FINANCIAL DISTRESS PREDICTION USING CUTTING-EDGE STATISTICAL TECHNIQUES:
A STUDY OF AUSTRALIAN REAL ESTATE SECTOR

FINANCIAL DISTRESS PREDICTION USING CUTTING-EDGE STATISTICAL TECHNIQUES: A STUDY OF AUSTRALIAN REAL ESTATE SECTOR

Sneha Raut[1], Milind Tiwari[1], Kuldeep Kumar[1]

ABSTRACT

"Financial Distress" is a condition where a company fails to meet its financial obligations. Financial Distress Prediction (FDP) is a process of understanding if the company is leading to success or failure. FDP has gained tremendous interest after the failure of high-profile companies like Enron and WorldCom. The aim of this paper is to study various financial ratios of Australian Real Estate companies to design a model effective for predicting distress. The experiment data included 164 listed (*successful*) and 12 delisted (*failed*) Australian Real Estate companies over two-year time frame from 2016-2018. Cutting-edge statistical techniques like Logistic Regression, Multivariate Discriminant Analysis, Artificial Neural Network, Hybrid Techniques, Decision Trees, Random Forest and Stochastic Gradient Boosting are used to construct various FDP models. The experiment results indicate that hybrid model combining Artificial Neural Network and Logistic Regression along with Stochastic Gradient Boosting had superior power at predicting if the company will be successful or a failure. These models can be effectively utilized by Australian Real Estate companies for distress prediction and by their investors to make correct investments decisions.

Key Words: *Financial Distress, Bankruptcy Prediction, Artificial Neural Network, Decision Trees, Financial Ratios, Real Estate, Australia*

I. INTRODUCTION

Real Estate is an independent industry, which significantly contributes to the financial positioning of financial institutions as a consequence of asset holding and mortgage loans. Financial sector is the backbone of any economy hence we can state that the fluctuations in real estate industry will have an impact on the economic growth of the country.

Australia has marked its 26th successive year of economic progress without hitting recession. Despite the ambiguity from previous years, the market for real estate industry looks positive in 2018 (*2018 Real Estate Outlook: The Australian Perspective, Deloitte*). According to *Emerging Trends in Real Estate ®Asia Pacific* by *PwC*, Australian market is one of the markets in Asia Pacific region to withstand the trend of diminishing rental growth.

As per the real estate trend by PwC, borrowing cost is high in Australia but the real estate market remains an attractive investment destination in Asia Pacific, as it's the only market where you can buy asset with long lease terms. Additionally, student housing is gaining popularity in Australia as the influx of international students is increasing. The only challenge faced in this industry is lack of choices for the investors.

[1] *Bond University, Australia.*

Financial Distress Prediction (FDP) can be used to indicate the financial health and soundness of a sector. It can be used to analyse the financial stability and predict the future outcome for a firm in a specific industry. Real estate industry can be potentially volatile because of the price movements indirectly having an impact on the economic growth. As mentioned above, the real estate industry is an attractive option for foreign investment and so it is important to understand risks associated with this industry and which firm is potentially at a risk of failure.

The aim of this paper is to investigate the real estate market in Australia. It will include analysing the ASX listed as well as delisted real estate companies in Australia for generating a model which can predict the financial distress using the financial ratios on the balance sheet. FDP can benefit the companies as well as the investor in long run. According to Kumar & Chaturvedi (2010), some of the benefits are:

1. Money lenders like banks and other financial institutions can use this when deciding of investment.
2. Businesses can use this model to prove their stability to their lenders, supplier, investors and other business associates.
3. Business investors can use this to make a confident investment decision.

Thus, FDP modelling can help reduce losses than can be incurred to the economy by distressed firms and extend support to stakeholders of the company. Such models can be used as warning symbols by the management of the company and take actions accordingly.

II. RESEARCH OBJECTIVES

This paper with the support of theoretical framework on financial distress prediction aims at identifying financial distress is Australian Real Estate companies. This research aims are:

1. Determine the reason for financial distress supported by the balance sheet financial ratios.
2. Construct a financial distress prediction model using the financial ratios to predict if the company will be successful or a failure.

III. LITERATURE REVIEW

There has been an extensive research around Financial Distress Prediction (FDP) since the early 1930s. Numerous techniques were developed since then, but the aim has always been to get a model with least error percentage. Type I and type II Error are widely used in the literature, where Type I Error means misclassification of a failing firm as a successful one and Type II Error means misclassification of a successful firm as a failed one (Gepp & Kumar, 2012).

1. Univariate Analysis – Beaver's Univariate Model

FitzPatrick (FitzPatrcik, 1932) was the first to initiate the study of comparing the ratios of a successful firm to that of a failed firm in the early 1930. Beaver (Beaver, 1966) extended on FitzPatrick's work to develop a Univariate Model for FDP to predict successful from the failed firms. He developed this model using 30 financial ratios and tested his model on 79 successful and 79 failed firms between the 1954 and 1964 (Gepp & Kumar, 2012).

Beaver's Model had 22% Type I Error and 5% Type II Error. This error percentage increased as the duration of prediction increased, thus being more useful for short-term predictions. Another problem in Beaver's Univariate Model was that it didn't contain an overall measure of financial distress causing different ratios resulting in conflicting prediction (Gepp & Kumar, 2012). Overall, Beaver's work inspired the future development of FDP using statistical models.

2. Discriminant Analysis – Altman's Multivariate Model

Following Beaver's Univariate Analysis, Altman pioneered the FDP modelling using Discriminant Analysis (DA) in 1968. It addressed the problem of Beaver's Model i.e. conflicting prediction with various ratios by incorporating a single weighted score (Z) for distress prediction (Altman, 1968).

The following discriminant function was used to calculate the score:

$$Z = \alpha 1X1 + \alpha 2X2 + \alpha 3X3 + \ldots\ldots\ldots\ldots\ldots\ldots + \alpha nXn + c$$

The Z score for Altman's Model (1968) was:

$$Z = 0.012(X1) + 0.0141(X2) + 0.033(X3) + 0.006(X4) + 0.999(X5)$$

Where,

X_1= Working Capital/Total Assets,

X_2= Retained Earnings/Total Assets

X_3= Earnings Before Interest and Tax (EBIT)/Total

Assets X_4= Market Value of Equity/Book Value of Total

Debt X_5= Sales/Total Assets

Altman's model performed better than Beaver's model with 95% accuracy rate for shorter period but dropped to 72% when used for longer period. Additionally, it was concluded that DA with more than two classification groups i.e. Multiple DA was more accurate than Linear DA which uses only two classification groups (Martin, 1977).

3. Logistic Regression

In 1980, Ohlson conducted a research on the probabilistic prediction of a firm's financial distress using Logistic Regression (LR) Analysis. Like Altman's Z-Score (1968), Ohlson's O-Score can be labelled as a distress prediction indictor, which is developed using a set of financial ratios (Halteh, Kumar & Gepp, 2018b).

LR Analysis requires four steps as follows (Ohlson, 1980):

1. Calculate a list of financial ratios;
2. Deduce the coefficient of ratios by fitting a LR model;
3. Establish a new variable 'y' by calculating the result of each coefficient; and
4. Calculate the probability of the firm's financial distress as $1/(1+e^{-y})$.

In his study of applying LR to FDP, Ohlson (1980) deduced three LR models to predict failure for 1, 2, and 3 years ahead. This study involved 14 ratios that included financial ratios, dummy variables based on balance sheet data and a variable representing the change in net income over the previous year. The results obtained from his model were not promising but, it revealed that LR analysis was statistically more significant and easier to interpret when compared to DA.

Further research on LR analysis stated that LR is somewhat superior empirically to DA in terms of prediction as well as classification accuracy (Laitinen & Kankaanpää, 1999). Nonetheless, Martin (1977), Collin and Green (1982) and Hamer (1983) they all established that overall classification of DA and LR analysis is not largely different (Gepp & Kumar, 2012). Additionally, Hall (1994) performed a LR analysis using non-financial variables to categorize distressed firms from the non-distressed one; he was successful in setting a LR.

4. Artificial Neural Networks

Artificial Neural Networks (ANN) is a non-parametric technique, which is made up of computing algorithms based on the inner- working of a human brain (Hertz *et al.*, 1991; Gepp Kumar, 2012). Warren McCulloch and Water Pitts took the first step to create a computing model based on human brain in 1943 (McCulloch & Pitts, 1943). Thereafter, this technique of prediction has been widely applied in the field of FDP. Algorithm usage helps the system to recognize patterns and help in notifying the data important for future prediction (Kumar & Chaturvedi, 2010). The ANN technique is usually preferred method in FDP because of its capability, adaptability and ease of use (Shah, 2014).

A number of studies were carried out in similar direction by Coats and Fant (1992), Salchenberger et al. (1992), Tam and Kiang (1992), and Fletcher and Goss (1993). Salchenberger et al. (1992), performed ANN technique on bankruptcy prediction of saving and loan institutions and found it was able to predict as good as or even better than Logit models across their time periods of 6,12 and 18 months (Shah, 2014).

5. Recursive Partitioning Techniques

Recursive Partitioning Techniques are better replacement for the parametric statistical methods (Zhang & Singer, 2010). Recursive partitioning is a set of statistical techniques used for multivariate analysis. They are non-parametric techniques developed to overcome the restrictive assumptions of traditional parametric techniques (Breiman, 1984).

Decision Trees (DT), Random Forest (RF) and Stochastic Gradient Boosting (SGB) are types of Recursive Partitioning Techniques. These techniques being relatively new are less prevalent in literature compared to traditional parametric methods, but they are gaining importance based on their superior prediction power (Gepp, 2015).

6. Hybrid Techniques

Hybrid models combine numerous individual statistical methods to maximize their advantages by simultaneously reducing the weakness of the model. The central idea is that the profit attained by accuracy and confidence by traditional models like discriminant analysis, logistic regression, artificial neural networks etc. is not justified by the cost (Ravi Kumar & Ravi, 2007). Usage of hybrid techniques is gaining importance in FDP modelling (Demyanyk & Hasan, 2010). There are numerous combinations possible and not restricted to artificial intelligent techniques (Ravi Kumar & Ravi, 2007).

IV. DATA

Historical data of Australian Real Estate Sector was obtained using the MorningStar® database. Historical data is easily available, thus providing a quick data extraction process that can be cumbersome otherwise (Shultz et al., 2005). The MorningStar® database accommodates real-time raw data on investment offerings, commodities, foreign exchange etc. (Morningstar, 2015; Halteh, Kumar & Gepp, 2018).

This paper utilizes data for both listed and delisted Real Estate firms in Australia. In Section A, of the search the scope of companies was selected as "all companies – includes both listed and delisted", the year range was from "2016-2018", the market scope was limited by selecting "Australian Securities Exchange (ASX)" and finally "Real Estate" was selected for Global Industry Classification Standard (GICS) sector input.

In Section B, Annual Ratio Analysis was selected as the section and 29 variables were selected in the sub-section. These 29 variables were added as a query in Section C to run the search. The 29 variables were selected based on financial ratios/variables used in previous empirical studies, have been mentioned in the literature and finally based on the availability in the database.

List of variables is shown in Table 1. *Variables omitted due to missing values are highlighted in blue.*

LIST OF VARIABLES
1. Net Profit Margin
2. EBIT Margin
3. Return on Equity (ROE)
4. Return on Asset (ROA)
5. ROIC
6. NOPLAT Margin
7. Inventory Turnover
8. Asset Turnover
9. PPE Turnover
10. Deprecation/PP&E
11. Deprecation/Revenue
12. Working Capital Revenue
13. Working Cap Turnover
14. Gross Gearing (D/E)
15. Financial Leverage
16. Current Ratio
17. Quick Ratio

18. Gross Debt/CF
19. Cash per Share ($)
20. Invested Capital Turnover
21. Net Gearing
22. NTA per Share ($)
23. BV per Share ($)
24. Receivables/Operating Revenue
25. Inventory/Trading Revenue
26. Creditors/Operating Reve
27. Sales per Share ($)
28. EV/EBITDA
29. PER

Table 1: List of Variables (Omitted Variables in Blue)

V. METHODOLOGY

The data search process resulted in 95 companies in which 86 were listed companies and 9 were delisted companies. This data was exported to an excel spreadsheet for cleaning. Subsequently, a binary variable was introduced to explain the status of the companies; where 1 represented listed (successful) companies and 0 represented delisted (failed) companies.

After cleaning of the data, the total row count was 190 that incorporated (86 x 2years = 172) successful companies and (9 x 2years = 18) failed companies for the 29 variables. Variables columns having more than 50% missing value was omitted from further analysis (Halteh, Kumar & Gepp, 2018), resulting in omission of 11 variables leaving 18 variables for analysis. Further, company rows having more than 50% missing values were omitted, this results in omission of 14 company rows (8 successful and 6 failed), leaving with 176 companies having 164 listed and 12 delisted companies.

IBM SPSS Statistics 25 was used for performing statistical techniques like Logistic Regression, Discriminant Analysis, Artificial Neural Networks and Hybrid Analysis. Salford Predictive Modeler (SPM® Version 8.2) was used to run Decision Tree (CART), Random Forest and TreeNet (Stochastic Gradient Boosting).

1. Logistic Regression

A Binary Logistic Regression was performed using all 18-independent continuous variable against a binary categorical dependent variable. Forward LR method was utilized to generate the output. Group membership and probabilities were demanded using the standardized residuals. Hosmer-Lemeshow goodness-of-fit test was requested as a part of the analysis.

2. Discriminant Analysis

Discriminant analysis was performed using the “Enter Independents Together” method. All the independent variables were entered in a single step having no categorical variables. The univariate

ANOVAs and the BOX's M were selected in the descriptive. Similarly, the unstandardized function coefficients were demanded as a part of the analysis.

3. Artificial Neural Network

IBM SPSS Statistic 25 has two methods of neural networks: i) Multilayer Perceptron and ii) Radial Bias Function out of which multilayer perceptron is widely used. This experiment used the multilayer perceptron method to analyse the 18 covariates against one binary dependent variable. The data was portioned into three segments i.e. training, testing and holdout in 7:2:1 ratio. For the output independent variable importance chart was requested along with the ROC curve.

4. Hybrid I & II

Two hybrid combinations were implemented in this study:

1. ANN + LR
2. ANN + DA

1. Probabilities and group membership obtained from logistic regression are saved along with the independent variables.
2. Probabilities obtained from discriminant analysis are saved along with the independent variables.
3. Multilayer perceptron (ANN) is run again where the covariates now include the logistic regression saved probabilities and group memberships along with the other 18 independent variables and the result is obtained following the same method used for ANN only.
4. Multilayer perceptron (ANN) is run again where the covariates now include the discriminant analysis saved probabilities along with the other 18 independent variables and the result is obtained following the same method used for ANN only.

5. Decision Trees

Classification and Regression Trees (CART) method was used using the Salford Predictive Modeler (SPM® Version 8.2) to generate the FDP tree. All 18-independnet variables were selected as predictors for the model. Gini splitting rule was applied based on its popularity and usage. A 10 cross-fold validation was applied to the analysis.

6. Random Forest

Salford Predictive Modeler (SPM® Version 8.2) was used to run the Random Forest Ensemble. Two parameters need to be set for random forest analysis (Bhattacharyya et al., 2011; Whiting et al., 2012):

1. Number of Trees: For this study a model was generated using 200, 500 and 1000 trees to determine the best for the future analysis.
2. Number of Variable: For this study the number of variables to be considered at each node was asked to be determined using the square root of the number of predictors (independent variables) i.e. $\sqrt{18} = 4.24$ hence rounding to 4.

7. Stochastic Gradient Boosting

Again, Salford Predictive Modeler (SPM® Version 8.2) was used to run TreeNet using all 18-independent variables. Like random forest analysis models were generated using 200, 500 and 1000 trees. As stated in literature, this analysis is based on the incremental improvement and thus, need to ensure that none of the individual trees are too large. Therefore, the standard setting of 6 maximum nodes per tree along with minimum 10 data points in each node was applied to this experiment. The default cross entropy was used for determining the incremental improvement.

8. Classification Cut-off Value

A default value of 0.5 is usually used as a cut-off value where companies having value greater than 0.5 can be classified as successful and those with values lower than 0.5 can be classified as failing one. This method is not preferred when class imbalance exists. Thus, the cut-off values for classification will be empirically optimized by training (Beneish, 1997; Bayley and Taylor, 2007; Perols, 2011). As this optimization will be carried for each model run the cut-off value may differ for each model and will fall between 0 and 1. These optimized cut-off values will create a balanced output and will be mentioned separately for each of the model generated by decision tree, random forest and TreeNet analysis.

VI. RESULTS

A list of statistical tests were performed on 18 independent variables (*ROE, ROA, Asset Turnover, PPE Turnover, Depreciation/PP&E, Working Cap Turnover, Gross Gearing, Financial Leverage, Current Ratio, Quick Ratio, Gross Debt/CF, Cash per Share, Net Gearing, NTA per Share, BV per Share, Sales per Share, EV/EBITDA and PER*) to develop a model for FDP of Real Estate companies in Australia. Traits of listed company helps predict the success of a company and similarly, traits of delisted company help predict failure of a company.

1. Logistic Regression (LR) Model

Forward LR was carried out to test the impact of the variables to predict if the success of a company, hence the last step values will be used for interpretation of the model. Out of the 18 independent variables only 2 were found to be significant in FDP of the real estate companies in Australia. ***Net Gearing*** and ***BV/Share*** were the two significant variables.

Table. 5 in *Appendix* ***Omnibus Test of Model Coefficients*** gives an overall indication of how well the model performs and can be referred to as "*goodness of fit*" test. For this set of results, the significance (sig.) should be less than .05. In the output as seen in the sig. value is .000 (which means p<.05) having chi-square of 21.965 with 3 degrees of freedom.

Table. 6 in *Appendix B* ***Hosmer and Lemeshow Test*** also helps to show if the model is worthwhile and acts as "*goodness of fit*" test. In this test the result is interpreted differently

from the omnibus test discussed above. For this test, poor fit is indicated by sig. value less than .05, hence values greater than .05 is needed for supporting this model. As seen the chi-square value for this test is 5.871 with a degree of freedom of 8 having sig. of .662. This value is greater than .05, hence indicating support for the model.

Table. 7 in Appendix ***Model Summary***, which also gives another piece of information about the usefulness of the model. The ***Cox & Snell R Square*** and the ***Nagelkerke R Square*** values provide an indication of the amount of variation in the dependent variable explained by the model (ranging from minimum value of 0 to maximum of 1). The *Cox & Snell R Square* and *Nagelkerke R Square* value is .206 and .505 respectively, suggesting 20.6% and 50.5% of the variability is explained by this set of variables.

Table. 8 in Appendix ***Classification Table***, provides us with an indication of how well the model can predict the correct category (Listed vs Delisted i.e. success vs failure) and can be used to calculate additional statistic like the *sensitivity* (the true positives) and *specificity* (the true negatives) of the model. The overall model correctly classified **93.75%** of the cases.

The *positive predicted value* i.e. the sensitivity is 95.6%, thus indicating that companies predicted to be as listed this model accurately picked 95.6% of them. Likewise, the *negative predicted value* i.e. the specificity is 60%, indicating that companies predicted to be delisted this model accurately picked 60% of them.

Table. 9 in Appendix ***Variables in the Equation***, gives us information about the contribution and importance of each predictor variable. Variables having sig. values less than .05 are significant in this model also known as the ***Wald*** test. The B values are the values that would be used in an equation to calculate the probability of a case for a specific category and the sign will tell about the direction of relationship.

Exp(B) values are the odds ratio for each variable. Net Gearing has sig. value of .013 having odds ratio of 6.845 and a positive B value. Likewise, BV/Share has sig. value of .001 with odds ratio of .532 and a negative B value.

Estimated equation for FDP model is using Logistic Regression:

$$Z = 4.004 + 1.924 * (\text{Net Gearing}) - 0.631 * (\text{BV/Share})$$

2. Discriminant Analysis

This test is performed to discriminate the successful and failed companies based on the 18 independent variables.

Six out of the eighteen independent variables were found significant as seen in Table. 10 in Appendix ***Test of Equality of Group Means.*** The variable that made statistically significant contribution in the model are ***Asset Turnover, Cash/Share, Net Gearing, NTA/Share, BV/Share*** and ***PER*** out of which NTA/Share and BV/Share were most important predictor variables to discriminate the function.

Table. 11 in Appendix **Eigenvalues** shows an eigenvalue of .66 accounting for 100% of explained variance. The Canonical Correlation linked to this function is .631. On squaring this value i.e. (.631^2 = 0.40), indicating that 40% of the variance in the dependent variable (success or failure) is explained by this model.

Wilk's Lambda is a measure that tells you how well the function separates the cases into specific groups. Table. 12 in Appendix we can see that the function is statistically significant with the Wilk's Lambda being .602.

Standardized Canonical Discriminant Function Coefficient allows you to compare the predictor variable measured in different scale. The coefficient having large absolute values indicate that they have grater discriminating ability. Table. 13 in Appendix we can see Net Gearing, NTA/Share and BV/Share make strong contribution to discriminate the companies into successful and failed groups.

Table. 14 in Appendix provides us with the ***Classification Function Coefficients*** with which the below models are constructed.

For Successful (1)

Y = -9.371 + 2.097 * (Net Gearing) + 0.091 * (NTA/Share) + 0.833 * (BV/Share) + 0.001 * (PER)

For Failed (0)

Y= -17.550 – 8.321 * (Net Gearing) + 0.737 * (NTA/Share) + 1.648 * (BV/Share) – 0.003 * (PER)

Table. 15 in Appendix ***Classification Results*** states that 106 out of 112 companies were correctly classified as being successful. 94.7% of the original grouped cases were correctly classified and resulted in 93.0% correct classification after cross-validation was performed.

3. Artificial Neural Network (ANN)

Multilayer Perceptron was applied to the data set fir getting the ANN output. Covariates in the case were rescaled using standardized method. The data was divided into training, testing and holdout having a ratio of 7:2:1 respectively. ***Net Gearing, Financial Leverage*** and ***Working Cap Turnover*** were found to be most important variables in FDP. Likewise, Asset Turnover was found to be of least importance.

Table. 16 in Appendix shows the distribution of testing, training and holdout sample. We can see class-imbalance problem in this output. This model is not efficient in predicting the failing companies (delisted) as seen in Table. 16 in Appendix. During the training, it has an accuracy of 100% to predict the successful companies with 0% accuracy to predict the failing one's. Similar results were seen on the testing and holdout sample. The overall correct percentage of prediction for training, testing and holdout sample was 94.2%, 89.2% and 94.7% respectively. This model is very efficient in predicting if the company will be successful but fails to predict the failing companies. To remedy this issue hybrid modelling was applied on the sample.

Two hybrid methods were applied to check which one is effective in predicting both the failing and successful firms. The two methods are: i) Hybrid of ANN and LR, and ii) Hybrid of ANN and DA.

4. Hybrid I: ANN + LR

The steps for ANN + LR hybrid analysis as mentioned in methodology was followed that involved using the probabilities and group memberships as second order independent variables. Unlike the results of ANN this hybrid model is efficient at predicting both the successful and failed companies having an accuracy of 100% in predicting the failed ones and 96% accuracy in predicting the successful one during the testing Table. 18 in Appendix. This model had an accuracy of 96.2% during testing. The variable importance chart shows that the probabilities and group membership values help in better prediction as they are the account for the top two important variables. ***Net Gearing*** and ***Gross Debt/CF*** are the most important variables for prediction following the saved probabilities and group memberships.

5. Hybrid II: ANN + DA

The steps for ANN + DA hybrid analysis as mentioned in methodology was followed that involved using the probabilities as second order independent variable. The result of ANN + DA hybrid are like those of ANN. We see the class imbalance in prediction persist in this model. During the training, this model had an accuracy of 100% to predict the successful companies with 0% accuracy to predict the failing one's as shown in Table. 20 in Appendix. Similar results were seen on the testing and holdout sample. The overall correct percentage of prediction for training, testing and holdout sample was 94.1%, 88.2% and 90.0% respectively. These percentages are slightly worse than that of ANN. ***PER*** followed by the saved probability were found to be of importance in this perdition. This model is good to predict successful companies only.

6. Decision Trees

CART method was utilized to build the decision trees. All eighteen variables were included to get the final model. Three variables; ***Financial Leverage, PPE Turnover*** and ***NTA/Share*** were selected as the best for FDP. From the tree diagram below in Fig.1 we can see that Financial Leverage is the predictor use for the primary split, forming Node 2 and Node 3. The split point is 1.07. Further, PPE Turnover is the variable for splitting Node 2 and NTA/Share is for splitting Node 3. This splitting resulted in four terminal nodes. This tree diagram clearly shows that the splitting has occurred using three variables Financial Leverage, PPE Turnover and NTA/Share.

The empirical optimization of the cut-off value on the training sample resulted in a cut of value of 0.93, which is close to 0.9 mark, thus selecting 0.9. The decision tree yielded an average accuracy rate of 53.76% having a specificity and sensitivity of 16.67% and 90.85% respectively. The overall accuracy rate for this model was 85.80%.

As per the variable importance graph as seen in Table. 21 in Appendix, PPE Turnover, EV/EBITDA, Financial Leverage and Asset Turnover were found to be most important having a score of more than 50%.

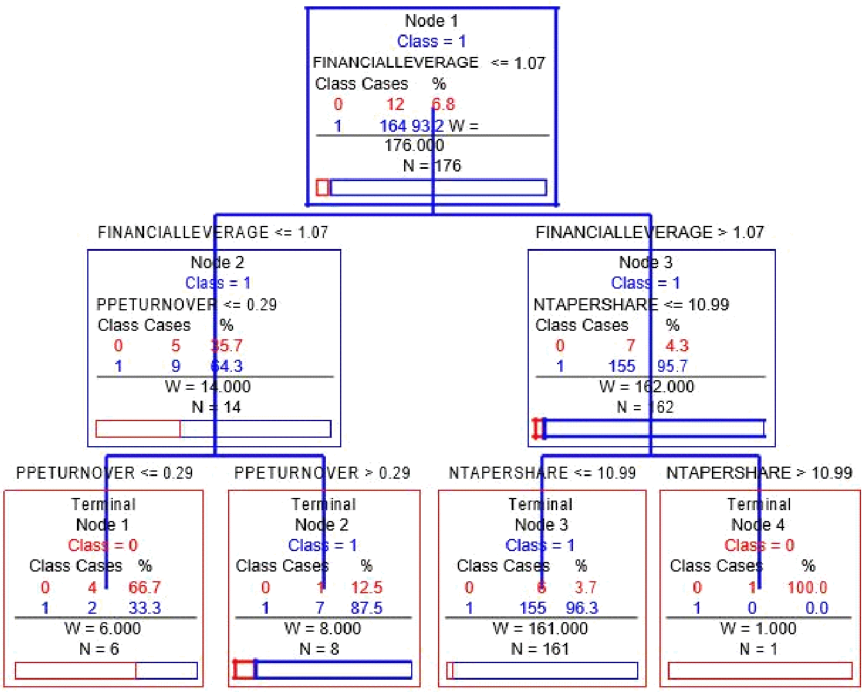

Figure 2: CART Tree Output

7. Random Forest

Experimentation was conducted by generating 200, 500 and 1000 trees, where the 100 trees yielded great results.

The empirical optimization of the cut-off value on the training sample resulted in a cut-off value of 0.5. The random forest model yielded an average accuracy of 66.46%, which is better than the single decision tree. The specificity and sensitivity percentages were 50% and 82.93%. Nevertheless, this model is slightly worse at predicting successful companies when compared to single decision tree, it has better predictive power to predict the failing companies. The overall accuracy rate of this model was 80.68%

As per the variable importance graph i.e. Table. 26 in Appendix, ***NTA/Share, ROE, Financial Leverage, EV/EBITDA, Gross Gearing, BV/Share, ROA*** and ***Net Gearing*** were found to be important having a score of more than 50%.

8. Stochastic Gradient Boosting

TreeNet analysis was performed to generate the output. Experimentation was conducted by generating 200, 500 and 1000 trees, where the 100 trees yielded great results.

The empirical optimization of the cut-off value on the training sample resulted in a cut of value of 0.93, which is close to 0.9 mark, thus selecting 0.9. We can see that the average accuracy rate of this model is 65.96%. This model has the best sensitivity of 90.24% when compared to decision tree and random forest but lags a bit on the specificity with 41.67%. The overall accuracy of this model is 86.93% which means it is better than single decision tree and random forest model.

As per the variable importance graph i.e. Table.29 in Appendix, ***ROE, Gross Gearing, Financial Leverage, NTA/Share, EV/EBITDA*** and ***BV/Share*** were found to be important having a score of more than 50%.

VII. DISCUSSION

FDP using financial ratios is widely used in various studies. This study was restricted in using financial ratios of Australian Real Estate companies publicly available from 2016-2018. This experiment utilized various cutting-edge statistical techniques to determine the best method for FDP. IBM SPSS Statistics 25 was used for performing statistical techniques like Logistic Regression (LR), Discriminant Analysis (DA), Artificial Neural Networks (ANN) and Hybrid Analysis (ANN + LR & ANN + DA). Salford Predictive Modeler (SPM® Version 8.2) was used to run Decision Tree (CART), Random Forest (RF) and TreeNet (Stochastic Gradient Boosting - SGB). The assumptions underlying all the methods were considered before analysis. The results of these various analysis can be discussed in three ways: i) Classification Percentage Comparison, ii) Specificity and Sensitivity Comparison and iii) Important Variable Comparison.

1. Classification Percentage Comparison

Eight out of nine statistical analysis provided classification percentage in two segments i.e. training, testing and holdout (was applicable in ANN analysis). From the Table. 2 below we can clearly see the overall classification accuracy is highest for the hybrid I model i.e. ***ANN + LR*** with ***96.2%*** for the testing sample.

Overall Classification Percentage			
Analysis	**Training**	**Testing**	**Holdout**
1. LR	93.7%	93.7%	-
2. DA	93.0%	93.0%	-
3. ANN	94.2%	89.2%	94.7%
4. ANN + LR	96.6%	96.2%	90.0%
5. ANN + DA	94.1%	88.2%	90.0%
6. CART	91.48%	85.80%	-
7. RF	80.68%	80.685	-
8. SGB	86.93%	86.93%	-

Table 2: Classification Comparison Table

2. Specificity and Sensitivity Comparison

Sensitivity is equivalent to the true positive rate in this case we will look for number of correctly identified successful companies by the model. Mathematically, it can be calculated as: Sensitivity = (Number of True Positive Outcome i.e. correctly identified successful companies)/(Total Positive Outcomes).

Specificity is equivalent to the true negative rate in this case we will look for number of correctly identified failed companies by the model. Mathematically, it can be calculated as: Specificity = (Number of True Negative Outcome i.e. correctly identified failed companies)/(Total Positive Outcomes).

From the Table. 3 below we see the sensitivity rate is best for the hybrid I model i.e. ANN + LR and likewise the specificity rate is best for TreeNet model i.e. SGB with 100% value. Looking at Table. 2 & 3 together we can see that two models perform well in predicting both the successful and failed companies. The two models which outperform other are***: i) Hybrid I Model (ANN + LR)*** and ***ii) Stochastic Gradient Boosting Model (TreeNet).*** These two models have great classification accuracy (ANN + LR = 96.2% & SGB = 86.93%) along with good rate of specificity (ANN + LR = 50% & SGB = 100%) and sensitivity (ANN + LR = 100% & SGB = 85.98%).

Analysis	Specificity	Sensitivity
1. LR	60.00%	95.60%
2. DA	0.00%	92.98%
3. ANN	0.00%	89.12%
4. ANN + LR	**50.00%**	**100.00%**
5. ANN + DA	0.00%	88.23%
6. CART	16.67%	90.85%
7. RF	50.00%	82.93%
8. SGB	**100.00%**	**85.98%**

Table 3: Specificity v/s Sensitivity

3. Important Variable Comparison.

Out of the nine statistical test eight provided list of important independent variables that can be used for predicting the model outcome. Table. 4 below helps compare the list of important variables for each analysis. Looking at the trend four variables were found to be repeated in five out of eight tests. Those variables are: ***i) Financial Leverage, ii) Net Gearing, iii) NTA/Share*** and ***iv) BV/Share.***

Independent Variables	Type of Analysis							
	LR	DA	ANN	ANN + LR	ANN + DA	CART	RF	SGB
Return on Equity							×	×
Return on Asset				×	×		×	
Asset Turnover						×		
PPE Turnover						×		
Deprecation/PP&E								
Working Cap Turnover			×					
Gross Gearing							×	×
Financial Leverage			×	×		×	×	×
Current Ratio				×				
Quick Ratio			×					
Gross Debt/CF				×				
Cash/Share								
Net Gearing	×	×	×	×			×	
NTA/Share		×		×	×		×	×
BV/Share	×	×		×			×	×
Sales/Share								
EV/EBITDA						×	×	×
PER		×	×		×			

Table 4: Important Variable Comparison

VIII. CONCLUSION & FUTURE WORK

Considering the volatility of the Real Estate sector and its impact on the overall economic growth this paper focused in constructing model which can help predict the future (success or failure) of Australian Real Estate companies. The models constructed in this paper can be used by the companies to determine their current financial status also it can be effectively used by investor to determine if their investing in the right company.

Cutting-edge statistical techniques were employed to find the best model for financial distress prediction. The results indicated that Stochastic Gradient Boosting Model is best at predicting the failing companies and likewise hybrid model of Artificial Neural Network and Logistic Regression are best at predicting succeeding companies. Out of the four variables found to be important: i) Financial Leverage, ii) Net Gearing, iii) NTA/Share and iv) BV/Share; these two models had all except Net Gearing in their important variable list.

This study was restricted to the Real Estate sector of Australia, but this can be easily extended to other sectors and different countries in the world. The size of the sample was not huge and hence this study can be conducted on a larger sample set in future to verify the outcome. Thus, taking this study as foundation lots of future research can emerge from the same.

This paper focused on predicting success or failure of Australian Real Estate company, but in future this kind of predictive models can be constructed for bankruptcy prediction or credit risk modelling etc. Hence, there is a huge scope of future work in this area.

IX. REFERENCES

- A research report submitted Kavir D. Rama. *An empirical evaluation of the Altman (1968) failure prediction model on south African JSE listed companies.*
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance, 23*(4), 589-609. Retrieved from http://www.econis.eu/PPNSET?PPN=374118043.
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*, 71-111. doi:10.2307/2490171.
- Beneish, M. D. (1997). Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance. *Journal of Accounting and Public Policy, 16*(3), 271-309. doi:10.1016/S0278-4254(97)00023-9.
- Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study doi://doi.org/10.1016/j.dss.2010.08.008.
- Breiman, L. (2017). Classification and regression trees (1st ed.). Boca Raton:
- Routledge. Retrieved from http://www.crcnetbase.com/isbn/9781351460491.
- Chandra, D. K., Ravi, V., & Bose, I. (2009b). Failure prediction of dotcom companies using.
- hybrid intelligent techniques. *Expert Systems with Applications, 36*(3), 4830-4837. doi:10.1016/j.eswa.2008.05.047.
- Chen, M. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications, 38*(9), 11261-11272. doi:10.1016/j.eswa.2011.02.173.
- Coats, P. K., & Fant, L. F. (1991). A neural network approach to forecasting financial distress. *Journal of Business Forecasting, 10*(4), 9.
- Collins, R. A., & Green, R. D. (1982). Statistical methods for bankruptcy forecasting. *Journal of Economics and Business, 34*(4), 349-354. doi:10.1016/0148-6195(82)90040-6.
- Conference Paper · July 1990.A neural network model for bankruptcy prediction doi:10.1109/IJCNN.1990.137710.
- Demyanyk, Y., & Hasan, I. (2010). Financial crises and bank failures: A review of prediction methods. *Omega, 38*(5), 315-324. doi:10.1016/j.omega.2009.09.007.
- Emerge Capital View. (a). *Australian real estate: Heating up.*
- Fantazzini, D., & Figini, S. (2009). Random survival forests models for SME credit risk measurement. *Methodology and Computing in Applied Probability, 11*(1), 29-45. doi:10.1007/s11009-008-9078-2.

- FitzPatrick, P. J. (1932). A comparison of the ratios of successful industrial enterprises with those of failed companies.
- Fletcher, D., & Goss, E. (1993). Forecasting with neural networks: An application using bankruptcy data. *Information & Management, 24*(3), 159-167. doi:10.1016/0378-7206(93)90064-Z.
- Gamesalingam, S., & Kumar, K. (2001). Detection of financial distress via multivariate statistical analysis. *Managerial Finance, 27*(4), 45-55. doi:10.1108/03074350110767132.
- Gepp, A. (2015). Financial statement fraud detection using supervised learning methods. ePublications@bond. Retrieved from http://epublications.bond.edu.au/theses/178.
- Gepp, A., & Kumar, K. (2012). Business failure prediction using statistical techniques:
- A review. ePublications@bond. Retrieved.
- from http://epublications.bond.edu.au/business_pubs/675.
- Gepp, A., Kumar, K., & Bhattacharya, S. (2010). Business failure prediction using decision trees. *Journal of Forecasting, 29*(6), 536-555. doi:10.1002/for.1153.
- Hall, G. (1994). Factors distinguishing survivors from failures amongst small firms in the UK construction sector. *Journal of Management Studies, 31*(5), 737-760. doi:10.1111/j.1467-6486.1994.tb00637.x.
- Halteh, K., Kumar, K., & Gepp, A. (2018a). Financial distress prediction of Islamic banks using tree-based stochastic techniques. *Managerial Finance, 44*(6), 759-773. doi:10.1108/MF-12-2016-0372.
- Halteh, K., Kumar, K., & Gepp, A. (2018b). Using cutting-edge tree-based stochastic models to predict credit risk. *Risks, 6*(2), 55. doi:10.3390/risks6020055.
- Hamer, M. M. (1983). Failure prediction: Sensitivity of classification accuracy to alternative statistical methods and variable sets. *Journal of Accounting and Public Policy, 2*(4), 289-307. doi:10.1016/0278-4254(83)90032-7.
- Hosmer, D. W., & Lemeshow, S. (2000). Applied logistic regression (2. ed. ed.). New York: Wiley.
- Hung, C., & Chen, J. (2009). A selective ensemble based on expected probabilities for bankruptcy prediction. *Expert Systems with Applications, 36*(3), 5297-5303. doi:10.1016/j.eswa.2008.06.068.
- *Investopedia.* (2012). Retrieved, 2018, from https://www.investopedia.com/
- Kumar, K., & Chaturvedi, A. (2010). Some Recent Developments in Statistical Theory and Applications. Florida: Brown Walker Press.
- Kumar, P., Ravi, V., Kasabov, N., & Srinivas, E. (2008). A semi-online training algorithm for the radial basis function neural networks: Applications to bankruptcy prediction in banks.
- Laitinen, T., & Kankaanpaa, M. (1999). Comparative analysis of failure prediction methods: The Finnish case. *European Accounting Review, 8*(1), 67-92. doi:10.1080/096381899336159.
- Martin, D. (1977). Early warning of bank failure: A logit regression approach. *Journal of Banking and Finance, 1*(3), 249-276. doi:10.1016/0378-4266(77)90022-X.
- McCulloch, W. S., & Pitts, W. (1990). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology, 52*(1), 99-115. doi:10.1016/S0092-8240(05)80006-0.
- MorningStar® Data Analysis; Retrieved from https://datanalysis.morningstar.com.au/ftl/search/currentcriteria?xtm-licensee=datpremium
- Nanni, L., & Lumini, A. (2009). An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications, 36*(2), 3028-3033. doi:10.1016/j.eswa.2008.01.018.
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research, 18*(1), 109-131. Retrieved
- from http://www.econis.eu/PPNSET?PPN=479611866.
- Paul Joseph. Fitzpatrick. (1932). A comparison of the ratios of successful industrial enterprises with those of failed companies. Washington:
- Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice & Theory, 30*(2), 19-50. doi:10.2308/ajpt-50009.
- Pranjut Gogoi. Knoldus Blog: First Interaction Artificial Neural Network. Retrieved.

- from https://blog.knoldus.com/first-interaction-artificial-neural-network/
- Ravi Kumar, P., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research, 180*(1), 1-28. doi:10.1016/j.ejor.2006.08.043.
- *Real estate outlook for 2018* (2018). London: Incisive Media Limited. Retrieved from https://search.proquest.com/docview/1987642248.
- Salchenberger, L. M., Cinar, E. M., & Lash, N. A. (1992). Neural networks: A new tool for predicting thrift failures. *Decision Sciences, 23*(4), 899-916. doi:10.1111/j.1540-5915.1992.tb00425.x.
- Shah, N. (2014). Developing financial distress prediction models using cutting edge recursive partitioning techniques: A study of Australian mining performance. *Review of Integrative Business and Economics Research, 3*(2), 103-143. Retrieved from https://search.proquest.com/docview/1553397564.
- Spathis, C., Doumpos, M., & Zopounidis, C. (2002). Detecting falsified financial statements: A comparative study using multicriteria analysis and multivariate statistical techniques.*European Accounting Review, 11*(3), 509-535. doi:10.1080/0963818022000000966.
- Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science, 38*(7), 926-947. doi:10.1287/mnsc.38.7.926.
- Zhang, G., Y. Hu, M., Eddy Patuwo, B., & C. Indro, D. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research, 116*(1), 16-32. doi:10.1016/S0377-2217(98)00051-4.
- Zhang, H., & Singer, B. (2010). Recursive partitioning and applications.

X. APPENDIX

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	8.473	1	.004
	Block	8.473	1	.004
	Model	8.473	1	.004
Step 2	Step	7.400	1	.007
	Block	15.873	2	.000
	Model	15.873	2	.000
Step 3	Step	6.092	1	.014
	Block	21.965	3	.000
	Model	21.965	3	.000

Table 5: Omnibus Tests of Model Coefficients

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	6.743	7	.456
2	6.008	8	.646
3	5.871	8	.662

Table 6: Hosmer and Lemeshow Test

Model Summary			
Step	**-2 Log likelihood**	**Cox & Snell R Square**	**Nagelkerke R Square**
1	41.509a	.085	.209
2	34.110b	.154	.376
3	28.018c	.206	.505
a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.			
b. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.			
c. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.			

Table 7: Model Summary (Logistic Regression)

Classification Table [a]					
Observed			**Predicted**		
			Listed/Delisted		**Percentage Correct**
			0	**1**	
Step 1	Listed/Delisted	**0**	1	6	14.3
		1	0	88	100.0
	Overall Percentage				93.7
Step 2	Listed/Delisted	**0**	2	5	28.6
		1	2	86	97.7
	Overall Percentage				92.6
Step 3	Listed/Delisted	**0**	3	4	42.9
		1	2	86	97.7
	Overall Percentage				93.7
a. The cut value is .500					

Table 8: Classification Table (Logistic Regression)

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	*Upper*
Step 1a	BV/Share	-.321	.111	8.351	1	.004	.725	.583	.902
	Constant	3.794	.717	27.979	1	.000	44.452		
Step 2b	Net Gearing	1.824	.740	6.074	1	.014	6.194	1.453	26.411
	BV/Share	-.480	.141	11.599	1	.001	.619	.470	.816
	Constant	4.083	.852	22.988	1	.000	59.308		
Step 3c	Asset Turnover	6.264	3.731	2.818	1	.403	525.291	.350	787691.628
	Net Gearing	1.924	.778	6.107	1	.013	6.845	1.489	31.473
	BV/Share	-.631	.184	11.784	1	.001	.532	.371	.763
	Constant	4.004	.997	16.141	1	.000	54.814		
a. Variable(s) entered on step 1: BV/Share.									
b. Variable(s) entered on step 2: Net Gearing.									
c. Variable(s) entered on step 3: Asset Turnover.									

Table 9: Variables in Equation (Logistic Regression)

Tests of Equality of Group Means					
	Wilks' Lambda	F	df1	df2	Sig.
ROE	1.000	.012	1	93	.913
ROA	.995	.511	1	93	.477
Asset Turnover	**.982**	**1.659**	**1**	**93**	**.021**
PPE Turnover	.989	.998	1	93	.320
Depreciation/PP&E	.999	.071	1	93	.791
Working Cap Turnover	1.000	.007	1	93	.932
Gross Gearing	.997	.319	1	93	.574
Financial Leverage	.999	.064	1	93	.801
Current Ratio	.999	.082	1	93	.775
Quick Ratio	1.000	.012	1	93	.912
Gross Debt/CF	.996	.355	1	93	.553
Cash/Share	**.980**	**1.918**	**1**	**93**	**.039**
Net Gearing	**.974**	**2.434**	**1**	**93**	**.022**
NTA/Share	**.892**	**11.258**	**1**	**93**	**.001**
BV/Share	**.875**	**13.310**	**1**	**93**	**.000**
Sales/Share	.993	.636	1	93	.427
EV/EBITDA	.991	.848	1	93	.359
PER	**.903**	**9.996**	**1**	**93**	**.002**

Table 10: Tests of Equality of Group Means

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.660a	100.0	100.0	.631
a First 1 canonical discriminant functions were used in the analysis.				

Table 11: Eigenvalues

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.602	42.577	4	.001

Table 12: Wilk's Lambda

Standardized Canonical Discriminant Function Coefficients	
	Function
	1
Net Gearing	-2.138
NTA/Share	.712
BV/Share	.700
PER	-.344

Table 13: Standardized Canonical Discriminant Function Coefficients

Classification Function Coefficients		
	Listed/Delisted	
	0	1
Net Gearing	-8.321	2.097
NTA/Share	.737	.091
BV/Share	1.648	.833
PER	-.003	.001
(Constant)	-17.550	-9.371
Fisher's linear discriminant functions		

Table 14: Classification Function Coefficients

Classification Results [a, c]					
		Listed/Delisted	Predicted Group Membership		Total
			0	1	
Original	Count	**0**	2	6	8
		1	0	106	106
	%	**0**	25.0	75.0	100.0
		1	.0	100.0	100.0
Cross-validated [b]	Count	**0**	0	8	8
		1	0	106	106
	%	**0**	.0	100.0	100.0
		1	.0	100.0	100.0
a. 94.7% of original grouped cases correctly classified.					
b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.					
c. 93.0% of cross-validated grouped cases correctly classified.					

Table 15: Classification Results (Discriminant Analysis)

Model Summary		
Training	Cross Entropy Error	21.859
	Percent Incorrect Predictions	5.8%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error [a]
	Training Time	0:00:00.06
Testing	Cross Entropy Error	11.466
	Percent Incorrect Predictions	10.8%
Holdout	Percent Incorrect Predictions	5.3%
Dependent Variable: Listed/Delisted		
a. Error computations are based on the testing sample.		

Table 16: Model Summary ANN

Classification				
Sample	**Observed**	**Predicted**		
		0	**1**	Percent Correct
Training	**0**	0	7	0.0%
	1	0	113	100.0%
	Overall Percent	0.0%	100.0%	94.2%
Testing	**0**	0	4	0.0%
	1	0	33	100.0%
	Overall Percent	0.0%	100.0%	89.2%
Holdout	**0**	0	1	0.0%
	1	0	18	100.0%
	Overall Percent	0.0%	100.0%	94.7%

Dependent Variable: Listed/Delisted

Table 17: Classification ANN

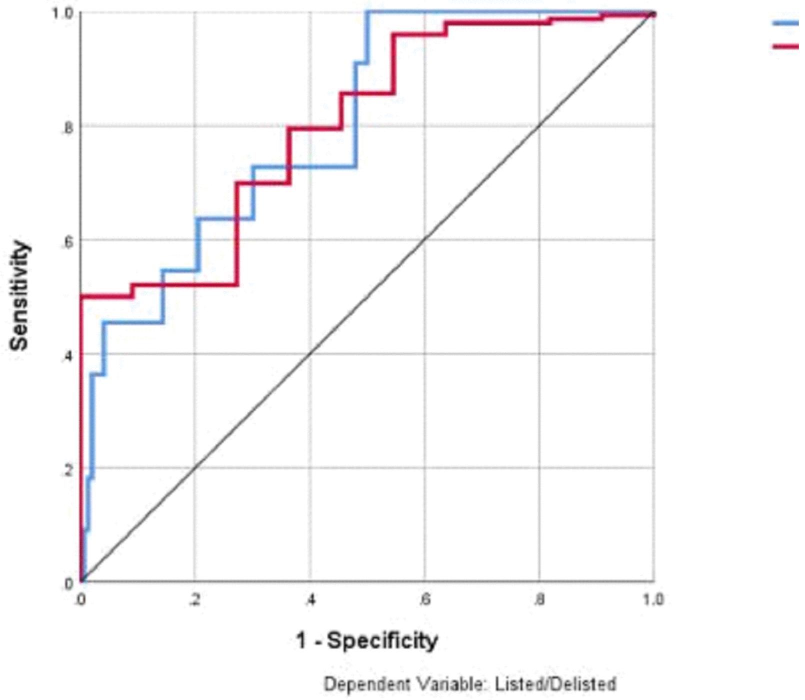

Figure 1: ANN Sensitivity vs Specificity

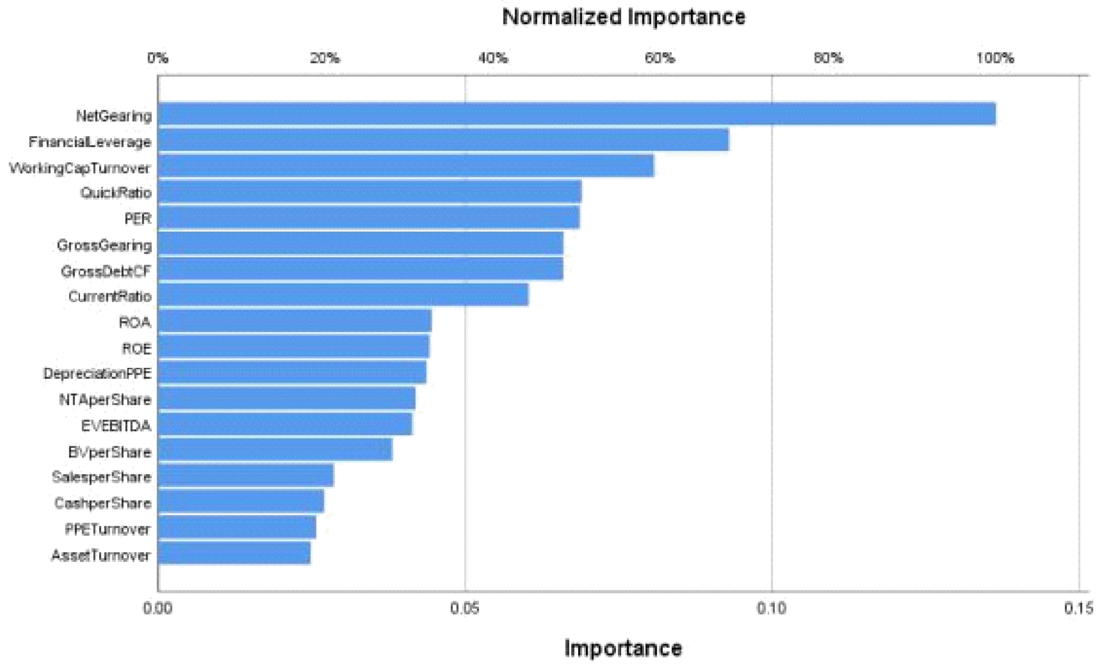

Figure 2: ANN Variable Importance Graph

Model Summary		
Training	Cross Entropy Error	3.473
	Percent Incorrect Predictions	3.4%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error [a]
	Training Time	0:00:00.02
Testing	Cross Entropy Error	1.296
	Percent Incorrect Predictions	3.8%
Holdout	Percent Incorrect Predictions	10.0%
Dependent Variable: Listed/Delisted		
a. Error computations are based on the testing sample.		

Table 18: Model Summary (ANN + LR)

Classification				
Sample	**Observed**	**Predicted**		
		0	**1**	**Percent Correct**
Training	**0**	4	0	100.0%
	1	2	53	96.4%
	Overall Percent	10.2%	89.8%	96.6%
Testing	**0**	1	0	100.0%
	1	1	24	96.0%
	Overall Percent	7.7%	92.3%	96.2%
Holdout	**0**	1	1	50.0%
	1	0	8	100.0%
	Overall Percent	10.0%	90.0%	90.0%
Dependent Variable: Listed/Delisted				

Table 19: Classification (ANN+LR)

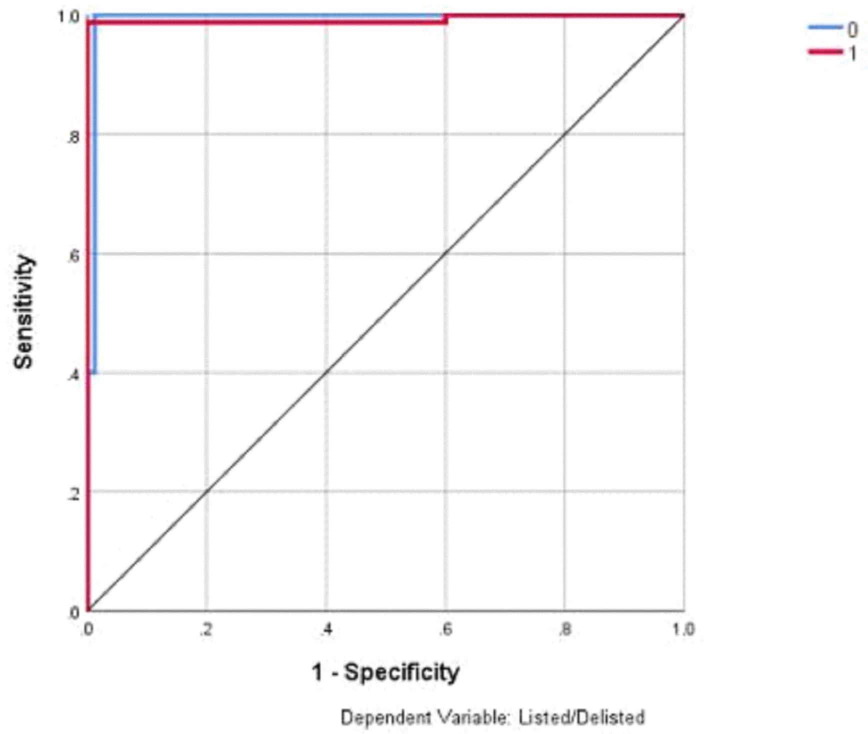

Figure 2: ANN+LR Sensitivity vs Specificity

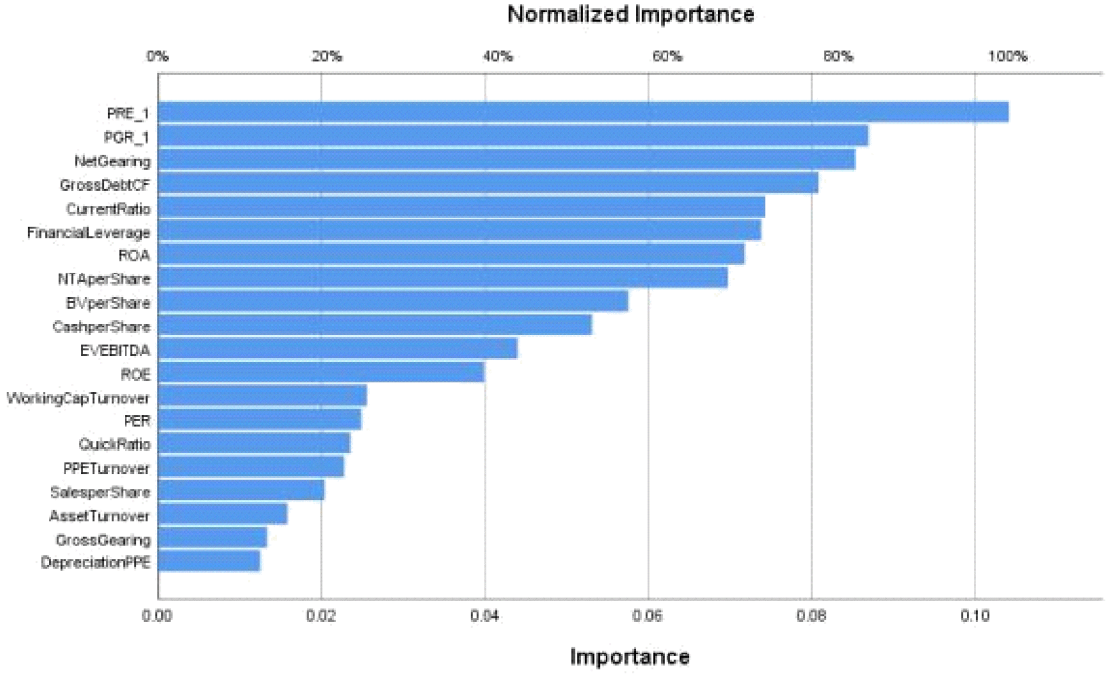

Figure 3: ANN+LR Variable Importance Graph

Model Summary		
Training	Cross Entropy Error	6.751
	Percent Incorrect Predictions	5.9%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error [a]
	Training Time	0:00:00.00
Testing	Cross Entropy Error	1.760
	Percent Incorrect Predictions	11.8%
Holdout	Percent Incorrect Predictions	10.0%
Dependent Variable: Listed/Delisted		
a. Error computations are based on the testing sample.		

Table 20: Model Summary (ANN+DA)

Classification				
Sample	**Observed**	**Predicted**		
		0	**1**	Percent Correct
Training	**0**	0	4	0.0%
	1	0	64	100.0%
	Overall Percent	0.0%	100.0%	94.1%
Testing	**0**	0	2	0.0%
	1	0	15	100.0%
	Overall Percent	0.0%	100.0%	88.2%
Holdout	**0**	0	1	0.0%
	1	0	9	100.0%
	Overall Percent	0.0%	100.0%	90.0%
Dependent Variable: Listed/Delisted				

Table 21: Classification (ANN+DA)

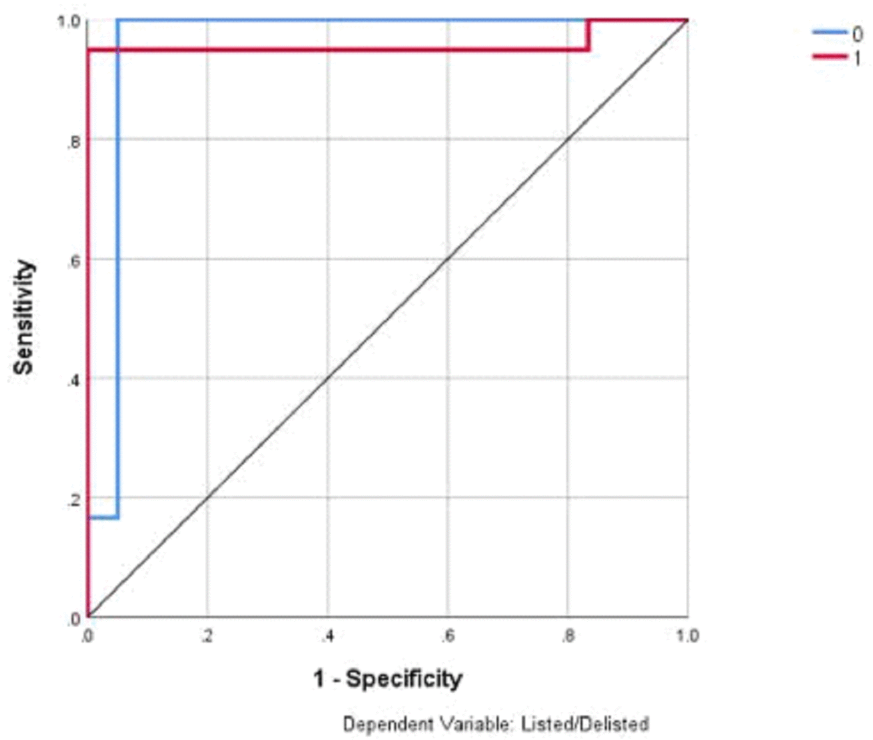

Figure 4: ANN+DA Sensitivity vs Specificity

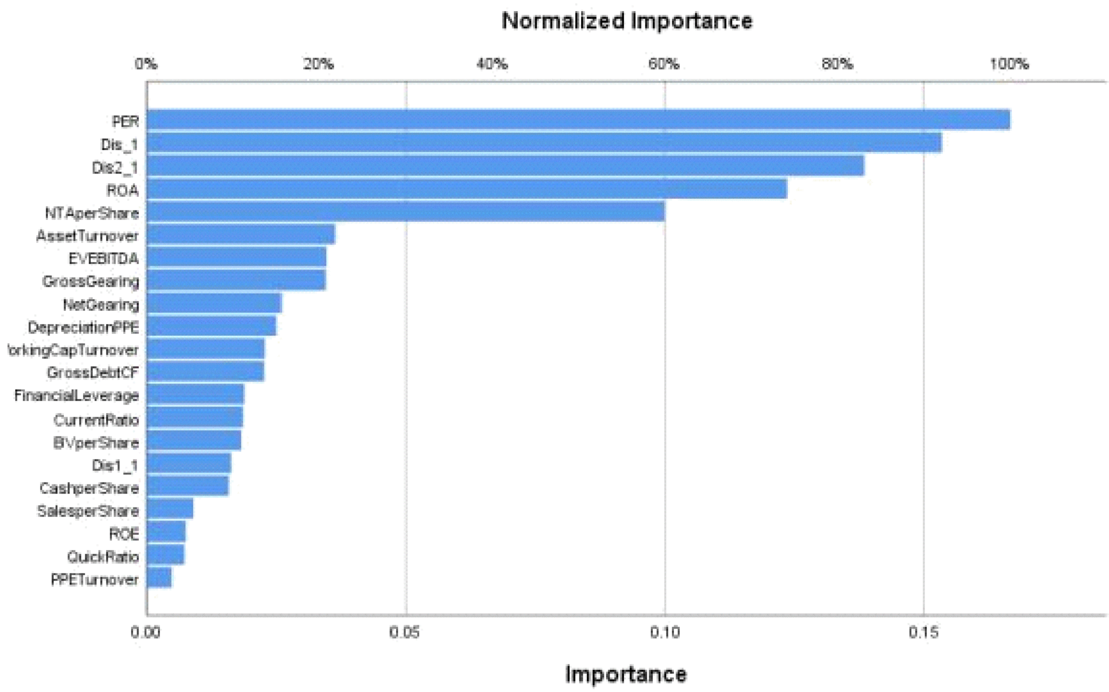

Figure 5: ANN+DA Variable Importance Graph

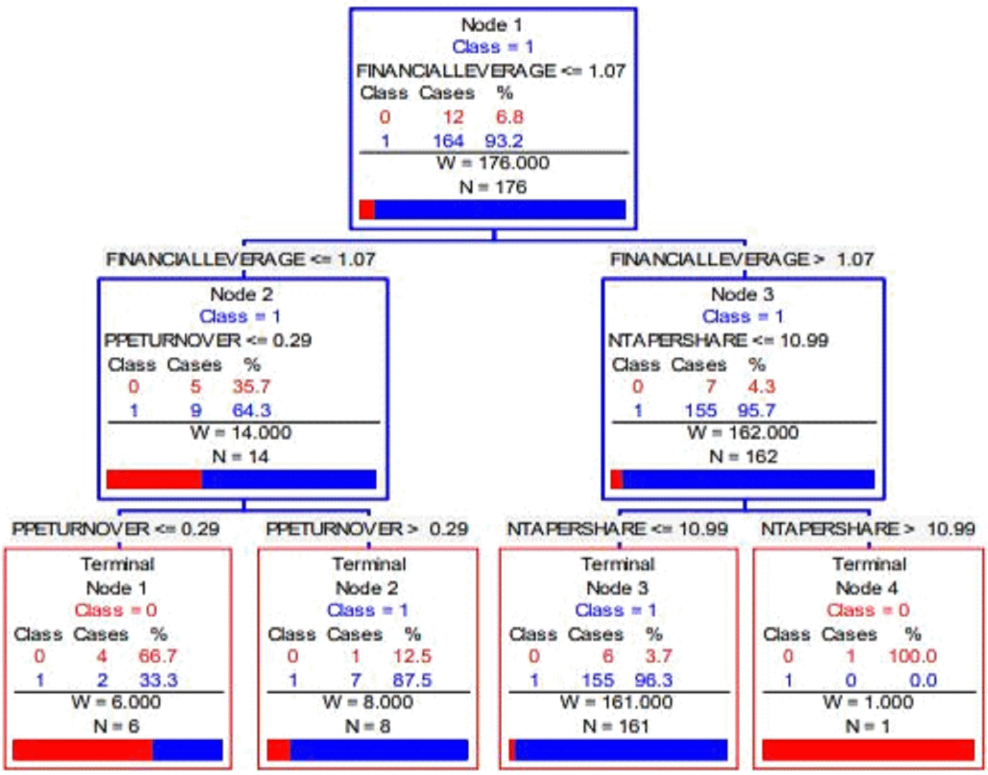

Figure 6: CART Tree Output

Variable	Score	
PPETURNOVER	100.00	
EVEBITDA	85.42	
FINANCIALLEVERAGE	71.12	
ASSETTURNOVER	47.52	
BVPERSHARE	43.74	
NTAPERSHARE	43.52	
CURRENTRATIO	22.81	
ROA	21.03	
QUICKRATIO	1.14	
NETGEARING	0.22	
GROSSGEARING	0.22	
ROE	0.10	

Table 22: CART Variable Importance

Actual Class	Total Class	Percent Correct	Predicted Classes	
			0 N = 15	1 N = 161
0	12	50.00%	6	6
1	164	94.51%	9	155
Total:	176			
Average:		72.26%		
Overall % Correct:		91.48%		
Specificity		50.00%		
Sensitivity/Recall		94.51%		
Precision		96.27%		
F1 statistic		95.38%		

Table 23: Confusion Matrix – Learn (CART)

	Actual Class	Total Class	Percent Correct	Predicted Classes	
				0 N = 17	1 N = 159
	0	12	16.67%	2	10
	1	164	90.85%	15	149
	Total:	176			
	Average:		53.76%		
	Overall % Correct:		85.80%		
	Specificity		16.67%		
	Sensitivity/Recall		90.85%		
	Precision		93.71%		
	F1 statistic		92.26%		

Table 24: Confusion Matrix – Test (CART)

	Class	N Cases	N Mis-Classed	Pct. Error	Cost
	0	12	6	50.00%	0.50000
	1	164	9	5.49%	0.05488

Table 25: Misclassification – Learn (CART)

	Class	N Cases	N Mis-Classed	Pct. Error	Cost
	0	12	10	83.33%	0.83333
	1	164	18	10.98%	0.10976

Table 26: Misclassification – Test (CART)

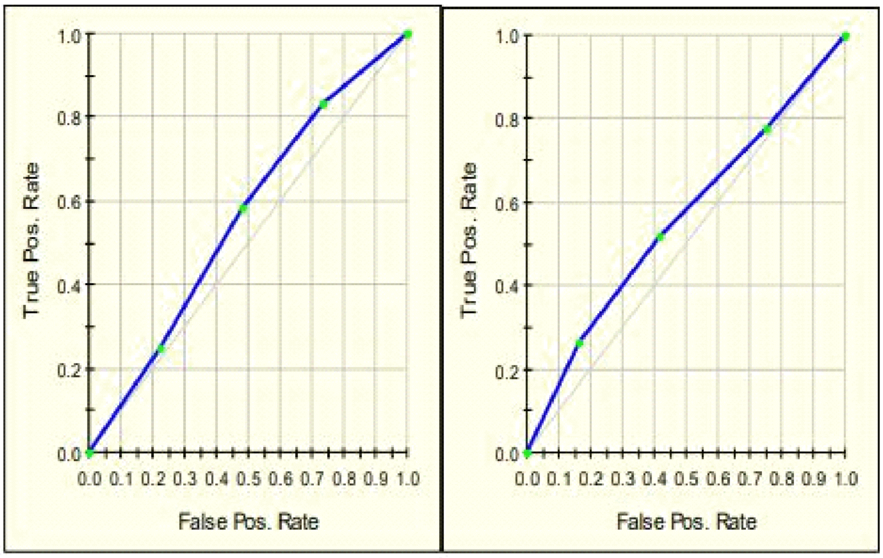

Figure 7 – Gains Chart – ROC, Target class: 0 vs 1 (CART)

Variable	Score	
NTAPERSHARE	100.00	
ROE	87.21	
FINANCIALLEVERAGE	81.91	
EVEBITDA	81.29	
GROSSGEARING	68.66	
BVPERSHARE	61.45	
ROA	51.81	
NETGEARING	49.63	
CASHPERSHARE	37.80	
PER	37.65	
CURRENTRATIO	35.67	
WORKINGCAPTURNOVER	35.11	
GROSSDEBTCF	32.35	
ASSETTURNOVER	29.36	
QUICKRATIO	28.78	
PPETURNOVER	17.67	
DEPRECIATIONPPE	15.34	
SALESPERSHARE	9.04	

Table 27: Random Forest Variable Importance

	Actual Class	Total Class	Percent Correct	Predicted Classes	
				0 N=34	1 N=142
	0	12	50.00%	6	6
	1	164	82.93%	28	136
	Total:	176			
	Average:		66.46%		
	Overall % Correct:		80.68%		
	Specificity		50.00%		
	Sensitivity/Recall		82.93%		
	Precision		95.77%		
	F1 statistic		88.89%		

Table 28: Confusion Matrix – Random Forest

	Class	N Cases	N Mis-Classed	Pct. Error	Cost
	0	12	0	0.00%	0.00000
	1	164	153	93.29%	0.93293

Table 29: Misclassification – Random Forest

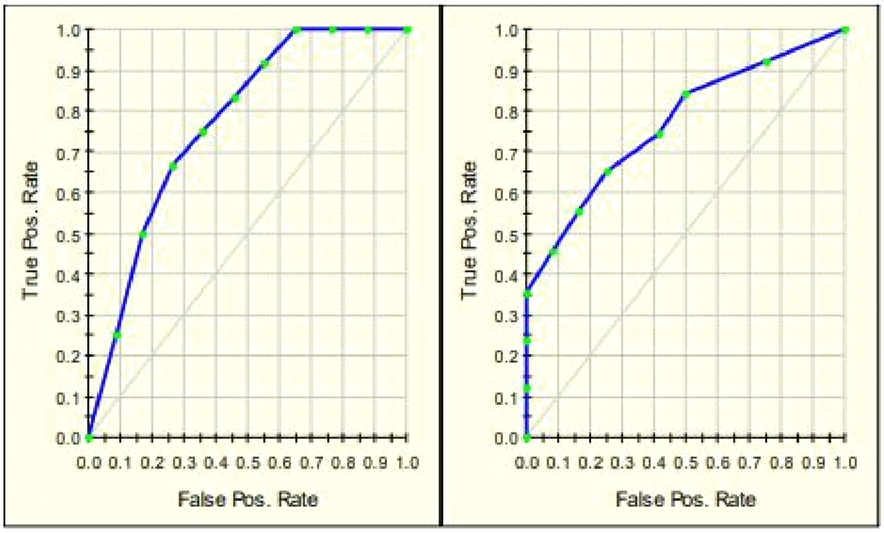

Figure 8: Gains Chart – ROC, Target class: 0 vs 1 (Random Forest)

Variable	Score
ROE	100.00
GROSSGEARING	81.11
FINANCIALLEVERAGE	75.84
NTAPERSHARE	71.70
EVEBITDA	67.86
BVPERSHARE	52.20
PER	46.95
CURRENTRATIO	41.67
ASSETTURNOVER	38.93
GROSSDEBTCF	37.96
CASHPERSHARE	36.29
PPETURNOVER	29.17
WORKINGCAPTURNOVER	25.54
ROA	25.02
NETGEARING	21.73
DEPRECIATIONPPE	20.47
SALESPERSHARE	16.50
QUICKRATIO	14.42

Table 30: TreeNet Variable Importance (SGB)

Actual Class	Total Class	Percent Correct	Predicted Classes	
			0 N = 35	1 N = 141
0	12	100.00%	12	0
1	164	85.98%	23	141
Total:	176			
Average:		92.99%		
Overall % Correct:		86.93%		
Specificity		100.00%		
Sensitivity/Recall		85.98%		
Precision		100.00%		
F1 statistic		92.46%		

Table 31: Confusion Matrix – Learn (SGB – TreeNet)

	Actual Class	Total Class	Percent Correct	Predicted Classes	
				0 N=21	1 N=155
	0	12	41.67%	5	7
	1	164	90.24%	16	148
	Total:	176			
	Average:		65.96%		
	Overall % Correct:		86.93%		
	Specificity		41.67%		
	Sensitivity/Recall		90.24%		
	Precision		95.48%		
	F1 statistic		92.79%		

Table 32: Confusion Matrix – Test (SGB – TreeNet)

	Class	N Cases	N Mis-Classed	Pct. Error	Cost
	0	12	0	0.00%	0.00000
	1	164	23	14.02%	0.14024

Table 33: Misclassification – Learn (SGB – TreeNet)

	Class	N Cases	N Mis-Classed	Pct. Error	Cost
	0	12	6	50.00%	0.50000
	1	164	34	20.73%	0.20732

Table 34: Misclassification – Test (SGB – TreeNet)

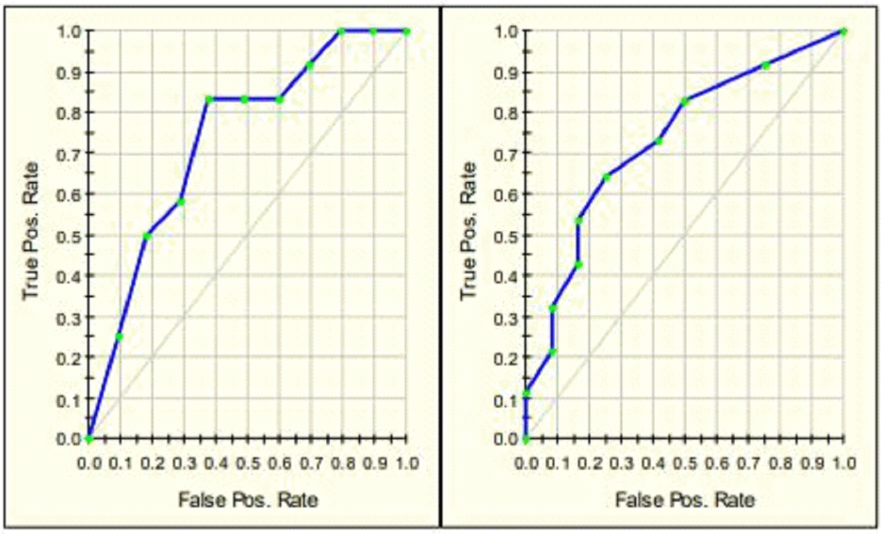

Figure 9: Gains Chart – ROC, Target class: 0 vs 1 (TreeNet – SGB)

❑ ❑ ❑