loans.So, the lower the Borrowing Dependency, it is better for the company. (Investopedia)

## **Unsupervised Learning: Clusters**

In order to have a better understanding of the data, we utilized the K-means clustering algorithm. First of all we wanted to decide about the number of clusters by looking at the Elbow Plot. Based on the Elbow Plot, we wanted to have 3 different clusters because K=3 clusters was the best option to minimize the Sum of Squares and Number of clusters. In other words, K=3 clusters is the left bottom part, elbow, of the Elbow Plot **(Exhibit 4)**. Finally, by looking at the most significant 7 variables, this exploratory work allowed us to create 3 clusters of companies based on their risk levels **(Exhibit 5)**.

**Low Risk Companies (Cluster 2):** We identified 16% of the data set (1068 companies) as low risk companies. These companies have the lowest Borrowing Dependency[[1]](#footnote-0) and the highest average of all ratios.

**Medium Risk Companies (Cluster 3):** We identified 81% of the data set (5531 companies) as medium risk companies. These companies have a medium range of selected ratios.

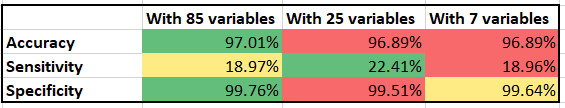
**High Risk Companies (Cluster 1):** We have identified 3% of the data set (220 companies) as high risk companies. These companies have the lowest average financial ratios except Borrowing Dependency. They have the highest borrowing dependency.

By using the Clustering algorithm we were able to allow the system to identify patterns within data sets on its own.

## **Logistics Regression**

We ran the logistics regression with all of the 85 variables, the 25 variables that we obtained from the stepwise regression and 7 variables based on the lowest P-values and AICs. The cut-of value is set at 0.5.

Below is the summary table of the confusion matrices obtained from the three models to evaluate their performance:

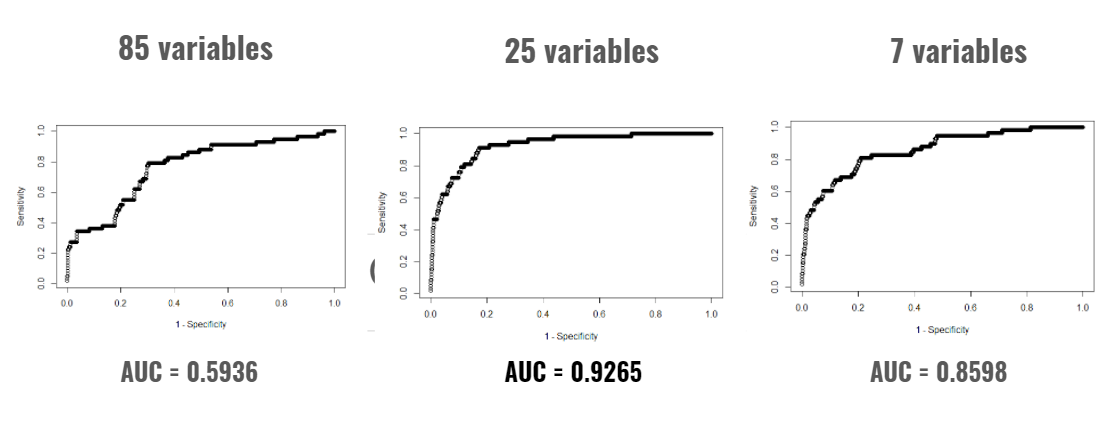
**Summary of the logistic regression confusion matrices**

We do not perceive much of a difference in the values in accuracy and specificity across three models. The model with 85 variables has a higher level of accuracy and specificity than the other two, suggesting that there are certain factors that the model with 85 variables could capture while the other two cannot. However, the discrepancy in total accuracy and specificity amongst the three models is very minimal. Therefore, this difference is negligible.

Sensitivity is low across three models, which is rather expected due to the imbalance between 0 and 1 observations in our dataset. Sensitivity is the metric that poses a more noticeable difference among the three models. The 25-variable model performs the best with 22.41% of class 1 observations (companies that go bankrupt) being correctly identified by the model. However, the difference between the 25-variable model and the other two is relatively small at around 3.45%.

In order to further evaluate the model, we investigate the classification ability of these three by graphing the ROC curve and measuring the area under the curve of each.

**Summary of three ROC curves and their AUC values**



The 85-variable scores a value of 0.5936 for its AUC, reflecting a relatively poor performance that is close to no discrimination in identifying the probability of a company going bankrupt.

At a cut-off point of 0.5, the model with 85 variables returns the lowest value of area under the curve. This could be explained by the fact that the model with 85 variables is overfitted, which leads to bias. Also, multicollinearity could exist amongst predictor variables when using 85 variables. The model with 25 variables has the highest AUC of 0.9265, which indicates a strong performance in distinguishing the two classes. With an AUC of 0.8598, the model with 7 variables is also at an outstanding performance in its ability to distinguish the two classes. It’s slightly behind the 25-variable model but the difference is also not too significant.

It can be concluded that adding more variables to the model does not remarkably contribute to the performance in classifying the two classes. Therefore, of all three models, the 7-variable model is the most efficient one with a high level of accuracy and area under the curve.

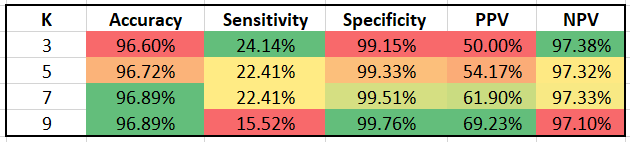
For the kNN and decision tree analysis that follow, we will use these 7 variables to provide the most simplified model yet still capture accurate information efficiently.

## **K-Nearest Neighbors (KNN)**

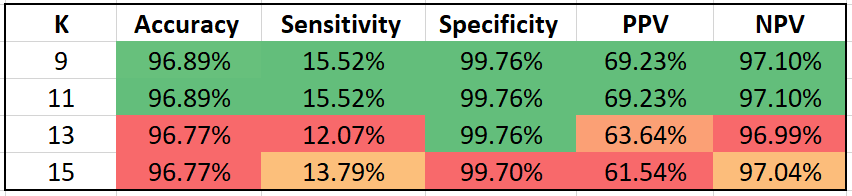
By using the 7 most significant variables and using the nearest neighbors corresponding to the values of predictors, we utilized the K-Nearest Neighbors algorithm in order to classify the observations. We ran our algorithm with different values of K namely 3, 5, 7 and 9. Among the results we found out, 7 and 9 as the number of nearest neighbors stood out as they yielded the highest accuracy which was 96.89%.

However, we decided to go with 7 neighbors instead of 9 as it provided us a better sensitivity. It is important for our dataset to yield higher sensitivity as much as possible since it indicates the true positive rate which corresponds to the possibility of companies going bankrupt. Since we believe that foreseeing the bankruptcy of a company is vitally important for both investors and company managers in the business world, we have taken sensitivity as our guide. Overall, based on this confusion matrix, K=7 has been the best performing number of neighbors for us, as seen below.

**Summary of the Confusion Matrices**



Apart from that, we deliberately did not use an even number as the nearest number of neighbors for our model because we did not want RStudio to make the decision on behalf of us whether a company will go bankrupt or not when a case is equidistant to both possibilities. Therefore, we only ran the kNN model with odd numbers which were 3, 5, 7 and 9. As regards the numbers over 9, the model was not able to provide a reliable outcome as values were similar to one another. The results of our kNN model with neighbor numbers 9 and higher is as follows below:



## 

## 

## 

## **Classification Trees**

When we created the decision trees for the 7 predictor variable dataset, we utilized balancing, classification trees and pruning. Regression trees was not an option since our target variable was a binary 0 or 1 variable. As already mentioned our data was quite imbalanced, in order to obtain more accurate results, we wanted to balance our data. In terms of balancing, we refer to down and upsampling the data, which can also be referred to as decimation and interpolation, respectively. All in all, we had three classification trees:

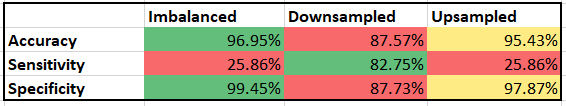
1. Imbalanced Data
2. Downsampled Data
3. Upsampled Data

Out of all the three confusion matrices, the one with the best accuracy and specificity was the imbalanced classification tree:

* Accuracy = 96.95% ≈ 97%
* Specificity = 99.45% ≈ 99%

Nonetheless, looking at the sensitivity measure, both imbalance and upsampled classification trees have the worst values. It might be assumed that the imbalanced classification tree is deemed best since it wins in two categories out of the three; however, the ratio between the sensitivity and specificity is something that one must consider to demonstrate more analytical thinking. Hence, although the downsampled classification tree has the worst accuracy and sensitivity, its sensitivity far exceeds those of imbalanced and upsampled, where there is a difference of almost 60%. In addition, downsampled classification tree’s ratio of sensitivity versus specificity is almost 1, which means a great AUC. Also, downsampled classification tree’s accuracy and specificity are high values nonetheless. All in all, we believe that the downsampled approach is the best one since it has already high values in all categories with the most important ratio and category being the strongest compared to the rest.

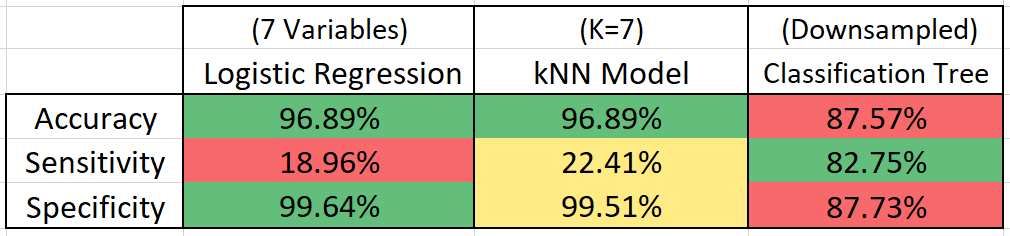
**Summary of the Confusion Matrices**



## **Optimal Model**

According to MoneyControl the average return in the Taiwnese Stock Exchange is 18.81%. It is slightly more than the S&P 500’s average return of 2020 which was 18.4% (Y-Charts). Based on this, if we assume that an investor who invests $10,000 in the Taiwanese Stock Market would make $11,881 at the end of the year. But if that company goes bankrupt the investor will be losing all of his investment of $10,000. In order to overcome one wrong investment, the investor needs to make more than 5 successful investments. This is the loss that happens in the case of a false negative, which is the scenario that a company is predicted to not go bankrupt but ends up going bankrupt. Meanwhile, if the investor chooses not to invest in a company that is predicted to go bankrupt but turns out not to, they only pay the opportunity of not investing in that company. Therefore, a false negative causes a bigger loss and hurts more than a false positive. As a result, sensitivity is an important metric for us to base our decision on as it reflects the ability to correctly identify class 1 value (companies that go bankrupt) of a model.

**Summary of All of the Confusion Matrices**



We can conclude that the optimal model is the classification tree with a downsampled dataset. Even though this model scores a lower value in accuracy and specificity, the difference is unremarkable, only at 9.32%. More importantly, the classification tree performs the best in sensitivity analysis at 82.75% while maintaining a good accuracy and specificity.

## **Findings and Recommendations**

### Data Findings

Through data processing and analysis, we managed to extract some data findings as below:

* The number of variables does not dictate the accuracy and efficiency of a model. Increasing variables does not necessarily return a better model. There are problems with overfitting, bias, and multicollinearity in a complex model that utilizes too many variables. Therefore, it is important to identify the most efficient model that captures the same level of accuracy with fewer variables.
* Accuracy is neither the most important nor the most reliable metrics when accessing a model. Performance evaluation depends on the objectives and goals of building a model. Therefore, it is essential to take the full picture with other metrics into consideration.
* Balancing the data is key in order to improve the performance of decision tree models.

### Business Findings

Based on our algorithms and models, the business findings are as follows:

* The finance industry has over 300 financial ratios. Initially our data set had 95 financial ratios, however, to predict bankruptcy, we found these 7 predictors to be the most significant ones:
  + Cash/Total Assets
  + Working Capital to Total Assets
  + Net Income to Total Assets
  + Working Capital/Equity
  + Retained Earnings to Total Assets
  + ROA(C) before interest and depreciation before interest
  + Borrowing Dependency
* Based on our decision trees, among these 7 financial ratios, we found Borrowing Dependency to be the most important financial ratio to predict bankruptcy.
* We found that investing into “High Risk” companies leads to bankruptcy, thus we do not recommend investing into such companies, despite the chance of higher returns. This is because all of the companies we identified as “High Risk” have bankrupted.

Lastly, we should take into account the possible limitations of these recommendations. These conclusions were drawn based on the 2000s Taiwanese Stock Exchange, with most companies operating in the manufacturing industry. Significance of financial ratios may vary for different industries, geographies and times.

## 

## 

## 

## **Appendix**

**Exhibit 1**

The Target Variable and The Predictors

Y - Bankrupt?: Class label

X1 - ROA(C) before interest and depreciation before interest: Return On Total Assets(C)

X2 - ROA(A) before interest and % after tax: Return On Total Assets(A)

X3 - ROA(B) before interest and depreciation after tax: Return On Total Assets(B)

X4 - Operating Gross Margin: Gross Profit/Net Sales

X5 - Realized Sales Gross Margin: Realized Gross Profit/Net Sales

X6 - Operating Profit Rate: Operating Income/Net Sales

X7 - Pre-tax net Interest Rate: Pre-Tax Income/Net Sales

X8 - After-tax net Interest Rate: Net Income/Net Sales

X9 - Non-industry income and expenditure/revenue: Net Non-operating Income Ratio

X10 - Continuous interest rate (after tax): Net Income-Exclude Disposal Gain or Loss/Net Sales

X11 - Operating Expense Rate: Operating Expenses/Net Sales

X12 - Research and development expense rate: (Research and Development Expenses)/Net Sales

X13 - Cash flow rate: Cash Flow from Operating/Current Liabilities

X14 - Interest-bearing debt interest rate: Interest-bearing Debt/Equity

X15 - Tax rate (A): Effective Tax Rate

X16 - Net Value Per Share (B): Book Value Per Share(B)

X17 - Net Value Per Share (A): Book Value Per Share(A)

X18 - Net Value Per Share (C): Book Value Per Share(C)

X19 - Persistent EPS in the Last Four Seasons: EPS-Net Income

X20 - Cash Flow Per Share

X21 - Revenue Per Share (Yuan ¥): Sales Per Share

X22 - Operating Profit Per Share (Yuan ¥): Operating Income Per Share

X23 - Per Share Net profit before tax (Yuan ¥): Pretax Income Per Share

X24 - Realized Sales Gross Profit Growth Rate

X25 - Operating Profit Growth Rate: Operating Income Growth

X26 - After-tax Net Profit Growth Rate: Net Income Growth

X27 - Regular Net Profit Growth Rate: Continuing Operating Income after Tax Growth

X28 - Continuous Net Profit Growth Rate: Net Income-Excluding Disposal Gain or Loss Growth

X29 - Total Asset Growth Rate: Total Asset Growth

X30 - Net Value Growth Rate: Total Equity Growth

X31 - Total Asset Return Growth Rate Ratio: Return on Total Asset Growth

X32 - Cash Reinvestment %: Cash Reinvestment Ratio

X33 - Current Ratio

X34 - Quick Ratio: Acid Test

X35 - Interest Expense Ratio: Interest Expenses/Total Revenue

X36 - Total debt/Total net worth: Total Liability/Equity Ratio

X37 - Debt ratio %: Liability/Total Assets

X38 - Net worth/Assets: Equity/Total Assets

X39 - Long-term fund suitability ratio (A): (Long-term Liability+Equity)/Fixed Assets

X40 - Borrowing dependency: Cost of Interest-bearing Debt

X41 - Contingent liabilities/Net worth: Contingent Liability/Equity

X42 - Operating profit/Paid-in capital: Operating Income/Capital

X43 - Net profit before tax/Paid-in capital: Pretax Income/Capital

X44 - Inventory and accounts receivable/Net value: (Inventory+Accounts Receivables)/Equity

X45 - Total Asset Turnover

X46 - Accounts Receivable Turnover

X47 - Average Collection Days: Days Receivable Outstanding

X48 - Inventory Turnover Rate (times)

X49 - Fixed Assets Turnover Frequency

X50 - Net Worth Turnover Rate (times): Equity Turnover

X51 - Revenue per person: Sales Per Employee

X52 - Operating profit per person: Operation Income Per Employee

X53 - Allocation rate per person: Fixed Assets Per Employee

X54 - Working Capital to Total Assets

X55 - Quick Assets/Total Assets

X56 - Current Assets/Total Assets

X57 - Cash/Total Assets

X58 - Quick Assets/Current Liability

X59 - Cash/Current Liability

X60 - Current Liability to Assets

X61 - Operating Funds to Liability

X62 - Inventory/Working Capital

X63 - Inventory/Current Liability

X64 - Current Liabilities/Liability

X65 - Working Capital/Equity

X66 - Current Liabilities/Equity

X67 - Long-term Liability to Current Assets

X68 - Retained Earnings to Total Assets

X69 - Total income/Total expense

X70 - Total expense/Assets

X71 - Current Asset Turnover Rate: Current Assets to Sales

X72 - Quick Asset Turnover Rate: Quick Assets to Sales

X73 - Working capital Turnover Rate: Working Capital to Sales

X74 - Cash Turnover Rate: Cash to Sales

X75 - Cash Flow to Sales

X76 - Fixed Assets to Assets

X77 - Current Liability to Liability

X78 - Current Liability to Equity

X79 - Equity to Long-term Liability

X80 - Cash Flow to Total Assets

X81 - Cash Flow to Liability

X82 - CFO to Assets

X83 - Cash Flow to Equity

X84 - Current Liability to Current Assets

X85 - Liability-Assets Flag: 1 if Total Liability exceeds Total Assets, 0 otherwise

X86 - Net Income to Total Assets

X87 - Total assets to GNP price

X88 - No-credit Interval

X89 - Gross Profit to Sales

X90 - Net Income to Stockholders’ Equity

X91 - Liability to Equity

X92 - Degree of Financial Leverage (DFL)

X93 - Interest Coverage Ratio (Interest expense to EBIT)

X94 - Net Income Flag: 1 if Net Income is Negative for the last two years, 0 otherwise

X95 - Equity to Liability

**Exhibit 2**

Summary of the outcome in the data set

| **Bankruptcy** | **Number of Companies** | **Percentage of Companies** |
| --- | --- | --- |
| 0 (No) | 6,599 | 97% |
| 1 (Yes) | 220 | 3% |
| **Grand Total** | **6,819** | **100%** |

**Exhibit 3**

Summary of Significant Predictors: Green color represents the common variables in columns

****

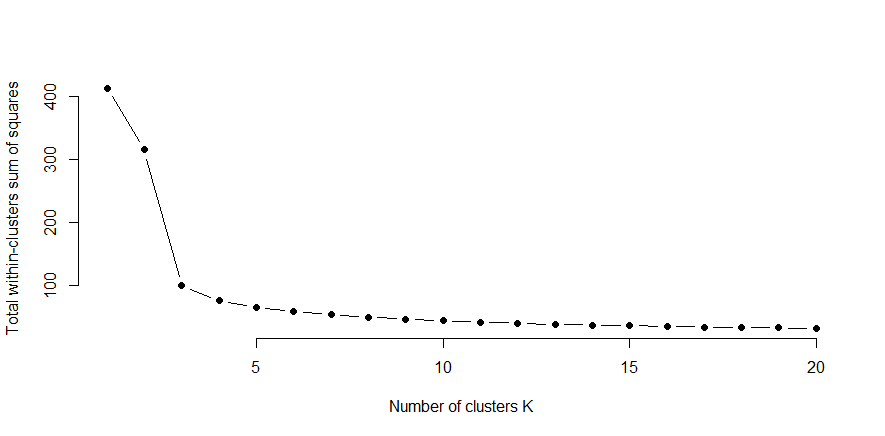
**Exhibit 4**

P-Values and AICs of each 85 Predictor

|  | **Variable Name** | **P Value** | **AIC** |
| --- | --- | --- | --- |
| 1 | ROA(C) before interest and depreciation before interest | 2.00E-16 | 1590.6 |
| 2 | ROA(A) before interest and % after tax | 2.00E-16 | 1597.6 |
| 3 | ROA(B) before interest and depreciation after tax | 2.00E-16 | 1595.2 |
| 4 | Operating Gross Margin | 2.49E-07 | 1911.3 |
| 5 | Realized Sales Gross Margin | 3.29E-07 | 1911.8 |
| 6 | Operating Profit Rate | 9.85E-01 | 1947.7 |
| 7 | Pre-tax net Interest Rate | 5.13E-01 | 1947.4 |
| 8 | After-tax net Interest Rate | 4.98E-01 | 1947.4 |
| 9 | Non-industry income and expenditure/revenue | 5.61E-02 | 1945.1 |
| 10 | Continuous interest rate (after tax) | 5.19E-01 | 1947.4 |
| 11 | Operating Expense Rate | 6.16E-01 | 1947.5 |
| 12 | Research and development expense rate | 4.61E-02 | 1943.5 |
| 13 | Cash flow rate | 1.00E-07 | 1915 |
| 14 | Interest-bearing debt interest rate | 1.01E-01 | 1941.9 |
| 15 | Tax rate (A) | 2.00E-16 | 1817.7 |
| 16 | Persistent EPS in the Last Four Seasons | 2.00E-16 | 1447.5 |
| 17 | Realized Sales Gross Profit Growth Rate | 9.70E-01 | 1947.7 |
| 18 | Operating Profit Growth Rate | 3.03E-01 | 1947 |
| 19 | After-tax Net Profit Growth Rate | 1.72E-02 | 1943.5 |
| 20 | Regular Net Profit Growth Rate | 1.88E-02 | 1943.6 |
| 21 | Continuous Net Profit Growth Rate | 1.45E-01 | 1946.2 |
| 22 | Total Asset Growth Rate | 2.72E-04 | 1935.1 |
| 23 | Net Value Growth Rate | 3.47E-01 | 1941 |
| 24 | Total Asset Return Growth Rate Ratio | 5.19E-02 | 1944.3 |
| 25 | Cash Reinvestment % | 2.18E-05 | 1932.9 |
| 26 | Current Ratio | 9.77E-01 | 1947.6 |
| 27 | Quick Ratio | 1.95E+03 | 0.072 |
| 28 | Interest Expense Ratio | 8.23E-01 | 1947.7 |
| 29 | Total debt/Total net worth | 3.56E-01 | 1947.1 |
| 30 | Debt ratio % | 2.00E-16 | 1563.3 |
| 31 | Net worth/Assets | 2.00E-16 | 1563.3 |
| 32 | Long-term fund suitability ratio (A) | 1.91E-01 | 1946.5 |
| 33 | Borrowing dependency | 5.60E-10 | 1866.4 |
| 34 | Contingent liabilities/Net worth | 2.28E-02 | 1937.3 |
| 35 | Operating profit/Paid-in capital | 2.00E-16 | 1656 |
| 36 | Net profit before tax/Paid-in capital | 2.00E-16 | 1459.2 |
| 37 | Inventory and accounts receivable/Net value | 2.68E-04 | 1930.7 |
| 38 | Total Asset Turnover | 7.53E-09 | 1904.3 |
| 39 | Accounts Receivable Turnover | 7.05E-01 | 1947.5 |
| 40 | Average Collection Days | 7.15E-01 | 1947.1 |
| 41 | Inventory Turnover Rate (times) | 9.10E-01 | 1947.7 |
| 42 | Fixed Assets Turnover Frequency | 5.66E-09 | 1918.8 |
| 43 | Net Worth Turnover Rate (times) | 8.62E-02 | 1945.4 |
| 44 | Revenue per person | 2.87E-02 | 1944.3 |
| 45 | Operating profit per person | 2.00E-16 | 1837.8 |
| 46 | Allocation rate per person | 8.16E-01 | 1947.7 |
| 47 | Working Capital to Total Assets | 2.00E-16 | 1693.4 |
| 48 | Quick Assets/Total Assets | 2.30E-12 | 1893.1 |
| 49 | Current Assets/Total Assets | 2.29E-04 | 1933.9 |
| 50 | Cash/Total Assets | 2.00E-16 | 1815.1 |
| 51 | Quick Assets/Current Liability | 9.74E-01 | 1947.5 |
| 52 | Cash/Current Liability | 1.91E-07 | 1929.8 |
| 53 | Current Liability to Assets | 2.00E-16 | 1746.1 |
| 54 | Operating Funds to Liability | 4.23E-12 | 1903.9 |
| 55 | Inventory/Working Capital | 8.73E-01 | 1947.7 |
| 56 | Inventory/Current Liability | 9.46E-01 | 1947.7 |
| 57 | Current Liabilities/Liability | 8.61E-02 | 1944.8 |
| 58 | Working Capital/Equity | 2.00E-16 | 1829.6 |
| 59 | Current Liabilities/Equity | 1.94E-09 | 1880.5 |
| 60 | Long-term Liability to Current Assets | 9.49E-01 | 1947.7 |
| 61 | Retained Earnings to Total Assets | 2.00E-16 | 1805.5 |
| 62 | Total income/Total expense | 2.00E-16 | 1614 |
| 63 | Total expense/Assets | 2.00E-16 | 1879.8 |
| 64 | Current Asset Turnover Rate | 3.25E-01 | 1946.8 |
| 65 | Quick Asset Turnover Rate | 3.36E-02 | 1943.4 |
| 66 | Working Capital Turnover Rate | 8.09E-01 | 1947.7 |
| 67 | Cash Turnover Rate | 1.37E-01 | 1945.4 |
| 68 | Cash Flow to Sales | 9.68E-01 | 1947.7 |
| 69 | Fixed Assets to Assets | 9.61E-01 | 1940.8 |
| 70 | Current Liability to Liability | 8.61E-02 | 1944.8 |
| 71 | Current Liability to Equity | 1.94E-09 | 1880.5 |
| 72 | Equity to Long-term Liability | 1.38E-04 | 1908.4 |
| 73 | Cash Flow to Total Assets | 1.19E-08 | 1918.3 |
| 74 | Cash Flow to Liability | 1.65E-04 | 1936.7 |
| 75 | CFO to Assets | 2.00E-16 | 1871.1 |
| 76 | Cash Flow to Equity | 1.45E-04 | 1926.6 |
| 77 | Current Liability to Current Assets | 8.27E-15 | 1874.1 |
| 78 | Net Income to Total Assets | 2.00E-16 | 1605.9 |
| 79 | Total assets to GNP price | 1.13E-02 | 1943.4 |
| 80 | No-credit Interval | 6.40E-01 | 1947.5 |
| 81 | Gross Profit to Sales | 2.48E-07 | 1911.3 |
| 82 | Net Income to Stockholder's Equity | 2.00E-16 | 1832.5 |
| 83 | Liability to Equity | 4.69E-11 | 1872.5 |
| 84 | Degree of Financial Leverage (DFL) | 4.23E-01 | 1947.3 |
| 85 | Interest Coverage Ratio (Interest expense to EBIT) | 6.46E-01 | 1947.5 |

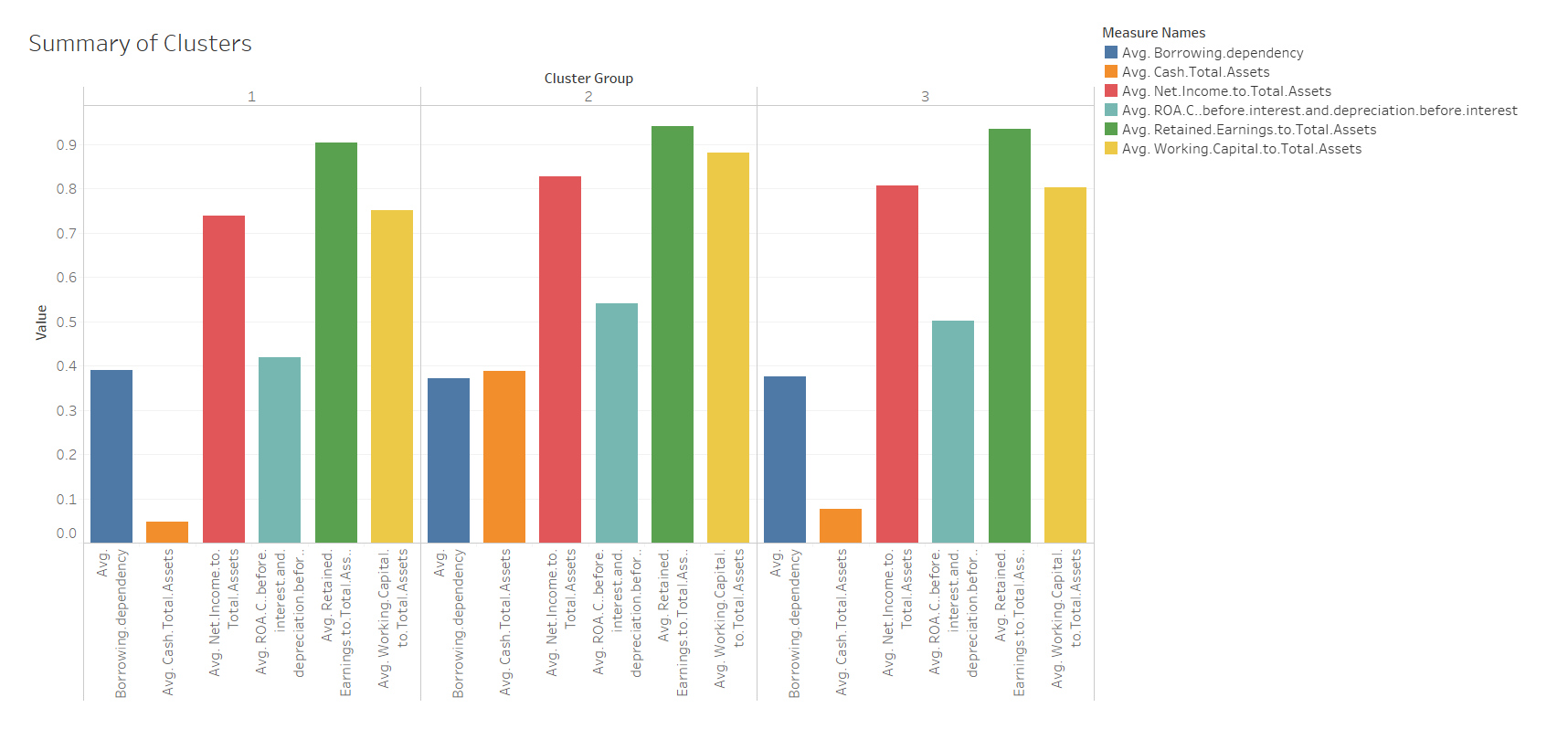
**Exhibit 5**

Elbow Plot for Clustering



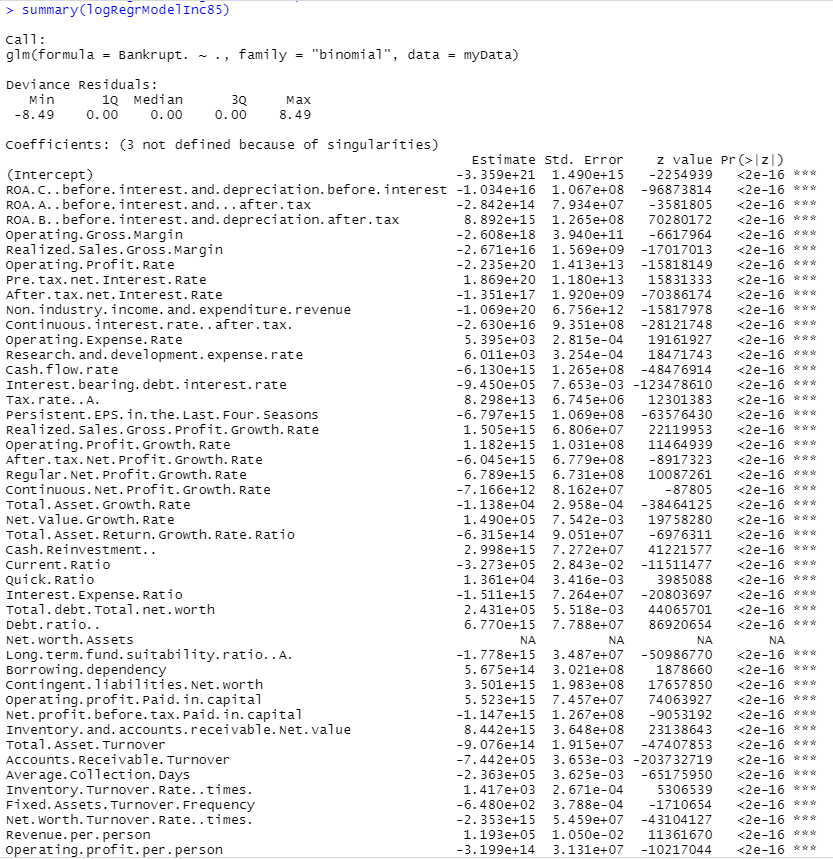
**Exhibit 6**

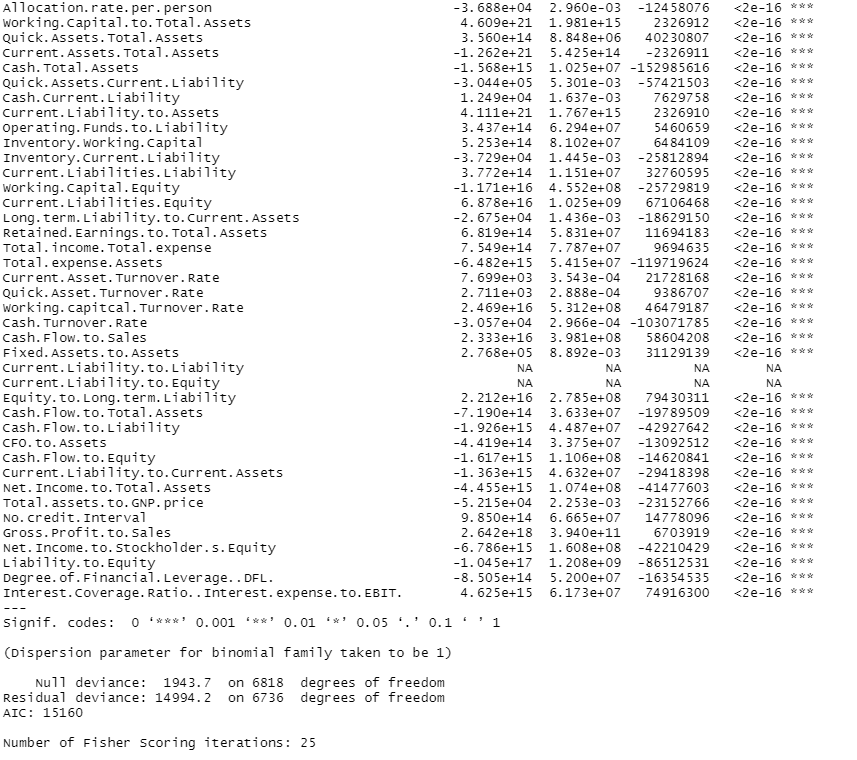
The detailed analysis of each cluster groups



**Exhibit 7**

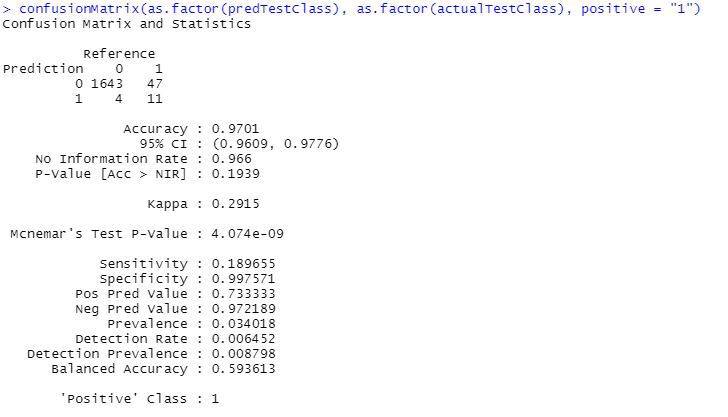
Results of 85-variable logistic regression model



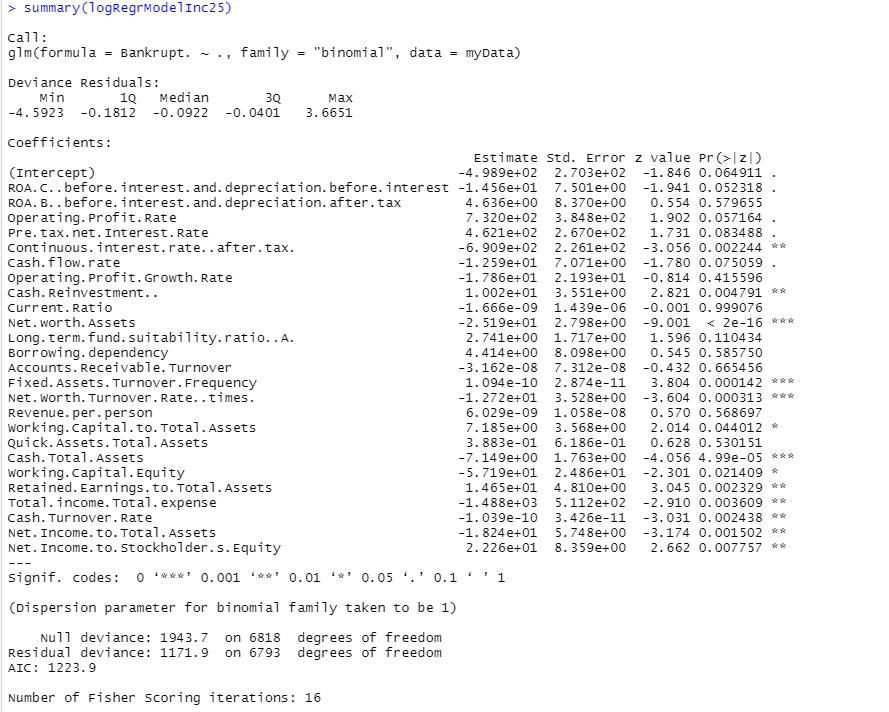


**Exhibit 8**

Confusion matrix of 85-variable logistic regression model

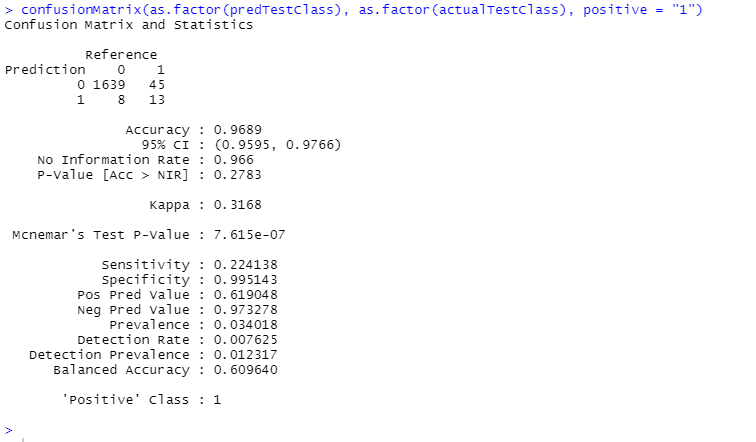


**Exhibit 9**

Results of 25-variable logistic regression model 

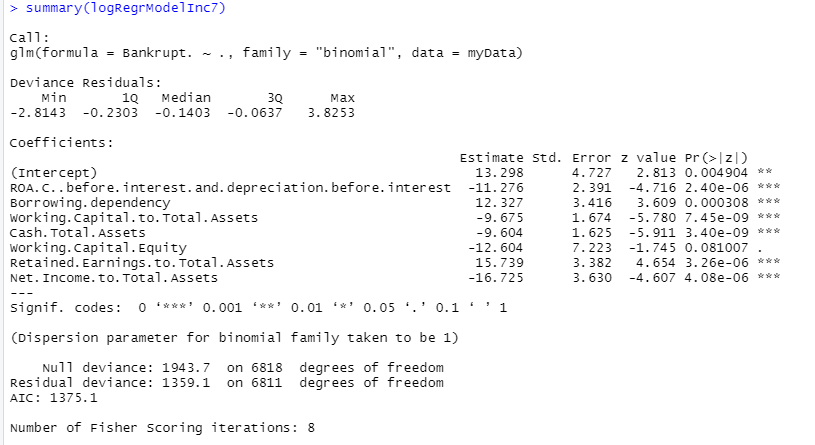
**Exhibit 10**

Confusion matrix of 25-variable logistic regression model



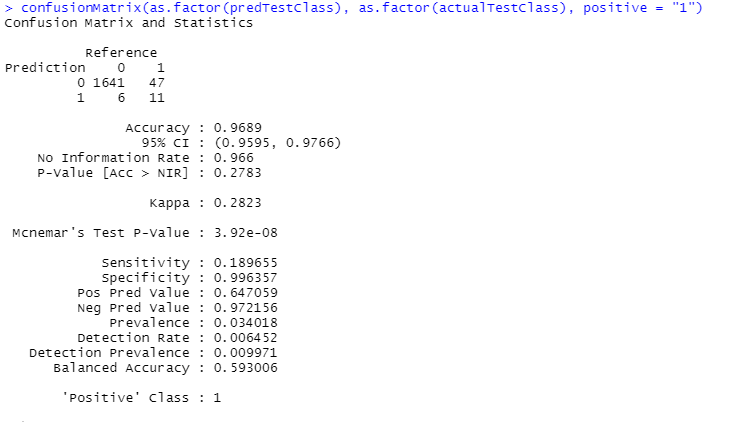
**Exhibit 11**

Results of 7-variable logistic regression model



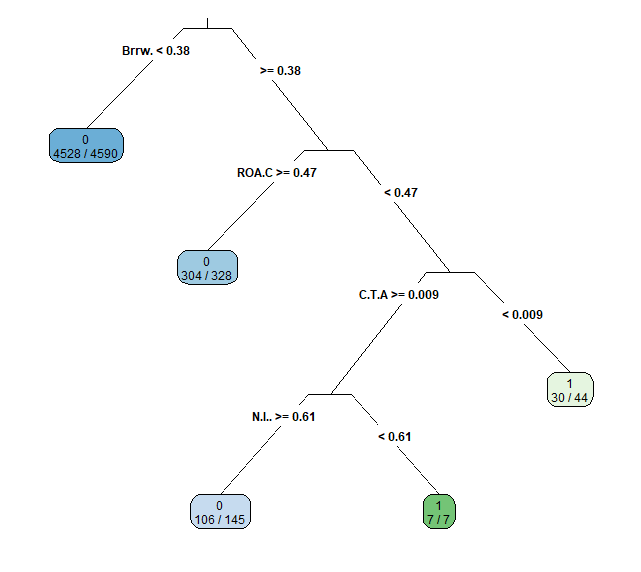
**Exhibit 12**

Confusion matrix of 7-variable logistic regression model



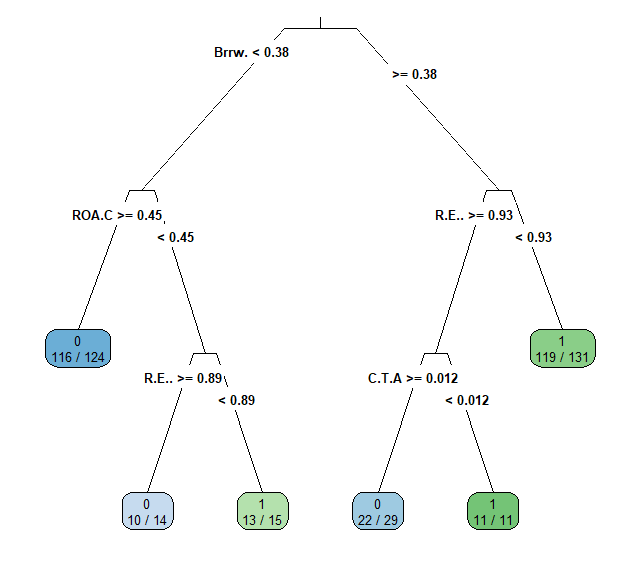
**Exhibit 13**

Imbalanced Classification Tree (Pruned)



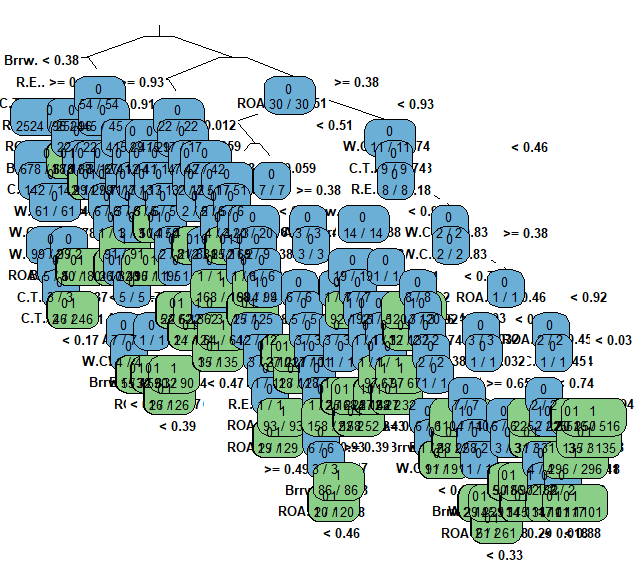
**Exhibit 14**

Downsampled Classification Tree (Pruned)



**Exhibit 15**

Upsampled Classification Tree (Pruned)



## 

## 

## 

## 

## **Works Cited**

Boyte-White, Claire. “How to Calculate Return on Assets (ROA) with Examples.” Investopedia, Investopedia, 8 Sept. 2021, https://www.investopedia.com/ask/answers/031215/what-formula-calculating-return-assets-roa.asp.

Editorial, Inc. “Financial Ratios - Encyclopedia - Business Terms.” Inc.com, Inc., 30 Nov. -1, https://www.inc.com/encyclopedia/financial-ratios.html#:~:text=Cash%20to%20total%20assets%3A%20Cash,may%20be%20viewed%20as%20inefficient.

Hayes, Adam. “Cost of Debt.” Investopedia, Investopedia, 2 Dec. 2021, https://www.investopedia.com/terms/c/costofdebt.asp.

Hindawi. “A Seasonal Time-Series Model Based on Gene Expression Programming for Predicting Financial Distress.” Table 4, https://www.hindawi.com/journals/cin/2018/1067350/tab4/.

“How Does Bankruptcy Affect the Economy?” Brent George Law, 6 Nov. 2017, https://www.brentgeorgelaw.com/bankruptcy-affect-economy/.

Liang , Deron, and Chih-Fong Tsai. “Company Bankruptcy Prediction.” Kaggle, 13 Feb. 2021, https://www.kaggle.com/fedesoriano/company-bankruptcy-prediction/metadata.

Moneycontrol.com. “Taiwan - Asian Market - Taiwan Index - Taiwan Market - Taiwan Stocks.” Moneycontrol, Moneycontrol, https://www.moneycontrol.com/live-index/taiwan.

Omondi, Sharon. “What Are the Biggest Industries in Taiwan?” WorldAtlas, WorldAtlas, 7 June 2019, https://www.worldatlas.com/articles/what-are-the-biggest-industries-in-taiwan.html.

S&amp;P 500 Annual Total Return, https://ycharts.com/indicators/sp\_500\_total\_return\_annual.

“Working Capital to Assets Ratio.” Appforfinance, https://www.appforfinance.com/working-capital-to-assets-ratio.html#:~:text=The%20working%20capital%20to%20total,the%20short%2Dterm%20company's%20solvency.

1. The lower the Borrowing Dependency, the better for the company [↑](#footnote-ref-0)