THE CAPCO INSTITUTE
JOURNAL
OF FINANCIAL TRANSFORMATION

TECHNOLOGICAL

Performance of using machine learning approaches for credit rating prediction: Random forest and boosting algorithms

W. PAUL CHIOU | YUCHEN DONG | SOFIA X. MA

ARTIFICIAL INTELLIGENCE

#58 NOVEMBER 2023

THE CAPCO INSTITUTE

JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

CONTENTS

TECHNOLOGICAL

OPERATIONAL

ORGANIZATIONAL

DEAR READER,

As the financial services industry continues to embrace transformation, advanced artificial intelligence models are already being utilized to drive superior customer experience, provide high-speed data analysis that generates meaningful insights, and to improve efficiency and cost-effectiveness.

Generative AI has made a significant early impact on the financial sector, and there is much more to come. The highly regulated nature of our industry, and the importance of data management mean that the huge potential of AI must be harnessed effectively – and safely. Solutions will need to address existing pain points – from knowledge management to software development and regulatory compliance – while also ensuring institutions can experiment and learn from GenAI.

This edition of the Capco Journal of Financial Transformation examines practical applications of AI across our industry, including banking and fintechs, asset management, investment advice, credit rating, software development and financial ecosystems. Contributions to this edition come from engineers, researchers, scientists, and business executives working at the leading edge of AI, as well as the subject matter experts here at Capco, who are developing innovative AI-powered solutions for our clients.

To realize the full benefits of artificial intelligence, business leaders need to have a robust AI governance model in place, that meets the needs of their organizations while mitigating the risks of new technology to trust, accuracy, fairness, inclusivity, and intellectual property. A new generation of software developers who place AI at the heart of their approach is also emerging. Both GenAI governance and these 'Developers 3.0' are examined in this edition.

This year Capco is celebrating its 25th anniversary, and our mission remains as clear today as a quarter century ago: to simplify complexity for our clients, leveraging disruptive thinking to deliver lasting change for our clients and their customers. By showcasing the very best industry expertise, independent thinking and strategic insight, our Journal is our commitment to bold transformation and looking beyond the status quo. I hope you find the latest edition to be timely and informative.

Thank you to all our contributors and readers.

Lance Levy, **Capco CEO**

PERFORMANCE OF USING MACHINE LEARNING APPROACHES FOR CREDIT RATING PREDICTION: RANDOM FOREST AND BOOSTING ALGORITHMS

W. PAUL CHIOU | Associate Teaching Professor of Finance, Northeastern University
YUCHEN DONG | Senior Engineer, MathWorks
SOFIA X. MA | Senior Engineer, MathWorks

ABSTRACT

Applying machine learning techniques to improve the accuracy and efficiency of predictions of credit risk rating is increasingly critical to the financial services industry. In this study, we apply MATLAB to investigate the performance of two approaches, decision forest and boosting algorithms, by using a wide range of financial data. The empirical outcomes suggest that both methods exhibit considerable performance but may be superior to each other in different scenarios. Boosting algorithms method exhibits an accuracy rate of approximately 67% across the credit rating categories. The random forests model generates lower accuracy rates for low and medium classifications than the boosting method, but the accuracy rate for high credit ratings reaches 79%, more accurate than the boosting method.

1. INTRODUCTION

Credit rating prediction is a critical task in the financial services industry, as the outcomes can affect investment decisions, corporate finance, and risk management. An accurate forecast of default risk provides an early warning system for identifying entities or investments that may pose financial, operational, or strategic risks. With recent systematic shocks, such as the COVID-19 pandemic, risk assessment and compliance required by regulators make predicting credit rating essential to avoid legal repercussions. As financial institutions use risk ratings to make lending decisions and to determine interest rates, accurate risk ratings can be critical to managing corporate finance. However, credit risk is modeled under assumptions of trackable borrower and market dynamics and does not account for unforeseen events, hence, leaving the models unable to produce reliable results.

Machine learning techniques have been widely applied for their potential to improve the accuracy and efficiency of predictions. However, several challenges – such as how to process data to ensure the quality for analysis, imbalanced numbers of defaults compared to non-defaults, identifying the relevant features, model validation to ensure accuracy and robustness, and selecting the appropriate algorithm – still need to be addressed in applying machine learning to forecast credit risk ratings.

This research contributes to the literature threefold. First, the levels of accuracy and interpretability of credit risk predictions may vary across different algorithms; consequently, identification of the appropriate approaches remains uncertain. Second, validation processes that consider performance metrics and testing scenarios can be complex and are relevant to the usefulness of the selected models. Third, identifying the relevant credit-related factors that can help minimize the impacts of feature overload, the curse of dimensionality, multicollinearity, and noise in data is crucial to risk management practice.

This study applies two methods, decision tree boosting and bagging algorithms, and highlights their relative strengths and applicability to credit rating prediction using a wide range of financial data as input features. Machine learning techniques have advanced credit risk assessment but can be highly complex, leaving implementation and interpretation of the outcomes challenging. We, particularly, apply the machine learning tools in MATLAB, specifically, bagged random forests and boosting algorithms, which are used in credit rating predictions. The findings suggest that both methods exhibit considerable performance but may be superior to one another in different scenarios. Boosting algorithms method exhibits accuracy rates of approximately 67% across the credit rating categories. The random forest model generates lower accuracy rates for low and medium classifications than the boosting method, but the accuracy rate for high credit ratings reaches 79%, more accurate than the outcome using the boosting method.

2. MODELS

2.1 Random forests

Random forests, or decision forests, that assemble a collection of decision trees working conjointly in predictions and classifications belong to a family of supervised machine learning models and algorithms. The methods offer numerous advantages, such as ease of configuration, native handling of diverse features, robustness to noisy data, and interpretability. Due to properties such as interpretability, scalability, resistance to overfitting, and handling missing data, decision forests are suitable for signal integration from tabular data, allowing for efficient aggregation of signals from multiple subsystems.

Random forests serve as a remedy for the overfitting issues related to the tree learning approach, such as low bias but high variance, resulting in decreased accuracy. Working on decision forests involves creating and training multiple decision trees with random subsets of data and features. By combining multiple deep decision trees trained on different subsets of the training data, this approach reduces variance while introducing a slight increase in bias and some loss of interpretability. The teamwork of many trees in a forest effectively enhances the overall performance, as compared to a single random tree, yielding more accurate and robust results for data mining tasks.

To build decision forests, the bagging method, repeatedly selecting random samples with replacements from the training set, is utilized in the training algorithm. Classification or regression trees are trained using these subsamples, and they predict unseen samples by either averaging the predictions from all individual regression trees or taking the majority votes from classification trees. The above procedure improves model performance by reducing variance while controlling the increase in bias. By creating fewer correlated trees through different training sets, bagging ensures that the average predictions of multiple trees are less sensitive to noise, as compared to a single tree. The process involves selecting a random subset of features at each candidate split, mitigating the issue of strong predictors dominating multiple trees and causing correlation.

Random forests are trained as a system that has few hyperparameters that can be of proper default values. This allows them to be more efficient in data preprocessing while reducing error sources and enhancing the accuracy of the results. A group of decision trees utilizes a random subset of features and data points from the training set, allowing numeric features to be natively handled. This enables the generation of robust results from highly stochastic data. Thus, the results of decision forests can be easily interpretable and understood.

As a supervised learning model, bagging is usually used to reduce the variance of the decision trees by averaging the prediction over a collection of bootstrap samples. Specifically, bagging is to create several subsets of data from the training samples chosen randomly with replacement. As a result, the prediction from sampled data will be more robust than using only one single decision tree. Suppose the training data is $\mathbf{Z} = \{(x_1, y_1), \ldots (x_N, y_N)\}$. Our goal is to determine the prediction $\hat{f}(x)$ given the data x. Denote the bootstrap samples as $\mathbf{Z}^{b}$, where $b = 1, 2, \ldots, B$. Here, B is the number of bootstrap sampled dataset. The bagging estimate is defined as

$$f_{\text{bag}}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{*b}(x) \quad (1)$$

Considering equation (1) as the Monte Carlo estimation, it converges to the true estimation as B goes to infinity. Let us consider the regression tree model as an example. Let $\hat{f}^{(x)}$ be the tree's prediction given the input data x. The trained tree model from bootstrapping samples typically involves different features than the original one. It might have a different number of terminal nodes as well. The bagged estimate $\hat{f}_{\text{bag}}(x)$ is the average prediction from B different trees.

In addition to the regression tree, the classification tree model is also popular and used in the following example. Suppose the target of the classification model takes value in 1, 2, ..., K, and we have m nodes and region R_m with N_m observations. We further define the proportion of class k observations in node m as

$$\hat{p}_{m,k} = \frac{1}{N_m} \sum_{x_i \in R_m} 1_{\{y_i=k\}} \qquad (2)$$

We classify the observations in node m to class $k(m) = argmax_K \, \hat{p}_{m,k}$, the majority class in node m. Considering the cost complexity criterion, we can select different measures of node impurity, such as misclassification error, Gini index, or cross-entropy based on different situations.

2.2 Boosting algorithms

Boosting algorithms is an ensemble learning technique that combines multiple weak decision tree models to enhance predictiveness through iterative learning steps. It is an extension of the boosting method, which is a general approach for improving the performance of a base learning algorithm by combining several weak learners in a weighted manner.

The boosting algorithm offers an effective solution for prediction tasks in both classification and regression fields. By iteratively adding decision trees to the ensemble, the framework allows each new tree to be trained to correct the errors of its predecessor. The trees are added sequentially, with each tree learning to fit the residual errors from the previous trees. In each iteration, coefficients, weights, or biases of input variables are adjusted to minimize the loss function, measuring the discrepancy between predicted and actual target values. The gradient represents incremental adjustments in each step, while boosting accelerates predictive accuracy improvements, reaching an optimal level. The final model is the weighted sum of all individual trees. By streamlining the objective and minimizing iterations, this method enhances the learning process, achieving a satisfactory optimal solution more efficiently.

The machine learning tools in MATLAB 2023a, such as bagged decision trees, are used in the domain of credit rating prediction in this study. The flexibility of the software, such as Deep Learning Toolbox and Database Toolbox, enables researchers to tailor and adapt the workflow delineated to their unique preferences and specific requirements.

3. DATA, PROCESS, AND PREPARATION FOR ANALYSIS

Financial ratios as predictors are used to forecast the credit rating as the response variable by fitting a bagged decision tree. Bagging, or bootstrap aggregation, consists of generating many random sub-samples, or bootstrap replicas from the dataset by sampling with replacement from the list of customers in the dataset. A decision tree grows from the replica. Each decision tree is a trained classifier on its own and could be used in isolation to classify new clients. The predictions of two trees grown from two different bootstrap replicas may be different. The ensemble aggregates the predictions of all the decision trees that are grown for all the bootstrap replicas.

Table 1: Variables used to predict credit rating

RATIO	MEDIAN	RATIO	MEDIAN
Current ratio	1.49	Debt equity ratio	1.65
Quick ratio	0.99	Debt ratio	0.64
Cash ratio	0.30	Effective tax rate	0.30
Days of sales outstanding	42.37	Free cash flow operating cash flow ratio	0.64
Net profit margin	0.06	Free cash flow per share	2.13
Pretax profit margin	0.08	Cash per share	3.69
Gross profit margin	0.41	Company equity multiplier	2.65
Operating profit margin	0.11	EBIT per revenue	0.09
Return on assets	0.05	Enterprise value multiple	9.27
Return on capital employed	0.07	Operating cash flow per share	4.35
Return on equity	0.12	Operating cash flow sales ratio	0.13
Asset turnover	0.70	Payables turnover	5.76
Fixed asset turnover	3.81		

If the majority of the trees predict one particular class for a new customer, it is reasonable to consider that prediction to be more robust than the prediction of any single tree alone. The information is still useful when a different class is predicted by a smaller set of trees. The proportion of trees that predict different classes is the basis for the classification scores that are reported by the ensemble when classifying new data.

3.1. Data

The quarterly data of 2,029 credit ratings between 2010 and 2016 are used in this study. Table 1 presents the financial ratios applied in this study to forecast credit rating and their medians. These widely applied measures collectively provide insights into a company's liquidity, profitability, asset management efficiency, and financial leverage, and are essential for assessing financial health, and risk profile, widely regarded as possible factors for predicting credit ratings, such as Altman's z-score (1968). First, it is natural to consider financial leverage that involves the use of debt to finance a company's operations. The debt/equity ratio of 1.65 and debt ratio of 0.64 represent, overall, the reliance on debt of companies in their capital structures. A higher debt/equity ratio indicates a relatively higher level of debt compared to equity, while the debt ratio illustrates the proportion of total assets financed by debt.

To measure the ability to meet short-term obligations, current ratio, quick ratio, and cash ratio provide insights into a company's liquidity position. Several ratios are used to measure profitability from different aspects: net profit margin, pretax profit margin, gross profit margin, operating profit margin, return on assets, return on capital employed, and return on equity. The fact that the company retained, on average, $0.06 in profit for every $1 in net sales revenue over the sample period, similar to the historical averages, validates the use of the data for credit risk analysis. In addition, utilization of assets is also considered in the analysis as operational performance can be critical to risk management. The "days of sales outstanding" suggests that the companies, overall, take an average of 42.37 days to convert sales into cash receipts. Asset turnover and fixed asset turnover values of 0.70 and 3.81, respectively, indicate the operating efficiency of the overall assets and fixed assets in generating revenue.

The distribution of sectors in this study includes a range of industries. The largest sectors by number of observations are energy, consumer services, public utilities, technology, and basic industries, together representing more than 60% of the sample analyzed. Other than the sectors above, the majority of companies included are from manufacturing industries. On the other hand, the study only includes 50 observations from the financial services sector. As the sample of this study reflects a comprehensive analysis of non-financial service industries, the financial ratios applied will be meaningful to determine the credit risk.

For the distribution of credit ratings, the majority falls within the investment-grade categories of BBB (671) and A (398), while higher credit ratings, such as AAA (7) and AA (89), are less common. Riskier credit ratings, such as CCC-rated or lower, comprise smaller portions of the dataset, representing less than 4% of the sample. The distribution of raters in the study shows that Standard & Poor's has the highest representation, followed by Moody's, and Egan-Jones, representing about 95% of the samples. Other rating agencies, like Fitch and DBRS, have comparatively fewer observations.

Table 2: The distribution of companies and rating agencies

SECTOR	N	RATING	N	RATING AGENCY	N
Basic industries	260	AAA	7	DBRS	3
Capital goods	234	AA	89	Egan-Jones	603
Consumer durables	73	A	398	Fitch	100
Consumer non-durables	132	BBB	671	Moody's	579
Consumer services	250	BB	490	Standard & Poor's	744
Energy	294	B	302		
Finance	50	CCC	64		
Healthcare	171	CC	5		
Public utilities	211	C	2		
Technology	234	D	1		
Transportation	63				
Miscellaneous	57				

3.2. Characteristics across various ratings

Table 3 presents a summary of various financial indicators across different credit ratings, providing a concise overview of key financial and operational indicators across different credit ratings and the first look at the relationships between creditworthiness and various performance metrics. The financial ratios in the empirical analysis include various measures of liquidity, profitability, asset management, and financial leverage. Companies with higher credit ratings tend to exhibit more favorable financial metrics. As shown in Panel A, higher credit ratings, such as AAA and AA, are associated with stronger liquidity, as seen in their higher current ratios and quick ratios compared to lower-rated categories. For instance, companies rated AAA showcase a high current ratio (CR) of 2.50, indicating a strong ability to cover short-term liabilities. Days of sales outstanding tend to decrease as credit ratings improve, indicating better management of receivables. Operating cash flow, generally, increases with higher credit ratings, except for a dip in the CCC category. Free cash flow per share and cash per share also tend to be more favorable in higher credit rating tiers, with the AAA-rated showing the strongest positions and the D-rated exhibiting the weakest.

Table 3: Summary statistics of some variables

PANEL A: LIQUIDITY										
FINANCIAL RATIO/RATING	AAA	AA	A	BBB	BB	B	CCC	CC	C	D
Current ratio	2.50	1.47	1.34	1.43	1.67	1.62	1.68	1.34	1.52	0.59
Quick ratio	2.30	0.97	0.85	0.93	1.14	1.06	1.17	0.57	0.55	0.39
Cash ratio	0.19	0.34	0.25	0.28	0.37	0.29	0.37	0.30	0.37	0.02
Days of sales outstanding	78.13	39.19	39.96	42.37	43.88	43.98	41.72	21.11	14.20	54.74
Operating cash flow per share	3.88	8.31	5.57	5.07	3.58	2.14	1.44	-1.51	4.54	3.22
Operating cash flow sales	0.37	0.18	0.15	0.15	0.12	0.10	0.06	0.00	0.03	0.08
Free cash flow operating cash flow	0.80	0.71	0.69	0.60	0.66	0.63	0.78	1.00	-0.33	0.42
Free cash flow per share ($)	3.22	3.61	3.20	2.42	1.65	0.75	0.01	1.75	1.94	1.36
Cash per share ($)	10.33	6.07	3.91	3.65	3.78	2.26	3.79	7.75	12.19	1.26
PANEL B: PROFITABILITY										
Net profit margin	0.20	0.11	0.09	0.07	0.05	0.02	-0.03	-0.23	-0.30	0.18
Pretax profit margin	0.08	0.15	0.12	0.09	0.06	0.02	-0.04	-0.10	0.03	0.13
Gross profit margin	0.69	0.59	0.49	0.38	0.35	0.41	0.76	0.92	0.11	1.00
Operating profit margin	0.30	0.15	0.14	0.12	0.08	0.06	0.02	-0.09	-0.01	-0.20
Return on assets	0.10	0.09	0.07	0.05	0.03	0.02	-0.02	-0.33	-0.34	0.04
Return on capital employed	0.06	0.17	0.13	0.08	0.05	0.03	-0.01	-0.61	-0.02	0.04
Return on equity	0.23	0.18	0.17	0.13	0.09	0.06	-0.01	0.91	0.64	-0.50
EBIT per revenue	0.08	0.15	0.13	0.10	0.06	0.03	-0.04	-0.10	0.03	0.13
PANEL C: ASSET MANAGEMENT										
Asset turnover	0.53	0.80	0.71	0.70	0.71	0.65	0.41	0.51	1.18	0.24
Fixed asset turnover	6.35	3.96	4.48	3.54	4.11	3.38	2.14	3.08	4.85	0.26
Effective tax rate	0.28	0.29	0.31	0.31	0.30	0.26	0.17	0.00	-0.48	-0.39
Payables turnover	4.19	5.48	4.99	5.76	6.46	6.08	1.84	1.12	11.82	3.65
PANEL D: FINANCIAL LEVERAGE										
Debt equity ratio	0.92	1.37	1.59	1.67	1.6	2.09	2.39	-3.46	-2.91	-12.41
Debt ratio	0.48	0.58	0.62	0.63	0.64	0.74	0.79	1.21	1.65	1.09
Company equity multiplier	1.92	2.28	2.56	2.68	2.6	3.16	3.46	-2.75	-1.91	-11.41
Enterprise value multiple	11.33	9.79	9.82	9.2	8.59	9.46	8.93	-1.69	1.68	20.82

The findings in Panel B suggest a correlation between the financial health and profitability of companies. Higher credit-rated companies, such as AAA and AA, exhibit stronger profitability metrics, including higher net profit margin, pretax profit margin, gross profit margin, and return on equity. In contrast, lower credit-rated categories, like B, CCC, and below, show weaker profitability indicators, with negative values observed for net profit margin, pretax profit margin, and operating profit margin in some cases. Return on assets and return on capital employed tend to decrease as credit ratings decline, while return on equity displays a mixed trend. Distinct trends of asset management efficiency across different credit ratings are evident in Panel C. The companies with higher credit ratings display more efficient asset utilization, evident from higher asset turnover and fixed asset turnover ratios, while maintaining a moderate effective tax rate. Conversely, companies in lower credit-rated categories, particularly C- and D-rated, exhibit varying asset management efficiency, with lower asset turnover and fixed asset turnover. The payables turnover ratio is less consistent but generally tends to be higher for higher credit-rated categories.

The financial leverage across different credit ratings reveals several patterns. In Panel D, companies with higher credit ratings, like AAA and AA, tend to have lower debt equity ratios, debt ratios, and company equity multipliers, indicating more conservative financial structures. On the other hand, the lower-rated exhibits higher financial leverage. Although the enterprise value multiple varies inconsistently across credit ratings, it appears to be lower for the higher-rated, confirming that higher credit-rated companies tend to have lower financial leverage structures.

From the first glance at the data, it becomes evident that the liability burden increases across lower credit ratings, suggesting a higher proportion of debt in their capital structures. Companies with low ratings exhibit poor profitability and inefficient operations, indicating a possibility of financial distress. Moreover, the cash per share (CPS) varies across the credit rating spectrum, reflecting the liquidity position of each category. Overall, the trends suggest a correlation between credit rating and financial health, underscoring the implicit information of these metrics in assessing a company's risk profile and stability. Thus, it makes sense to apply these indicators to determine the financial health and operational efficiency of companies within each rating tier.

3.3. Data preprocessing

The process of data cleaning and selection in the context of machine learning is critical to ensure the effectiveness and reliability of the resulting models. Since not all available data may exhibit large and representative characteristics, data preprocessing stands out as a pivotal stage in the machine learning algorithm. After observing the data, we find that certain credit ratings, such as AAA and D, have insufficient representation. For instance, AAA rating only has seven data points and rating D only has one data point. Thus, data consolidation into fewer categories is essential to address this issue. After the rating is reorganized as demonstrated in Table 4, the risk data is representative, as each classification is large. We further let 80% of data, or 1,623 observations, be tested while the others are used for training.

Table 4: Reorganized credit rating

	CREDIT VALUE	#	%
1	Low	494	24.35
2	Medium	671	33.07
3	High	864	42.58

3.4. Classifying new data classification

The previously constructed classification can be used for the assignment of credit ratings to new companies. Since the ratings for existing companies also require reviews that account for variations in their financials, the dataset includes a list of such customers. To predict the credit ratings for the new data, the classifier's "predict" method is invoked. This method yields two essential outputs: the predicted class and the associated classification score. Both output arguments are acquired as the classification scores furnish valuable insights into the confidence level associated with the predicted ratings. Certain advanced computational software, such as MATLAB, facilitates an easily applied tool for the report generation of the classification process.

Preserving records of predicted ratings and their corresponding scores can be used to prove the benefits of periodic assessments of classifier performance. This information can be efficiently stored within a table and further archived through means such as saving to a comma-delimited text file or direct integration into a database system.

3.5. Back-testing: profiling the classification process

The assessment of model performance and the validation of credit ratings are applied in this research. The first measurement centers on the accuracy of predicted ratings relative to actual ratings. Predicted ratings are derived from automated classification processes, while actual ratings are assigned by a credit committee amalgamating various information sources. The second measurement evaluates the accuracy of actual ratings retrospectively. Specifically, it examines whether the actual ratings effectively mirror the credit risk of customers. The ex-post analysis, generally conducted over a one-year horizon, identifies companies that experienced defaults during the period, assessing actual rating accuracy.

The research leverages ex-post credit rating data, encompassing subsequent developments for the same companies considered previously. This dataset includes ratings assigned by the credit committee, along with a default flag indicating whether a given company defaulted within a year of the rating process.

Enhanced accuracy of predicted ratings translates into enhanced efficiency in reviewing these predictions. Consequently, it is plausible that the credit risk committee seeks periodic evaluations to gauge alignment between predicted and final ratings, potentially recommending re-training of the automated classifier in case of significant disparities. To facilitate the comparison of predicted versus actual ratings, a confusion matrix is employed. The matrix can be normalized by standardizing values as percentages by dividing the number of observations with true ratings.

4. EMPIRICAL RESULTS

4.1. Boosting algorithms method

We employ the AdaBoostM2 model, a technique for multiclass classification to conduct boosting for the projection of credit risk values (low, medium, high) by using the information of financial ratios. This ensemble method involves the aggregation of multiple weak learners (decision trees), iteratively refining their predictive power. Our dataset comprises 1,623 observations, out of which 150 are used for training. AdaBoostM2 employs a weighted pseudo-loss function to measure classification accuracy, particularly suitable for multiple classes. Initially, the ensemble prioritizes low pseudo-loss values in the early training steps, indicating strong performance from the first few learners. Subsequently, as the ensemble grows, the learning rate diminishes, gradually approaching a pseudo-loss value of 0.5, which represents random chance.

As presented in Figure 1, it is observed that a reduction in error decreases when leaf size increases. Specifically, the impact of altering decision leaf sizes ranging from 1 to 25 can be found as the error drops by 27%, from an initial 0.37 to 0.27. However, the error rate improvement diminishes as the number of trees increases in the ensemble, indicating the diminishing returns of additional trees.

We next evaluate the performance of the model by applying the confusion matrix for each class, expressed as a percentage of the true rating. Specifically, the matrix aims to present the effectiveness of the automated classification process in predicting credit ratings compared to the ratings assigned by human credit rating agents. The first metric is the accuracy of predicted ratings, generated through automated classification, juxtaposed with the actual ratings determined by human agents. These human assessments incorporate a wide array of information, including economic conditions, news, subjective judgment, and potentially other pertinent

Figure 1: Train classification error

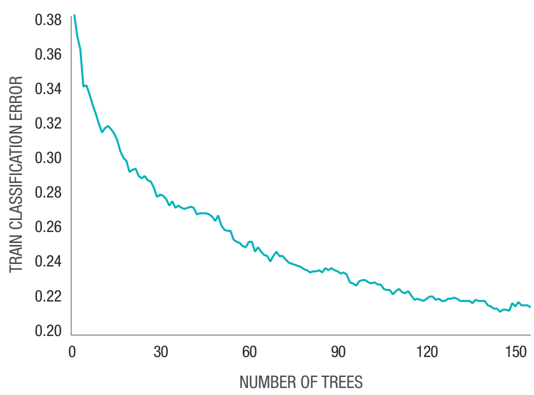


Figure 2. Normalized confusion matrix in percentage: boosting method

TRUE CLASS	1-low	2-medium	3-high	Accuracy	Error
1-low	67.6%	30.5%	1.9%	67.6%	32.4%
2-medium	12.5%	67.2%	20.3%	67.2%	32.8%
3-high	2.9%	30.6%	66.5%	66.5%	33.5%
	Predicted classes				

Figure 3: Classification error for different leaf sizes

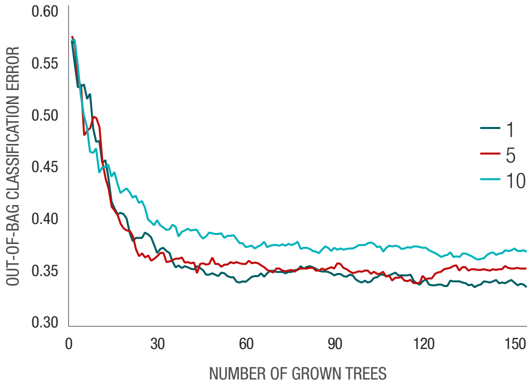
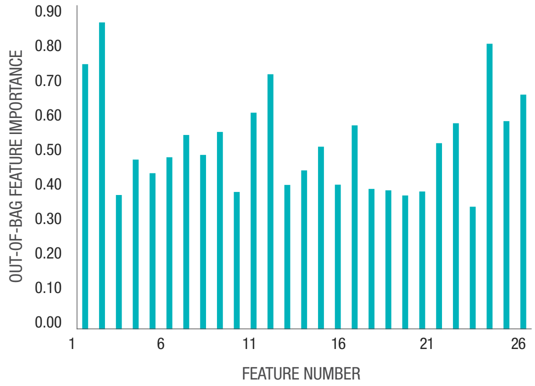


Figure 4: Feature importance analysis



data. The second metric pertains to the accuracy of actual ratings that evaluate the extent to which these ratings reflect the default risk of companies. This ex-post analysis involves scrutinizing which companies experienced defaults within a specified period, typically one year. The analysis encompasses follow-up information on previously evaluated companies, encompassing the ratings assigned and a binary flag indicating whether a company defaulted within a year of the rating process.

Since the primary objective of employing an automated classifier is to expedite the work of the credit committee, enhancing accuracy in predicted ratings can improve the efficiency of reviewing these ratings. Consequently, regular assessments are essential to ascertain the alignment between predicted and final ratings. Any significant disparities may trigger recommendations for retraining the automated classifier.

The confusion matrix, illustrated in Figure 2, compares the predicted and actual ratings. We normalize the values by dividing them by the number of observations with the same true rating to ensure a meaningful assessment. An ideal outcome would manifest as values in the main diagonal dominating the other entries in each row, ideally approaching 1. Our model exhibits accuracy rates of approximately 67% across the three credit rating categories.

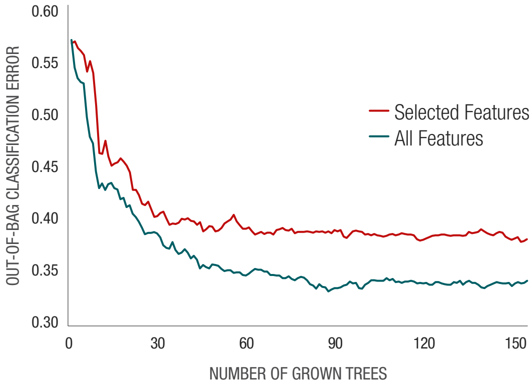
4.2. Decision forest method

The first phase of constructing the classification ensemble, or tree bagger, is to determine an optimal leaf size for individual trees. In this setting, there is no requirement to partition the data into distinct training and test subsets, as this partitioning process occurs internally and implicitly during the sampling procedure. The classification error trends for various leaf sizes (1, 5, and 10 in this analysis) are examined while considering a maximum of 150 trees in the ensemble. To ensure reproducibility and facilitate comparisons, the random number generator is reset for each iteration, allowing for the resampling of data for classifier construction. Figure 3 demonstrates the comparable errors observed across the three leaf-size options and suggests that a leaf size of ten is preferable, as it leads to the development of more streamlined trees and enhances computational efficiency.

The training set consists of the bootstrap replica for each bootstrap iteration. For any “out-of-bag” samples, the observations left out are employed as test points to estimate the out-of-bag classification error. To maintain the efficiency of computational processes and yield leaner trees, a sample size of 10 is employed in this study.

The subsequent step involves the assessment of feature importance to discern their contribution to improving the accuracy of the risk classifier. As presented in Figure 4, certain features emerge as pivotal among the feature set. Specifically, the rating agency (Feature 1), industry (Feature 2), return on capital employed (Feature 12), operating cash flow per share (Feature 25), operating cash flow over sales ratio (Feature 26), and payables turnover (Feature 27) stand out as the most influential predictors within this dataset. It is noteworthy that the inclusion of these features underscores their substantial role in predictive accuracy. Furthermore, the significance of these features aligns with established structural models, such as Merton (1974), wherein the assessment of default risk hinges on the relationship between a firm’s equity value and its level of liabilities. Consequently, these influential features bear relevance to foundational models of credit assessment and underscore their pertinence in the present analysis.

Figure 5: Comparison of classification error: all features and selected features



While some features may not exhibit significance as pronounced as the above, they hold potential importance in the predictive model. For instance, the positive correlation between retained earnings and a firm's age suggests that these variables warrant closer consideration. The process of feature selection aims to identify the most influential predictors and, in this context, those exceeding a predefined threshold of 0.7 merit inclusion. Subsequently, a novel classification ensemble is trained to utilize solely the selected highly important features, and its classification error is subjected to comparison with the error derived from the preceding classifier employing all available features. This comparative analysis serves to illuminate the performance discrepancy between two distinct predictor sets, denoted as "all features" and "selected features", respectively. The aim is to discern whether a refined feature selection strategy can enhance classification accuracy and to what extent these additional features contribute to model refinement.

Figure 5 presents a comparison of classification errors between using all features and utilizing selected high-importance features. The classification accuracy exhibits minimal deterioration when less crucial features are excluded from the ensemble, indicating that featuring selected predictors is suitable for subsequent predictions. The process of feature selection can be time-intensive when the initial set comprises a multitude of variables. However, its success hinges on a judicious blend of quantitative tools and the discernment of the analyst. The variable importance measure, thus, serves as a mechanism of ranking to assess the relative impact of each feature by evaluating the extent to which random permutation of its values affects the predictive accuracy of risk classification.

The method discerns features that significantly contribute to predictive power. As indicated in Figure 5, the selected features do not perform better than no feature selection. This can be caused by information loss due to the reduction in dataset dimensionality resulting from the exclusion of specific features. The complex interactions of the features and some non-selected features that can be relevant to these interactions make the predictions less accurate. When dealing with two strongly correlated features of importance, both may receive high ranks in the analysis. In such cases, retaining just one of these features may suffice for accurate predictions, but this determination may not solely be derived from ranking results. Under this situation, one may need to consider additional examination of feature correlations or expert judgment. Consequently, while quantitative tools are useful in feature selection, the informed judgment of the human analyst remains an indispensable component of the process.

Figure 6. Normalized confusion matrix in percentage: decision tree

TRUE CLASS	1-low	2-medium	3-high	Accuracy	Error
1-low	62.9%	28.6%	8.6%	62.9%	37.1%
2-medium	10.2%	59.4%	30.5%	59.4%	40.6%
3-high	2.3%	18.5%	79.2%	79.2%	20.8%
	Predicted classes				

Once the model parameters have been determined, the classifier can be saved for future sessions when it is necessary to rate new clients. To predict the credit rating for new data, the "predict" method is invoked on the classifier and yields the predicted class and the associated classification score as key outputs. Among them, the classification scores provide insights into the degree of confidence associated with the predicted ratings.

Figure 6 presents a normalized confusion matrix utilizing the decision tree method. The accuracy rates for low and medium classifications stand at 63% and 59%, respectively, which are marginally lower than those achieved through the boosting method. Conversely, the accuracy rate for high credit ratings reaches 79%, surpassing the corresponding rate achieved using the boosting method. These insights underscore the nuances in classifier performance across different credit rating categories.

5. CONCLUSION

This study evaluates the effectiveness of decision forest and boosting algorithms in predicting credit ratings. By leveraging financial ratios as input variables, various machine learning tools in MATLAB are used in this study. Using the quarterly financial data of 2,029 credit ratings in 12 industries that were rated by five different agents between 2010 and 2016, we first reclassify the ratings to avoid the issue related to too few observations.

The empirical findings demonstrate that both these methods exhibit considerable performance but may be superior to each other in different areas. Boosting algorithms method exhibits accuracy rates of approximately 67% across the credit rating categories. The random forests model generates lower accuracy rates for low and medium classifications than the boosting method, but the accuracy rate for high credit ratings reaches 79%, more accurate than the outcome using the boosting method.

This study exemplifies how to apply appropriate machine learning models in forecasting default risk by using financial data. Furthermore, we show the usefulness of both methods exhibiting robustness when handling noisy data as they expedite training with large datasets and enhance the interpretability of the findings. The results can be useful to practitioners aiming to integrate machine learning algorithms into credit rating prediction tasks.

REFERENCES

Altman, E. 1968, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy" Journal of Finance 23:4, 589-609

BCBC, 2005, "Studies on the validation of internal rating systems," Basel Committee on Banking Supervision, Bank for International Settlements (BIS), https://tinyurl.com/5dbcxtrk

BCBS, 2006, "International convergence of capital measurement and capital standards: a revised framework," Basel Committee on Banking Supervision, Bank for International Settlements (BIS), https://tinyurl.com/4tnbm7wh

Loeffler, G., and P. N. Posch, 2007, Credit risk modeling using Excel and VBA, Wiley Finance

Hastie, T., R. Tibshirani, and J. Friedman, 2008, The elements of statistical learning, Springer

Merton, R., 1974, "On the pricing of corporate debt: the risk structure of interest rates," Journal of Finance 29:2, 449-470

ABOUT CAPCO

Capco, a Wipro company, is a global technology and management consultancy focused in the financial services industry. Capco operates at the intersection of business and technology by combining innovative thinking with unrivalled industry knowledge to fast-track digital initiatives for banking and payments, capital markets, wealth and asset management, insurance, and the energy sector. Capco's cutting-edge ingenuity is brought to life through its award-winning Be Yourself At Work culture and diverse talent.

To learn more, visit www.capco.com or follow us on Facebook, YouTube, LinkedIn and Instagram.

WORLDWIDE OFFICES

APAC
Bangalore – Electronic City
Bangalore – Sarjapur Road
Bangkok
Chennai
Dubai
Gurgaon
Hong Kong
Hyderabad
Kuala Lumpur
Mumbai
Pune
Singapore

EUROPE
Berlin
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
London
Milan
Munich
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA
Charlotte
Chicago
Dallas
Hartford
Houston
New York
Orlando
Toronto
Washington, DC

SOUTH AMERICA
Alphaville
São Paulo

THE COVER IMAGE WAS CREATED USING JASPER AI, AN AI ART GENERATOR

FSC
www.fsc.org
MIX
Paper from responsible sources
FSC® C004309

WWW.CAPCO.COM

CAPCO 25
a wipro company