

Corporate distress prediction using random forest and tree net for india

¹Arvind Shrivastava, ¹Nitin Kumar, ¹Kuldeep Kumar, ¹Sanjeev Gupta

¹The views expressed in this paper are strictly personal and not related to authors' respective institutional affiliation.

Corresponding author.

Correspondence: Arvind Shrivastava

E-mail: arvind1aug@gmail.com

Article info

Received 2nd January 2020

Received in revised form

15th February 2020

Accepted

20th February 2020

Keywords

financial distress,
insolvency risk,
prediction, random forest,
tree net

Abstract

The paper deals with the Random Forest, a popular classification machine learning algorithm to predict bankruptcy (distress) for Indian firms. Random Forest orders firms according to their propensity to default or their likelihood to become distressed. This is also useful to explain the association between the tendency of firm failure and its features. The results are analyzed vis-à-vis Tree Net. Both in-sample and out of sample estimations have been performed to compare Random Forest with Tree Net, which is a cutting edge data mining tool known to provide satisfactory estimation results. An exhaustive data set comprising companies from varied sectors have been included in the analysis. It is found that Tree Net procedure provides improved classification and predictive performance vis-à-vis Random Forest methodology consistently that may be utilized further by industry analysts and researchers alike for predictive purposes.

JEL Codes: G33, C53, L2

1. Introduction

Entrepreneurship and robust corporate institutions are vital assets of modern states that create value and employment opportunities [1, 2]. However, delay in identifying and taking corrective steps for weak performers may lead to insolvency and avoidable intervention. Timely detection of stressed firms is useful input for appropriate policy actions. Accordingly, a range of statistical methodologies have been applied by researchers to address this issue like z-scores, econometric tools, operations research framework, other multivariate techniques etc.

In its seminal work, Altman [3] (1968) devised a measurement score viz., Z-score based on multiple discriminant technique to assess the bankruptcy. Extensive usage of logistic modeling has also been applied for categorizing and predicting failure of firms [4-7] (Shrivastava, 2018; Hua et al., 2007; Altman et al., 1994; Ohlson, 1980). Li et al. [8] dwelled on a unique approach, wherein they applied Data Envelopment Analysis to derive efficiency. Thereafter, failure/success was modeled as logistic regression with derived efficiencies. Bapat and Nagale [9] performed an assessment of forecasting prowess of various statistical techniques for listed firms in Indian exchanges like logistic, discriminant, and neural network. Advanced statistical strategies like decision trees, Random Forest (RF), and stochastic gradient boosting has been applied by Halteh et al. [10] to improve the understanding about credit-risk phenomenon.

Bagging, also known as bootstrap aggregation is a widely known and experimented machine learning methodology for decreasing the volatility for projected mathematical formulation applied by another school of researchers. It is established to provide satisfactory output in case of elevated volatility and small-bias techniques e.g. presence of trees in the data set. Boosting dominates bagging in many relevant cases and therefore is a popular and preferred choice among researchers [11, 12]. In a significant enhancement of bagging method is the RF approach proposed by Breiman [13], wherein large collection of de-correlated trees are built and subsequently those are averaged. The RF methods perform well in many real problems as compared to boosting and is convenient to train and tune as per situation. This has resulted in wide popularity of RF technique, implemented in variety of scenarios.

Strong financial performance of firm is crucial not only for its existence in competitive market but also for shareholders and society at large. However, complexity involved in operations of firm and

asymmetric market information may lead to financial distress and bankruptcy. Effective management of random shocks leads to successful concerns. Ignoring the risks may lead to insolvencies for vulnerable companies. The study enriches understanding of the bankruptcy (distress) prediction for Indian corporate which otherwise has scant studies delving into RF approach. It applies the said technique that is profusely applied procedure for machine learning. Utilizing various firm specific parameters, both RF and Tree Net (TN) methodologies have been reviewed for both out of sample and in-sample examination. It is found that TN method has been producing improved classification vis-a-vis RF approach. TN displays superior predictive performance in contrast to RF method outperforming it consistently for all future periods. The remainder of paper is arranged as below. The next section explains the technical approach followed by data explanation in Section 3. The result of analysis is presented in Section 4 with Section 5 summarizing the entire study.

2. Methodology

This section elaborates both the RF and TN classification techniques, which are common ensemble learning procedures for grouping data points applied herein. Beginning with RF, it ranks covariates taking into consideration the performance of predictive ability. The work of Breiman [14] is landmark for RF algorithm. Additionally, remarkable contribution in the development of RF algorithms have been done by Dietterich (2000), Ho (15) and Amit and Geman (16) using randomization to grow a forest. The RF procedure has been applied in varied fields. To construct credit risk models, Khandani et al. (17) has applied technique of machine learning programming for modeling consumer behavior. The main benefit of the RF algorithm is its strength to manage efficiently and effectively several relevant features of the firm bankruptcy prediction. This is unlike logistic model that generally becomes unstable for data wherein share of failed firms is low. As an example, since bankruptcies are rare events described by many responsible covariates, RF reduces the set of variables which have the maximum reasonable importance to estimate the bankruptcy effectively.

Random forest procedure essentially evolves to obtain the class vote from each tree. The goal is to make non-overlapping hyper-cubes or regions viz., of predictors by splitting the predictor space. Subsequently, classification is carried out by majority vote. Simple averages are utilized as forecast from target point from each tree in case of regression. RF employs out-of-bag (OOB) sample for generating estimates. Bagging involves creating multiple copies of original training data set using bootstrap, fitting separate tree for each copy, and combining them to create a single predictive model. The estimator of the predicted value for a given x is given by:

$$\hat{y}(x) = \frac{1}{N} \sum_{n=1}^N \hat{y}_n(x)$$

We can employ an arbitrary number of trees in bagging so that the error variance stabilises beyond a critical number of trees. The estimates obtained by N-fold cross validation and OOB process are nearly similar. Due to this feature, RF can be applied in single iteration with multiple cross validations. This property differentiates RF from other non-linear iterative procedures. The training course may be stopped once the variation in the estimates as obtained by OOB is minimized to satisfactory level, calculated as below.

$$\frac{\sum_{n=1}^N (\hat{y}_n(x) - \hat{y}(x))^2}{N}$$

Here, MSE is Mean Square Errors and n is the number of observations in the validation set. RF is likely to perform poorly for cases wherein share of related parameters is less vis-à-vis total number of variables.

The features of selecting significant and important variable are one of the most exciting features in RF method. RF builds multiple decision trees and aggregates them together to derive stable estimates. RF analyses all the available variables to choose which one is the most important in hierarchical structure however Boosting ignores some variables completely to determine it. The split-variable tactics improves the likelihood of obtaining single variable selection through RF. Further, RF employs subsamples to derive importance measure for variables that intends to assess the forecasting prowess of varied parameters separately. As and when the tree is developed, its forecasting correctness is noted and the procedure is pursued for further subsamples. Subsequently, arbitrary parameter is drawn to assess the precision. The mean across all trees due to the decline in exactness of variable selection is calculated. It reflects the RF variable significance. The final value so obtained is depicted in chart as percentage with respect to the maximum score.

One of the relevant and most informative outputs of RF is a proximity plot. The proximity matrix is generated for the training information set through RF. Here, the proximity plot is augmented by unity for a pair of observations having same node. The two dimensional matrix of proximity incorporates multidimensional scaling. The essence of proximity plots is to provide an impression about closeness of observations that move jointly amongst large number of variables. The RF procedure depicts plots in terms of original values from the mixture information that enables improved comprehension about their characterization.

The RF algorithm has many benefits over logit regression and various additional common sorting methodologies in predicting the firm bankruptcy. Breiman (2001) advocated that RF is very sensitive to tweaking of values for sub sample size etc. Prediction based on default settings also produced reasonably good and tuning of parameters not required furthermore. Techniques like Support Vector Machines (SVM) require lot of simulation with parameter values for satisfactory outcome. Alternatively, RF classification also generates diagnostics that enable choosing the relevant variables. The individual coefficient estimates are not produced by RF method unlike classical regression approach. Nevertheless, it produces variable importance ranks that provide better forecast. The variable hierarchy list ranked the sufficiently significant variable in many cases is mostly calcified accurately. Mostly these are the same variables that work well in other models linear and logistic model. As compared to other classification methods RF also works fine for examining unbalanced information sets and it dominates according to classify the majority classes. However, Lee and Urrutia (18) has discussed the standard method of dealing with such data via sub-sampling of the majority class. The main gain in sub-sampling is simplifying the problem by reducing sample size to improve the precision of quantile estimation. The estimated values of quantiles that are near to either to 0 or 1 are not very accurate but sub-sampling improves its accuracy (19). The major disadvantages of the sub-sampling approach are that it decreases the predicative power. However, sub-sampling also is an important procedure inbuilt in the RF algorithm. The introduction of proper weights can also improve the accuracy of quantile estimates. The shrinkage and ridge regression (20, 21) and Lasso regression (22) or elastic net (23) could be much useful to overcome the insufficient sample size. The latter two approaches are also utilized in doing significant variable selection. Predominantly, the strategies centered around logistic regression share an underlying presumption of linear relation for non-bankrupt and bankrupt firms. The assumption of linear decision boundary can be relaxed to an extent by including interaction effects into the modeling exercise. There are always computational challenges present in the lasso and other types of shrinkage methods. But in RF algorithms a non-linear decision boundary can be constructed to deal with large number of predictor variables.

As regarding TN, it is one of the most contemporary and sophisticated research areas of machine learning for forecasting bankruptcy. An extensive exposition on the topic can be found in Ravi et al. (24) and Mukkamala et al. (25). In the Salford System's TN is the most efficient and dominant data mining tool, capable of constantly generating more accurate performing models. TN algorithm generates numerous tiny decision trees created in an ordered manner to correct the errors to congregate to an accurate model. Its robustness also includes data polluted with inaccurate target labels. This kind of data inaccuracy is mostly tricky to handle employing usual data mining tools and is inadequate for regular boosting. In contrast, TN is usually insulated to such imperfections. On the contrary, the degree of precision provided by TN is generally not realized by linear models or system like bagging or conventional boosting. TN has the advantage on ANN (Artificial Neural Network) by not being sensitive to erroneous data and requires minimal data preparation time, imputation of missing values, or pre-processing (Mukkamala et al., 2008). Ravi et al. (2008) discuss about a system with a multi-faceted statistical technique constituency to predict financial distress of banks. They adopted a novelty method to use Tree Net for feature selection (selecting the top five predictor variables), and then added them to the fuzzy rule based classifier. Their results yielded lower Type I and Type II errors vis-à-vis the constituent models in stand-alone mode. The feature of proximity plot is also available in tree net which ranks the important variable with highest predictive ability in order.

3. Data

The data is compiled from Capital IQ database that is repository of varied characteristics of Indian public limited companies collated from their Annual reports. A panel of 628 firms are included in our sample for a period encompassing 2006 to 2015 covering varied sectors, such as manufacturing, services, mining, and construction. As per the classification criteria based on company financials, 312 firms out of 668 firms are categorized as stressed. The remaining 356 firms are non-stressed. Eventually, the unbalanced data set comprises of 4539 observations.

Distress Grouping Criteria

The procedure employed for labelling of firms as distressed/non-distressed is elaborated hereunder, which is roughly as followed in per Shrivastava et al. (2018). The process that is employed avoids low proportion of distressed units and is shown to provide robust results. The grouping approach of firms into appropriate class is performed on the basis of various company financial parameters that is elucidated as below.

- (a) Interest coverage ratio (<1): It depicts a firm's capacity to fulfil various interest commitments. High figure for this ratio implies smooth business operations.
- (b) EBITDA (Gross earnings) to expense ratio (<1): A low ratio signals distress status wherein expenses are more than the earnings.
- (c) Leverage: A figure more than unity for networth to debt ratio is a reasonable situation, however lower value signifies greater debt burden thereby higher level of risk of failure.
- (d) Networth growth (negative for two consecutive years): A non-positive networth growth implies dented performance, which may lead to bankruptcy.

As per above, if at least three conditions are met, then the firm is marked as distress unit. Else, if a company clears all aforementioned criterions, then the company is labelled as healthy firm. Over period when a firm falls in distress category, it is considered to be its end and thereafter its data is not considered for analysis. Moreover, companies not classified in any of the groups are dropped off the dataset.

Sample Selection Procedure

To examine the performance of modelling strategies, the whole sample is bifurcated into two parts viz., training and testing. The training dataset is employed for model estimation whereas the testing sample is utilized to assess and compare the performance of modelling strategies. Around 20% of initial data with similar share of distress and non-distress firms comprise our testing data sample. The training set aggregates to 533 firms with 254 distress firms and 279 non-distress firms. Likewise, testing sample overall contains 135 firms with 77 non-distress and 58 distress firm. Both RF and TN are applied to evaluate and assess their respective classification and prediction abilities both. The classification ability is the accuracy with which a technique labels a firm as distress/non-distress in the training sample, also called as ‘in-sample’ category. Correspondingly, the degree of accuracy of a methodology with which a firm is marked as success/failure in testing sample states its prediction capability.

4. Empirical Analysis and discussions

Variable selection

A sizable number of firm level parameters reflecting characteristics like financial, age, performance, size etc. have been considered for the analysis. The complete list of such factors is shown in Table 1. Initially, the correlation matrix was tabulated to examine level of similarity amongst the variables in order to drop highly related variables. Subsequently, the proximity plots are derived that provide variable importance measure for both RF and TN to select the final variables in Table 2 and 3 respectively. Proximity plots gauges the average percentage change in predictive accuracy when the variable is included/excluded in model. The proximity plot shows that parameters like EBIT margin, profit margin, retained earnings to asset ratio, debt to assets and others are vital indicators for prediction¹. Accordingly, a set of eight firm variables each for RF and TN are chosen for further analysis.

Table 1: List of variables

Variable	Description
EBIT_TAXPAID	Earnings before interest and tax/Tax paid
PROFIT_MARGIN	Net income/Net sales
RETA_EARNING_ASSET	Retained earnings/Assets
DBT_AST	Debt/Assets
REC_CUR_AST	Trade receivable/Current assets
CASH_CURRENT_LIB	Cash/Current liability
WRK_CP_AST	Working capital/Assets
LN_NW	Natural log(Net worth)
INVENTORY_SALES\$	Inventory/Sales
LOG_INVENTORY\$	Natural log(Inventory)
LOG_ASST\$	Natural log(Assets)
LOG_INVESTMENT\$	Natural log(Inventory)
LOG_AGE\$	Natural log(Time since incorporation)
LOG_PUC\$	Natural log(Paid up capital)

Note: \$ indicates the variable dropped due to low ranking as per variable import

Table 2: Proximity plot – Random Forest

¹ Variables with a score of more than 10 are included for further analysis.

Variable	Score	Plot
EBIT_TAXPAID	100.00	
DBT_AST	95.02	
PROFIT_MARGIN	76.90	
WRK_CP_AST	39.25	
RETA_EARNING_ASSET	26.88	
CASH_CURRENT_LIB	11.28	
REC_CUR_AST	10.68	
LN_NW	10.14	
INVENTORY_SALES	8.89	
LOG_AGE	8.83	
LOG_ASST	5.68	
LOG_INVESTMENT	3.19	
LOG_INVENTORY	1.02	
LOG_PUC	0.61	

Table 3: Proximity plot - Tree-Net

Variable	Score	Plot
CASH_CURRENT_LIB	100.00	
EBIT_TAXPAID	74.49	
LN_NW	46.15	
REC_CUR_AST	32.70	
RETA_EARNING_ASSET	18.64	
WRK_CP_AST	10.88	
PROFIT_MARGIN	10.85	
DBT_AST	10.22	
LOG_AGE	6.55	
LOG_ASST	6.19	
INVENTORY_SALES	4.95	
LOG_INVENTORY	4.45	
LOG_PUC	3.20	
LOG_INVESTMENT	1.87	

The snapshot of selected indicators included for modelling procedure both for the training and entire sample is tabulated in Table 4 and 5 respectively. At a glance, the mean and standard deviation of all parameters is observed to be broadly similar for both the samples. A closer look reveals lower profitability figures for distressed companies indicating distress build-up. Borrowing pressure is also reflected through leverage that is higher for stressed firms. Likewise, working capital to assets ratio is very low, whereas receivable to current ratio is higher for strained firm. In order to attest to the differentiating ability of the above mentioned techniques to accurately group the firms as distressed or failure, relevant statistical test has also been carried out. The results showed that the figures are significantly different between the two groups vindicating the classification methodology. However, tables are not presented here as per space concern.

To test the performance of RF method, we have used the above mentioned dataset based on Indian companies. Chart 1 enables to assess the number of trees in the RF methods to reduce the error. It clearly shows that after the generation of 30 trees, the method stabilizes with the lowest balance error rate.

Table 4: Variable Summary—Training Sample

Variable	Group	Average	SD
EBIT_TAXPAID		10.71	35.14
	1	-57.39	198.76
PROFIT_MARGIN		0.15	0.12
	1	-0.16	0.29
RETA_EARNING_ASSET		0.06	0.05
	1	-0.06	0.12
DBT_AST		0.13	0.13
	1	0.20	0.19
REC_CUR_AST		0.37	0.26
	1	0.32	0.18
CASH_CURRENT_LIB		0.26	0.59
	1	0.13	0.35
WRK_CP_AST		0.13	0.21
	1	0.01	0.35
LN_NW		6.05	0.70
	1	5.61	0.65

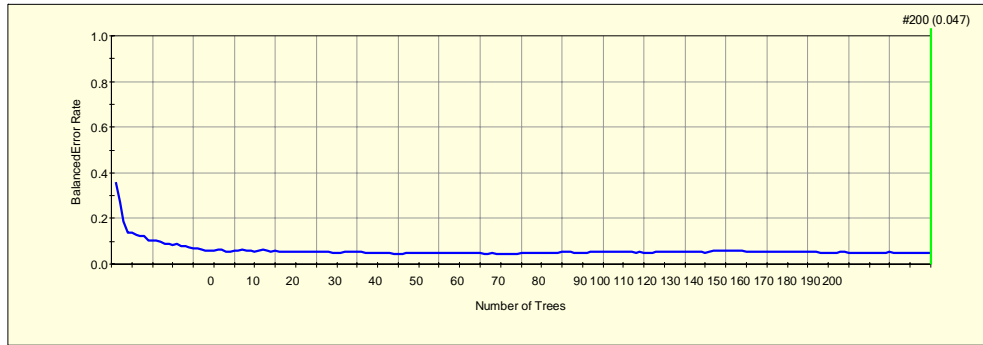
0 stands for healthy whereas 1 stressed firms.

Table 5: Variable Summary—Entire Sample

Variable	Category	Mean	SD
EBIT_TAXPAID		10.89	35.24
	1	-65.34	227.36
PROFIT_MARGIN		0.15	0.11
	1	-0.16	0.29
RETA_EARNING_ASSET		0.06	0.05
	1	-0.07	0.12
DEBT_ASST		0.13	0.13
	1	0.20	0.21
RECEIVABLE_CURR_ASST		0.37	0.24
	1	0.31	0.20
CASH_CURRENT_LIB		0.25	0.53
	1	0.11	0.31
WRK_CP_AST		0.13	0.21
	1	-0.03	0.41
LN_NW		6.07	0.73
	1	5.57	0.65

0 stands for non-distressed whereas 1 distressed firms.

Chart 1: Error rate vis-à-vis number of trees



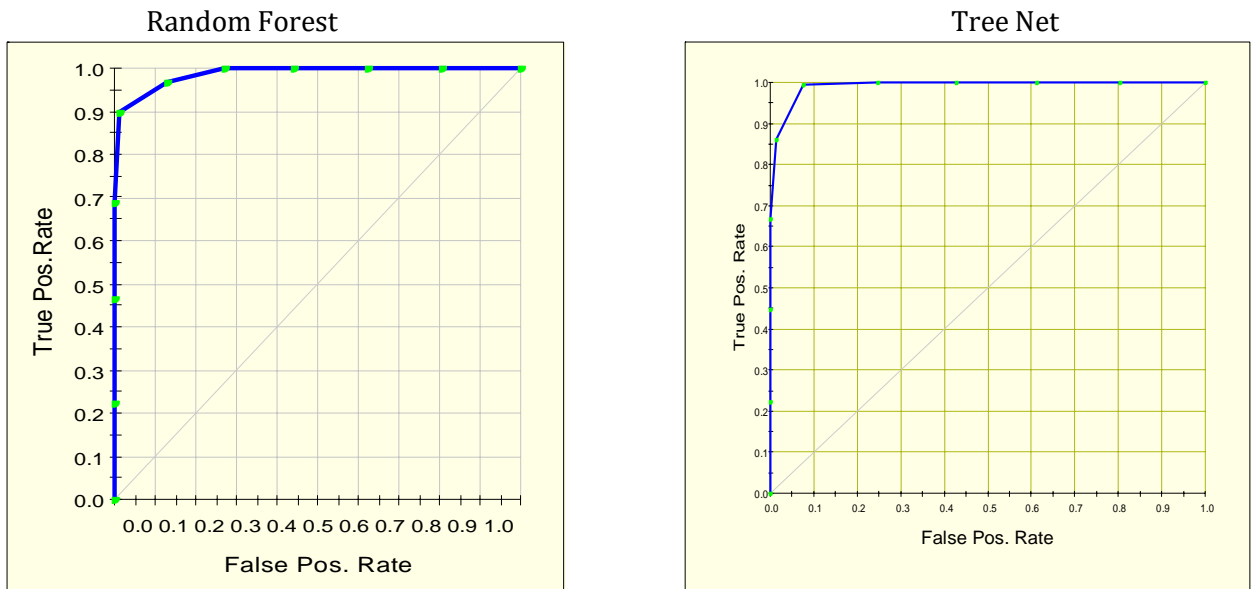
Model Evaluation

This section compares both in-sample classification and prediction accuracy assessment. Table 6 contrasts the classification ability RF vis-à-vis TN. It portrays that in-sample classification estimates of TN surpasses that of RF by approximately 3.0%.

The Chart 2 depicts the ROC curve for both RF and TN that shows higher area under curve in case of TN. The area in case of TN is 98.1% as compared to 96.4%, which implies that TN provides better estimates by nearly 1.7% in contrast to RF.

The forecasting accuracy has also been assessed for RF and TN separately. On the basis of training sample, N period ahead forecasting has been performed. The results are summarized in Chart 3. It is observed that although Type I error of both RF and TN increases as we move farther ahead. However, TN has provided consistently better results in comparison to RF that re-verifies the superiority of TN.

Chart 2: Receiver operating characteristic (ROC) curve



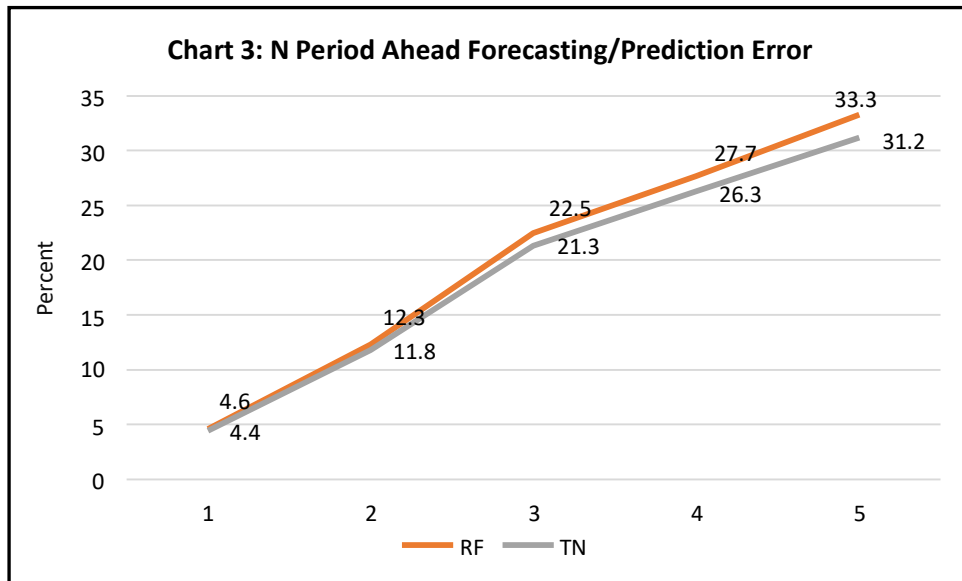


Table 6: Classification Matrix

Random Forest-Training Sample					Tree-Net-Training Sample				
Actual Class	Total Class	Percent Correct	Predicted Classes		Actual Class	Total Class	Percent Correct	Predicted Classes	
			Non-distressed	Distressed				Non-distressed	Distressed
Non-distressed	279	94.98%	265	14	Non-distressed	279	97.13%	271	8
Distressed	254	95.67%	9	243	Distressed	254	99.61%	1	253
Total:	533				Total:	533			
Overall % Correct:		95.31%			Overall % Correct:		98.31%		

Random Forest-Testing Sample					Tree-Net-Testing Sample				
Actual Class	Total Class	Percent Correct	Predicted Classes		Actual Class	Total Class	Percent Correct	Predicted Classes	
			Non-distressed	Distressed				Non-distressed	Distressed
Non-distressed	77	92.21%	71	6	Non-distressed	77	97.40%	75	2
Distressed	58	94.83%	3	55	Distressed	58	96.55%	2	56
Total:	135				Total:	135			
Overall % Correct:		93.33%			Overall % Correct:		97.04%		

5. Conclusion

The Random Forest classification is a popular machine learning algorithm to predict bankruptcy (distress) of firms or business that has been considered in this study and evaluated against Tree Net algorithm that is also an extensively applied machine learning algorithm not only in bankruptcy prediction

but also in Information Technology and other fields. Random Forest orders firms according to their propensity to default or to become distressed.

The relative superiority of the different approaches has been verified in this study employing exhaustive information set from corporate India. The firms covering varied sectors like manufacturing, services etc. from 2006 to 2015 have been chosen for the purpose. On comparison amongst the two, it is observed that the Tree Net methodology has been producing better 'in-sample' classification accuracy in contrast to Random Forest methodology translating in estimation gain of around 3%. Furthermore, Tree Net is showing superior predictive performance in contrast to the Random Forest consistently. The analysis provides useful insights for possible tools that may be used by management, regulators and researchers alike to forecast and ascertain the financial health of firms.

Statement

The views expressed in this paper are strictly personal and not related to authors' respective institutional affiliation. All the errors, omissions etc., if any, are the sole responsibility of authors. The authors may be contacted at arvind1aug@gmail.com, nitin_005us@yahoo.com, kkumar@bond.edu.au and sanjeev2678@gmail.com respectively.

References

1. J. Tirole, *The theory of industrial organization*, MIT press, (1988).
2. A. Demirgüç-Kunt, and R. Levine, *Financial structure and economic growth: A cross-country comparison of banks, markets, and development*. MIT press, (2004).
3. E.I. Altman, *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy*, *The journal of finance*, 23 (1968) 589-609.
4. A. Shrivastava, K. Kumar, and N. Kumar, *Business Distress Prediction Using Bayesian Logistic Model for Indian Firms*, *Risks*, 6 (2018) 113.
5. Z. Hua, Y. Wang, X. Xu, B. Zhang, and L. Liang, *Predicting corporate financial distress based on integration of support vector machine and logistic regression*, *Expert Systems with Applications*, 33 (2007) 434-440.
6. E.I. Altman, G. Marco, F. Varetto, *Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)*, *Journal of banking & finance*, 18 (1994) 505-529.
7. J.A. Ohlson, *Financial ratios and the probabilistic prediction of bankruptcy*, *Journal of accounting research*, 18 (1980) 109-131.
8. Z. Li, J. Crook, and G. Andreeva, *Chinese companies distress prediction: an application of data envelopment analysis*, *Journal of the Operational Research Society*, 65 (2014) 466-479.
9. V. Bapat, and A. Nagale, *Comparison of bankruptcy prediction models: evidence from India*, *Accounting and Finance Research*, 3 (2014) 91-98.
10. K. Halteh, K. Kumar, and A. Gepp, *Using Cutting-Edge Tree-Based Stochastic Models to Predict Credit Risk*, *Risks*, 6 (2018) 55.
11. E. Bauer, R. Kohavi, *An empirical comparison of voting classification algorithms: Bagging, boosting, and variants*, *Machine learning*, 36 (1999) 105-139.
12. T.G. Dietterich, *An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization*, *Machine learning*, 40 (2000) 139-157.
13. Breiman, L. *Statistical modeling: The two cultures (with comments and a rejoinder by the author)*, *Statistical science*, 16 (2001) 199-231.
14. L. Breiman, *Some infinity theory for predictor ensembles*, Technical Report 579, Statistics Dept. UCB (2000).

15. T.K. Ho, The random subspace method for constructing decision forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (1998) 832-844.
16. Y. Amit, and D. Geman, Randomized Inquiries About Shape: An Application to Handwritten Digit Recognition (No. TR-401), Department of Statistics, University of Chicago (1994).
17. A.E. Khandani, A.J. Kim, and A.W. Lo, Consumer credit-risk models via machine-learning algorithms, *Journal of Banking & Finance*, 34 (2010) 2767-2787.
18. S.H. Lee, and J.L. Urrutia, Analysis and prediction of insolvency in the property-liability insurance industry: A comparison of logit and hazard models, *Journal of Risk and Insurance*, 63 (1996) 121-130.
19. D. Mease, A.J. Wyner, and A. Buja, Boosted classification trees and class probability/quantile estimation, *Journal of Machine Learning Research*, 8 (2007) 409-439.
20. A.E. Hoerl and R.W. Kennard, Ridge regression: applications to non-orthogonal problems, *Technometrics*, 12 (1970) 69-82.
21. R. Hastie, Problems for judgment and decision making, *Annual review of psychology*, 52 (2001) 653-683.
22. R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological)*, 58 (1996) 267-288.
23. H. Zou, and T. Hastie, Regularization and variable selection via the elastic net, *Journal of the royal statistical society: series B (statistical methodology)*, 67 (2005) 301-320.
24. V. Ravi, P.R. Kumar, E.R. Srinivas, and N.K. Kasabov, A semi-online training algorithm for the radial basis function neural networks: applications to bankruptcy prediction in banks, In: V. Ravi (Ed.) *Advances in Banking Technology and Management: Impacts of ICT and CRM*, (2008) 243-260.
25. S. Mukkamala, A. Vieira, and A. Sung, Model selection and feature ranking for financial distress classification, In *Proceedings of the 8th International Conference on Enterprise Information Systems INSTICC–Institute for Systems and Technologies of Information, Control and Communication* (2008).