

An empirical evaluation of machine learning performance in corporate sales growth prediction

Miho Saito¹, Takaya Ohsato² and Suguru Yamanaka^{3,4*}

¹ Graduate school of science and engineering, Aoyama Gakuin University, 5-10-1 Fuchinobe Chuo-ku, Sagami-hara-shi, Kanagawa 232-5258, Japan

² Center for data science education and research, Shiga University, 1-1-1 Banba, Hikone-shi, Shiga 522-8522, Japan

³ College of science and engineering, Aoyama Gakuin University, 5-10-1 Fuchinobe Chuo-ku, Sagami-hara-shi, Kanagawa 232-5258, Japan

⁴ Institute of innovative research, Tokyo institute of technology, 4259 Nagatsutacho Midori-ku, Yokohama-shi, Kanagawa 226-8503, Japan

*Corresponding author: syamanaka@gem.aoyama.ac.jp

Received February 26, 2021, Accepted March 22, 2021

Abstract

Corporate performance prediction has attracted considerable research interest in the investment and financial risk management fields. This study comprehensively examines the ability of machine learning algorithms to integrate analysis of sales growth prediction, with specific focus on random forest, weighted random forest, gradient boosting decision tree, and support vector machine, as well as a least-squares probabilistic classifier. We carried out an experimental comparison study over a dataset comprising real corporate data on the effectiveness of these five machine-learning algorithms. The results showed sufficient performance for some machine-learning algorithms.

Keywords sales growth, classification, machine learning

Research Activity Group Mathematical Finance

1. Introduction

Decision-makers, including business analysts, creditors, investors, and financial managers have used financial ratios, a traditional yet powerful tool, for predicting corporate performance. Various methodologies have been developed to predict financial performance in association with financial ratios. These studies have focused on predicting bankruptcy and stock returns using various statistical and data-mining techniques, including machine learning algorithms. In this study, we focus on predicting sales growth and conduct a comprehensive experimental comparison study on the effectiveness of machine learning algorithms.

There are many studies on machine learning methods applied to predict future firm performance. Bankruptcy prediction and credit rating classification studies use traditional statistical methods and machine learning techniques, including random forests [1], support vector machines [2–6], and gradient boosting decision trees [7, 8]. In addition, return on equity (ROE) studies use neural network models [9] and decision trees [10]. The idea of employing random forest for firm performance prediction is also suggested in [11].

This study is a comparative analysis of sales growth predictions using financial ratios. We focus on widely used machine learning algorithms, that is, random forest, weighted random forest, gradient boosting decision tree, and support vector machine algorithms. In addition to these, we employ a least-squares probabilistic

classifier, which shows high classification accuracy for the credit rating classification problem [12]. We compare their performance in analyzing four different sectors of the Japanese economy. It is noteworthy that we employ a database that includes samples of unlisted firms, although most previous studies have only analyzed listed firms.

The remainder of this paper is organized as follows. Section 2 briefly describes the formulation of our classification problem and discusses the methodology, including machine learning algorithms, to be examined in this study. Section 3 presents the empirical results on sales growth classification of our dataset, which comprises real Japanese corporate data. Section 4 concludes.

2. Methods

2.1 Formulation of classification problem

In this study, we treat the sales-growth prediction problem as a binary classification problem. Let $y_i \in \{0, 1\}$ be the response variable value for observation i , $i = 1, 2, \dots, n$. Here, $y_i = 1$ and $y_i = 0$ indicate firm i 's future sales growth and non-sales growth, respectively. The sales-growth classification aims to determine the value of label y_i at a particular time, given the values of the covariates x_i . Here, vector $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ represents the sample-specific covariates associated with growth label y_i . The list of covariates x includes the target firms' financial ratios, which are calculated using the financial statements of the firm.

2.2 Data preprocessing

Data preprocessing is an integral step in machine learning, as data quality and useful information affects our model’s ability to learn directly. The first step addresses the null or NaN values of the covariates. We handle this problem by simply removing samples that include null or NaN values. Then, we transform the covariates so that they can be recognized samples obtained from the standard normal distribution, with a mean of 0 and a standard deviation of 1. Specifically, we employ the Yeo-Johnson transformation proposed in [13]. Next, we employed the synthetic minority over-sampling technique (SMOTE) [14] to handle the imbalance of the data. SMOTE attempts to construct new minority class samples by interpolating and selecting a near minority-class neighbor randomly.

We employ the least absolute shrinkage and selection operator (lasso) [15] to select covariates. Lasso is a regression method that enables simultaneous estimation and variable selection in a non-orthogonal setting. Under a suitable choice of penalty power, the lasso selects covariates by forcing some coefficients to zero and shrinking others. In our empirical study, we applied five-fold cross-validation for the lasso and obtained five sets of estimated coefficients. We then averaged the five estimated coefficients for all covariates and detected the most important covariate as that with the largest absolute value of the averaged coefficients. Finally, we selected covariates for which the absolute values of the averaged coefficients were larger than 10% of the largest coefficient, for machine learning inputs.

2.3 Machine learning algorithm

In our comparative analysis, we employed five machine learning algorithms, that is, random forests (RF), weighted random forests (WRF), gradient boosting decision tree (GBDT), support vector machine (SVM), and least-square probability classifier (LSPC). In the following, we briefly describe the machine learning algorithms.

RF [16], WRF [17], and GBDT [18] are considered advanced decision tree techniques. A decision tree is a classification approach to analyzing data based on a tree structure. A decision tree classifies an instance by sorting it through the tree to the appropriate leaf node, that is, each leaf node represents a classification. Each node represents an attribute of the instance, and each branch corresponds to one of the possible values for this attribute. Instead of using a single classification tree, RF, WRF, and GBDT grow a large set of trees, and the final prediction is obtained as the average of the predictions stemming from the individual trees. The RF model grows the set of trees using a different bootstrapped sample of the original dataset for each tree and selecting the best split at each branch using only a randomly selected subset of the covariates.

The WRF model aims to reduce the bias by assigning different weights to different classes for imbalanced training datasets. In weighted random forests, misclassifying a class with a higher weight carries a higher penalty than misclassifying a class with a lower weight. Thus, the error rate calculation is more heavily weighted for a

higher weighted class. The class weights are incorporated to evaluate the criterion for finding splits and to determine the terminal class by a weighted majority vote. Imbalances in error rates among classes can be used to select the weights.

GBDT has gained attention in machine learning in recent years and is popular for its use in solving classification and regression problems. The GBDT grows the set of trees recursively using a learning-from-mistakes approach, where classification errors from the previous trees are used as the dependent variable to grow the next tree. Specifically, we employed the Light GBM (Light Gradient Boosting Machine) [19] for the GBDT implementation. LightGBM is a quick and efficient GBDT algorithm designed by Microsoft Research Asia in 2016, in an open-source promotion framework. This algorithm is used in sorting, classification, regression, and many other machine-learning tasks.

SVM [20] is another popular supervised machine learning algorithm and has been extensively applied in the field of credit scoring because of its powerful predictive capabilities. It projects the input data into a high-dimensional feature space and then finds a hyperplane supported by the support vectors to separate the two classes with a maximal margin.

LSPC [21] is a probabilistic classification that employs a linear combination of kernel functions, and its parameters are learned by least-squares fitting of the true class-posterior probability.

For the implementation of RF, WRF, and SVM, we employed the machine learning package, `scikit-learn`, for Python. We employed the `light-gbm` package for Python for the GBDT implementation. We originally implemented LSPC with Python, without using any machine learning packages. The hyperparameters of each machine learning algorithm were determined using 3-fold cross-validations.

2.4 Performance metrics

To evaluate the accuracy of the classification results, we introduce a widely used measure called the accuracy ratio (AR), weighted F1-scores, and the area under the ROC curve (AUC). The AR measures the overall effectiveness of the model and is defined as the number of correct predictions over the number of predictions, written as:

$$\text{AR} = \frac{tp + tn}{tp + tn + fp + fn}$$

where tp is the number of true positive predictions, tn is the number of true negative predictions, fp is the number of false-positive predictions, and fn is the number of false-negative predictions. A higher AR implies a better prediction performance, and the AR of the random prediction is 0.5. The AR of practical default prediction models is $0.6 \sim 0.8$ [22].

However, the AR can be misleading for very imbalanced data, where a classification missing all minority classes can still achieve a high overall accuracy. To overcome this obstacle, we also report the performance using the F1-score and the AUC, which focus on the incorrect

Table 1. Top five important covariates selected by Lasso.

Software industry:	
Variables	Definition
Sales growth rate	(Current period sales – Prior period sales) / Prior period sales
Total capital growth rate	(Current period capital – Prior period capital) / Prior period capital
Sales to administrative expense ratio	Sales / Administrative expenses
Shareholder equity to capital ratio	Shareholder equity / Capital
Asset turnover	Sales/[(Opening assets+ closing assets)/2]
Service industry:	
Variables	Definition
Sales growth rate	(Current period sales – Prior period sales) / Prior period sales
Total capital growth rate	(Current period capital – Prior period capital) / Prior period capital
Sales to administrative expense ratio	Sales / Administrative expenses
Shareholder equity to capital ratio	Shareholder equity / capital
Asset turnover	Sales/[(Opening assets+ closing assets)/2]
Food industry:	
Variables	Definition
Sales growth rate	(Current period sales – Prior period sales) / Prior period sales
Total capital growth rate	(Current period capital – Prior period capital) / Prior period capital
OI growth rate	(Current period OI– Prior period ordinary income) / Prior period OI
Annual increase of OI to asset	Current period OI/ Total assets – Prior period income/ Total assets
Annual increase of return on asset	Current period net income/ Total assets – Prior period net income/ Total assets
Manufacturing industry:	
Variables	Definition
Sales growth rate	(Current period sales – Prior period sales) / Prior period sales
Total capital growth rate	(Current period capital – Prior period capital) / Prior period capital
Operating ratio	(Operating expenses + Cost of goods sold)/Net sales
Sales to administrative expense ratio	Sales / Administrative expenses
Non-OI to sales ratio	Non-OI/ Sales

Remark: OI is an abbreviation for operating income.

classification of the minority class. F1-scores are the harmonic mean of precision and recall, that is,

$$F1 = \frac{2(\textit{precision} \times \textit{recall})}{\textit{precision} + \textit{recall}}$$

where $\textit{precision} = tp/(tp + fp)$ and $\textit{recall} = tp/(tp + fn)$. The value of the weighted F1-score is the weighted average of the F1-scores for each class label, using the number of true instances for each class label for weights.

The AUC is also a widely used performance metric for classification problems, and is considered a better test of classification than the AR for determining which of the models predicts the class best. The ROC curve is obtained by plotting the hit versus false alarm rates. The AUC of the random prediction is 0.5, and the higher the model’s AUC, the better the performance. Previous studies assert that a model possesses sufficient prediction power if its AUC exceeds 0.7 [23].

3. Comparative analysis

3.1 Data

We use an extensive dataset of financial statements and sales growth indicators for Japanese listed and unlisted firms in the software, service, food, and manufacturing industries in 2015. The samples are taken from Teikoku Data Bank, Ltd’s database. Each sample firm is labeled as a sales-growth firm ($y_i = 1$) in a given year if the sales growth ratio from 2015 to 2018 is greater than one standard deviation point from the average on the distribution of the observed sales growth ratio over

all sample firms in the industry. “The numbers of sales-growth firm samples/The numbers of all samples” for the software, service, food, and manufacturing industries are 365/1804, 1440/10502, 51/1192, and 878/12082, respectively.

3.2 Results

Table 1 shows the selected covariates for each industry. Remarkably, the sales growth rate and total capital growth rate are selected for all four industries. In addition, the sales-to-administrative expense ratio tends to be selected.

Table 2 shows the classification performance of each model. The AR and AUC are above 50% in most cases, which indicates that the machine learning predictions are better than random prediction. In particular, the AR for the food industry classification is above 90%, which is considered a comparatively good result. No single classifier dominated among those that showed the best results in each performance metric; SVM showed the best performance in AR, although RF showed the best performance three times in F-score and AUC.

4. Concluding remarks

In this study, we examined the effectiveness of five machine learning algorithms for sales growth prediction. We applied all methods to Japanese sample firms, including unlisted firms, from four industries. For data preprocessing, we employed the data normalization technique of the Yeo-Johnson transformation and selected variables

Table 2. Classification results.

Software:

	AR	F-score	AUC
RF	0.7345	0.4333	0.7054
WRF	0.7311	0.4326	0.7047
GBM	0.7533	0.3498	0.5969
SVM	0.7761	0.0685	0.6240
LSPC	0.6652	0.3195	0.5684

Service:

	AR	F-score	AUC
RF	0.7597	0.4155	0.7760
WRF	0.7606	0.4149	0.7760
GBM	0.8066	0.3619	0.6356
SVM	0.8447	0.1159	0.6886
LSPC	0.7513	0.3148	0.6188

Food:

	AR	F-score	AUC
RF	0.9161	0.1937	0.7626
WRF	0.9127	0.2008	0.7638
GBM	0.9321	0.1767	0.5702
SVM	0.9329	0.1321	0.6463
LSPC	0.8649	0.1476	0.5718

Manufacturing:

	AR	F-score	AUC
RF	0.7561	0.1949	0.6745
WRF	0.7575	0.1975	0.6739
GBM	0.8480	0.1428	0.5375
SVM	0.9051	0.0466	0.6213
LSPC	0.7925	0.1288	0.4669

automatically according to the lasso results. The machine learning algorithms show classification accuracy in the range of 70%~90%. In addition, some machine learning algorithms showed AUC values above 0.7. To some extent, the results imply the feasibility of a machine learning system for sales growth prediction. The realization of a sales-growth prediction system would encourage investment in Japanese firms and lead to the growth of the Japanese economy.

Acknowledgments

Teikoku Databank, Ltd. supported our research by providing data related to Japanese business firms and by financially supporting the Center for TDB Advanced Data Analysis and Modeling, Tokyo Institute of Technology for academic research purposes. This work was partially supported by JSPS KAKENHI (Grant Number JP18K12818).

References

[1] M. Ryser and S. Denzler, Selecting credit rating models: a cross-validation-based comparison of discriminatory power, *Financial Mark. Portf. Manag.*, **23** (2009), 187–203.
 [2] Z. Huang, H. Chen, C.-J. Hsu, W.-H. Chen and S. Wu, Credit rating analysis with support vector machines and neural networks: a market comparative study, *Decis. Support Syst.*, **37** (2004), 543–558.
 [3] L. Cao, L. K. Guan and Z. Jingqing, Bond rating using support vector machine, *Intell. Data Anal.*, **10** (2006), 285–296.
 [4] Y.-C Lee, Application of support vector machines to corporate credit rating prediction, *Expert Syst. Appl.*, **33** (2007), 67–74.
 [5] K. Kim and H. Ahn, A corporate credit rating model using

multi-class support vector machines with an ordinal pairwise partitioning approach, *Comput. Oper. Res.*, **39** (2012), 1800–1811.
 [6] K. Tanaka and H. Nakagawa, A method of corporate credit rating classification based on support vector machine and its validation in comparison of sequential logit model (in Japanese), *Trans. Oper. Res. Soc. Jpn.*, **57** (2014), 92–111.
 [7] Y.-C Chang, K.-H Chang and G.-J Wu, Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions, *Appl. Soft Comput.*, **73** (2018), 914–920.
 [8] T. Le, B. Vo, H. Fujita, N.-T Nguyen and S. W. Baik, A fast and accurate approach for bankruptcy forecasting using squared logistics loss with GPU-based extreme gradient boosting, *Inform. Sci.*, **494** (2019), 294–310.
 [9] M. Lam, Neural network techniques for financial performance prediction: integrating fundamental and technical analysis, *Decis. Support Syst.*, **37** (2004), 567–581.
 [10] D. Delen, C. Kuzey and A. Uyar, Measuring firm performance using financial ratios: A decision tree approach, *Expert Syst. Appl.*, **40** (2013), 3970–3983.
 [11] D. Miyakawa, Enterprise information processing device, enterprise event prediction method and prediction program, JP Patent 6611068, 2019-11-27.
 [12] M. Saito and S. Yamanaka, Performance evaluation of least-squares probabilistic classifier for corporate credit rating classification problem, *JSIAM Lett.*, **13** (2021), 9–12.
 [13] I.-K. Yeo and R. A. Johnson, A new family of power transformations to improve normality or symmetry, *Biometrika*, **87** (2000), 954–959.
 [14] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, *J. Artif. Intell. Res.*, **16** (2002), 321–357.
 [15] R. Tibshirani, Regression shrinkage and selection via the Lasso, *J. R. Statist. Soc. B*, **58** (1996), 267–288.
 [16] L. Breiman, Random forests, *Mach. Learn.*, **45** (2001), 5–32.
 [17] C. Chen, A. Liaw and L. Breiman, Using random forest to learn imbalanced data, Technical Report 666, University of California, Berkeley, 2004.
 [18] J. H. Friedman, Greedy function approximation: A gradient boosting machine, *Ann. Statist.*, **29** (2001), 1189–1232.
 [19] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye and T.-Y Liu, LightGBM: a highly efficient gradient boosting decision tree, *Adv. Neural. Inf. Process. Syst.*, **30** (2017), 3149–3157.
 [20] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
 [21] M. Sugiyama, Superfast-trainable multi-class probabilistic classifier by least-squares posterior fitting, *IEICE Trans. Inf. & Syst.*, **E93-D** (2010), 2690–2701.
 [22] S. Yamashita and K. Miura, Shinyo risuku moderu no yosoku seido [Prediction accuracy for credit risk models] (in Japanese), Asakura Shoten, Tokyo, 2011.
 [23] T. Ono, S. Yamashita and H. Tsubaki, Default distribution model truncated by stochastic credit decision (in Japanese), *Proc. Inst. Statist. Math.*, **59** (2011), 3–23.