



Comparison of Empirical Methods for the Reproduction of Global Manufacturing Companies' Credit Ratings

Sergei Grishunin¹, Svetlana Suloeva², Alexandra Egorova¹, Ekaterina Burova^{2*}

¹National Research University, Higher School of Economics, 20 Myasnitskaya Ulitsa, Moscow, 101000, Russia

²Peter the Great St. Petersburg Polytechnic University, 29 Politechnicheskaya Ulitsa, St. Petersburg, 195251, Russia

Abstract. The quantitative assessment of the credit quality of manufacturing companies is a task of great interest to researchers and practitioners. This is underpinned by the elevated credit risk of these companies stemming from rapid technological changes. However, few studies have addressed this issue specifically for manufacturing companies. This study aimed to fill this research gap by comparing the predictive power of various methods in reproducing manufacturing companies' public credit ratings from available financial and non-financial data. The sample included 109 manufacturing companies from developed and emerging markets over the period 2005–2016. The analysis included three methods: ordered logistic regression (OLR) and two machine learning techniques, random forest and gradient boosting. The results showed that machine learning techniques outperformed OLR in terms of predictive power. In the best specification model, random forest had an accuracy of 50%, followed by gradient boosting (47%) and OLR (25%). We also tested two types of sampling in the training and test sets: random and time-dependent. The results showed that the models' predictive power was greater with random sampling. The inclusion of macroeconomic variables did not improve the models' predictive power due to the rating agencies' preferred through-the-cycle rating approach. The study's findings have implications for the development of manufacturing firms' internal credit ratings. They can also be useful for researchers exploring the accuracy of empirical models in predicting industrial firms' insolvency and creditworthiness.

Keywords: Credit rating; Machine learning; Manufacturing companies; Rating agencies; Risk management

1. Introduction

Over the last decade, the manufacturing industry has been influenced by a new wave of scientific and technological progress, which has been called “the fourth industrial revolution” (Liao et al., 2017). However, these opportunities come with new threats that increase the credit risks of manufacturing firms. Thus, selecting the credit risk assessment model with the highest accuracy has become increasingly important. For manufacturing firms, these models are used by investors to make funding decisions. They are also needed for the calculation of bad debt provisions as required by Basel III (Basel Committee on Banking Supervision, 2017) or the international accounting standard IFRS9. Lastly, they are used by manufacturing firms' stakeholders to control, diagnose, and monitor their credit

*Corresponding author's email: burova_ev@spbstu.ru, Tel.: +7-921-750-61-40
doi: [10.14716/ijtech.v11i6.4424](https://doi.org/10.14716/ijtech.v11i6.4424)

quality (Cuny and Dube, 2017).

Public credit ratings (PCRs) are the most common measure of creditworthiness. These ratings are assigned by domestic and international rating agencies, such as Moody's Investors Service, Fitch Ratings, and Standard & Poor's Financial Services (Karminsky and Peresetsky, 2009). The ratings provide a consistent global framework for accurate assessment and comparison of companies' and countries' credit quality (Karminsky and Polozov, 2016). However, ratings are often assigned to large, diversified manufacturers mainly from developed markets. This is due to the high cost of the rating, the need to provide agencies with a large amount of information, and the reluctance of small companies to publish low ratings. Other limitations of PCRs are their long update intervals, as well as errors and inefficiencies of rating agencies which were found by various studies (Langohr and Langohr, 2008). The errors included maintenance ratings at the investment grade level for companies with high credit risks or slow response to the credit crisis. To address these issues, investors develop internal credit rating models (ICRs). A common approach is to reproduce the PCRs of unrated companies from available financial and non-financial data using empirical methods. ICRs have proven to be objective and low-cost. However, their predictive power varies significantly depending on the underlying models (Karminsky and Peresetsky, 2007). The task of comparing various models and selecting the optimal one has therefore become extremely important.

The aims of this study were: (1) to compare the predictive power of various empirical methods in reproducing Moody's ratings specifically for manufacturing firms; and (2) to identify the optimal model in terms of data availability, forecast accuracy, and outcome interpretability. We also examined whether the addition of macroeconomic factors to the set of explanatory variables increases the models' prediction accuracy. This study contributes to the literature in several ways. First, it focuses on manufacturing companies, which are not the focus of most studies. Second, it applies explanatory variables that match those used in Moody's methodology. Third, it uses data spanning a long period (from 2005 to 2016). The study's findings have implications for the development of manufacturing firms' internal credit ratings. The results can be useful for researchers exploring the ability of empirical models to predict industrial firms' insolvency and creditworthiness.

2. Methods

2.1. Literature Review

Previous studies have explored a wide range of empirical methods for the assessment of credit quality and reproduction of PCRs. However, most studies have focused on financial institutions, whereas little research has focused on non-financial companies, including industrial companies.

Until the 1980s, univariate methods, ordinary least square regression, and linear and multiple discriminant analyses (DA) were commonly used (Altman et al., 1977; Bhushan and Reddy, 2016). However, these methods were limited in predictive power (Demeshchev and Tikhonova, 2014). Since the 1980s, ordered logistic regression (ORL) and ordered probit regression (OPR) models have been used for credit rating reproduction (Amato and Furfine, 2004; Kamstra et al., 2001; Karminsky, 2011). A limitation of these models is their moderate predictive power, ranging between 37% and 43%. Higher accuracy can be achieved by combining OPR with the least absolute shrinkage and selection operator (Serpiniis et al., 2018). Another limitation of these models is that they are susceptible to multicollinearity and overfitting. In recent years, applied machine learning (ML) methods, such as support vector machines (SVM), artificial neural networks (ANN), and classification trees, have gained popularity due to their higher predictive power (Huang et al., 2004; Lee, 2007; Saitoh, 2016). Ye et al. (2008) reproduced Moody's PCRs for companies from diverse

industries between 1999 and 2001 using various ML techniques. SVM had the highest accuracy (64%), followed by decision tree learning (47%), OPR (28%), and OLR (23%). [Kumar and Bhattacharya \(2006\)](#) reported that an ANN model was more accurate than linear DA in forecasting Moody's ratings (79% vs. 33%). The disadvantages of ML methods are low interpretability and overfitting ([Tsai and Chen, 2010](#)). The models built with ML methods provide insights into the captured dependencies using such tools as the relative variable influence and partial dependence plots. However, these tools cannot explain how the variables affect PCRs.

Most studies do not focus on a particular industry but include diverse sets of industries, and they only consider companies from individual countries. Other limitations include limited sample time spans and the application of explanatory variables that do not match those used by rating agencies. The aim of our study was to fill these gaps.

2.2. Model, Data, and Explanatory Variables

The theoretical basis of this research was informed by the studies of [Karminsky and Peresetsky \(2007, 2009\)](#), [Demeshev and Tikhonova \(2014\)](#), [Amato and Furfine \(2004\)](#), [Saitoh \(2016\)](#), [Huang et al. \(2004\)](#), [Sermpinis et al. \(2018\)](#) and [Tsai and Chen \(2010\)](#). We modeled long-term credit ratings assigned by Moody's and applied three methods: OLR, random forest (RF), and gradient boosting (GB). Moody's agency expresses ratings using letters and letter-number combinations ([Moody's, 2020](#)). We converted ratings to integers in an ascending order: the lower the rating, the higher the number (Table 1).

Table 1 Numerical scale for dependent variables

Rating letter	Rating number
AAA	1
Aa1	2
Aa2	3
Aa3	4
A1	5
A2	6
A3	7
Baa1	8
Baa2	9
Baa3	10
Ba1	11
Ba2	12
Ba3	13
B1	14
B2	15
B3	16
Caa1–Caa3	17
C–Ca	18

The dataset contained 880 observations of financial and non-financial metrics as well as Moody's ratings of 109 manufacturing companies from developed and emerging economies over the period 2005–2016. These companies served primarily metals and mining, telecommunications, energy, medical, agriculture, and construction end-markets. The time structure of the dataset covered a full credit cycle from a low number of defaults to a peak in 2008 and then a return to previous levels in 2011–2013 (Figure 1). That 2008 peak resulted from bursting of houses bubble in the United States damaging financial institutions and non-financial companies over the world. The need to account for a full

credit cycle is underpinned by the fact that rating agencies assign ratings on a through-the-cycle basis (Karminsky and Polozov, 2016). However, our sample is unbalanced in terms of ratings (Figure 2) due to the specific features of the manufacturing industry: the business-to-business nature of operations, high capital intensity, volatility of demand, and long production cycles.

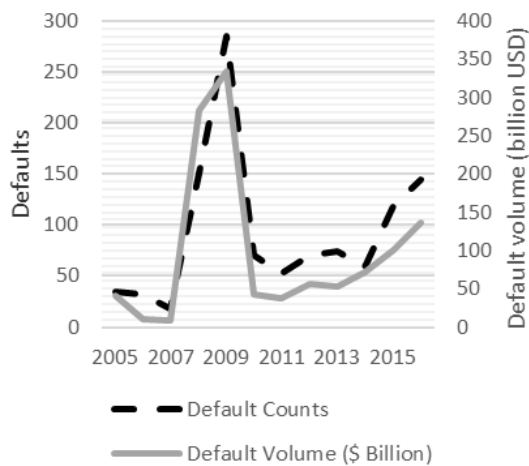


Figure 1 Main credit cycle in 2005–2016
Source: Moody's (2020)

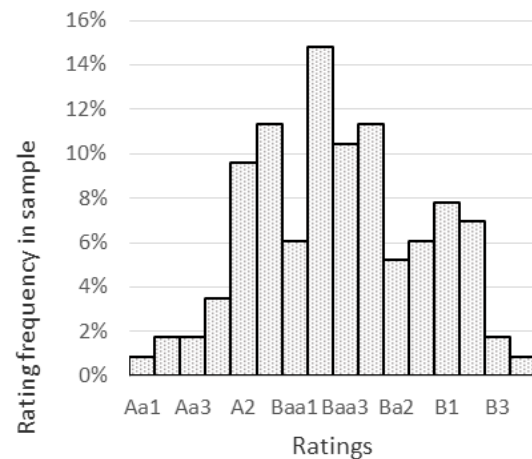


Figure 2 Rating distribution in the sample
Source: Moody's (2020)

The explanatory variables included the financial ratios reflecting the companies' performance and macroeconomic variables in their countries of origin. To ensure the models' greatest predictive power, we brought the list of variables as close as possible to the set of factors used by Moody's in its rating process. We believe that this list fully reflects the properties of manufacturing companies, such as long production cycles, moderately volatile demand and prices, and a high share of fixed costs in their cost structures. The financial ratios were obtained from Moody's methodology for manufacturing companies (Moody's, 2020). Financial data and ratings were taken from Thomson Reuters Eikon. Macroeconomic variables were retrieved from the World Bank's website (World Bank Open Data, 2020). The list of variables and their descriptive statistics and expected signs are displayed in Table 2.

Table 2 Explanatory variables in the model

Variable	Unit	Expected sign	Mean	Standard deviation
Business ratios				
Share of gross investments in GDP	%	-	20.8	20.4
Share in global manufacturing	%	-	0.14	0.21
Time trend	years	+	5	3
Economic downturn flag	1/0	+	—	—
Private company flag	1/0	+	—	—
Resident in developed country flag	1/0	-	—	—
Quality of fixed assets	%	+	8.1	3.5
Market value to sales multiple	—	-	1.76	1.54
Market value to EBITDA multiple	—	-	7.3	5.8
Interest paid	%	+	4.25	1.16
Profitability ratios				
Return on average equity	%	-	13.7	30.4
EBITDA margin	%	-	15.0	6.0

Variable	Unit	Expected sign	Mean	Standard deviation
Leverage and debt coverage ratios				
Net debt/EBITDA	—	–	2.9	4.5
Debt/Book Capitalization	%	+	61.8	12.7
Debt/Market Capitalization	%	+	27.32	15.3
Cash ratio	%	–	47.4	79.5
Retained cash flow to net debt	%	–	75.2	23.2
Available retained cash flow debt coverage	%	–	16.4	22.0
EBITDA interest coverage	—	–	8.6	11.5
Liquidity ratios				
Current ratio	—	–	1.9	0.7
Quick ratio	—	–	1.1	0.5
Macroeconomic variables				
Real GDP growth	%	–	1.6	2.1
Gross investments to GDP	%	–	20.7	20.4
Inflation	%	uncertain	1.7	1.4
Rule of law	—	–	—	—
Government effectiveness	—	–	—	—
Control of corruption	—	–	—	—

GDP: gross domestic product; EBITDA: earnings before interest, taxes, depreciation, and amortization.
Source: The authors, based on [Moody's \(2020\)](#)

2.3. Research Design

To solve the problem of multicollinearity in the OLR model, we built a correlation matrix to exclude the most closely correlated variables (with pairwise cross-factor correlations above 0.8). For the remaining variables, we estimated the variance inflation factor (VIF) ([Senaviratna and Cooray, 2019](#)) and excluded from consideration all variables with a VIF of over 5. To assess the predictive power of explanatory variables, we applied principal component analysis ([Abdi and Williams, 2010](#)). Conversely, when modeling ratings using ML techniques, we applied the entire set of explanatory variables. The model was built on the training sample and tested on the remaining sample. We divided the data into the training and testing samples in two ways. The first way was to split the data randomly in the ratio of 70%-30% while the second assumed that the earliest 70% of the data assigned to the training set and the most recent 30% assigned to the test set.

2.4. Research Hypotheses

Our first hypothesis (H1) was that of the three models tested, GB would have the highest predictive power, followed by RF, while OLR would have the lowest. This hypothesis was based on the results of [Ye et al. \(2008\)](#) and [Chopra and Bhilare \(2018\)](#). Another argument against OLR was that the coefficients are estimated with maximization of the likelihood function. As our sample was unbalanced, its results would be skewed toward the most frequent ratings.

Our second hypothesis (H2) was that randomly dividing the data into a training and a test set would yield greater predictive power than dividing them in a time-dependent manner (the earliest 70% of the data assigned to the training set and the most recent 30% assigned to the test set). Given the unbalanced nature of the sample, it was expected that random sampling would yield higher accuracy.

Our third hypothesis (H3) was that the addition of macroeconomic variables to the models would improve their predictive power. This hypothesis was based on the findings of [Karminsky and Peresetsky \(2007\)](#) and [Karminsky \(2011\)](#), who showed that the use of macroeconomic variables resulted in a statistically significant increase in predictive power.

To test this hypothesis, we applied the models with and without macroeconomic explanatory variables.

Our fourth hypothesis (H4) was that in the OLR, the actual signs of statistically significant variables would match our expectations about the nature of the correlations (positive or negative) between the ratings and the explanatory variables (Table 2).

3. Results and Discussion

Table 3 presents the results of the model estimation and credit rating reproduction. To assess the models' predictive power, we applied multiclass classification metrics (Hossin and Sulaiman, 2015; Fachrurrazi and Munirwansyah, 2017).

Table 3 Model accuracy estimation results

Model	Accuracy	Modified Accuracy	Kappa Accuracy	McFadden R ²	AIC	Precision	Recall	F1 Score
Time-dependent sample with macro regressors								
Naïve	7.63%	12.70%	-1.57%	—	—	5.88%	5.53%	14.96%
OLR	22.88%	41.52%	14.92%	22.33%	3174	18.40%	19.40%	32.68%
RF	37.29%	46.61%	31.15%	—	—	45.04%	37.35%	41.29%
GB	39.01%	50.54%	32.59%	—	—	39.74%	36.23%	40.26%
Random sample with macro regressors								
Naïve	9%	16.85%	-0.20%	—	—	4.16%	4.66%	12.37%
OLR	26.97%	39.32%	18.23%	22.45%	2924	36.72%	20.51%	37.32%
RF	47.75%	55.61%	42.24%	—	—	58.99%	50.06%	55.80%
GB	48.88%	57.30%	43.65%	—	—	53.74%	47.57%	52.54%
Time-dependent sample without macro regressors								
Naïve	7.63%	12.70%	-1.57%	—	—	5.88%	5.53%	14.96%
OLR	23.73%	41.52%	15.61%	20.8%	3220	20.37%	20.41%	33.94%
RF	45.76%	51.69%	40.16%	—	—	52.49%	45.53%	47.02%
GB	40.11%	55.49%	33.75%	—	—	39.57%	38.10%	40.32%
Random sample without macro regressors								
Naïve	9%	16.85%	-0.20%	—	—	4.16%	4.66%	12.37%
OLR	25.28%	38.58%	16.23%	0.209	2964	27.30%	19.20%	34.39%
RF	50.56%	64.04%	45.33%	—	—	56.79%	52.83%	55.99%
GB	47.21%	53.04%	44.75%	—	—	53.17%	49.63%	54.01%

OLR: ordered logistic regression; RF: random forest; GB: gradient boosting; AIC: Akaike information criterion

H1 was partially confirmed. GB and RF outperformed OLR, as well as the naïve forecast, by all accuracy metrics. However, in several metrics, GB did not outperform RF. A plausible explanation lies in the differences between the ensemble formation strategies of the modeling techniques. In this unbalanced sample with observations from different countries over 11 years, the estimation error should be unpredictable, and the results of GB might converge with those of RF. Further research into this phenomenon is required.

H2 was confirmed. Random division of the data into the training and test sets provided higher model accuracy by all metrics except for the modified accuracy of OLR. The randomly sampled training and test datasets had similar class distributions, which resulted in more accurate forecasts. In contrast, dividing the data in a time-dependent manner exacerbated the existing rating distribution imbalances.

H3 was rejected. The addition of macroeconomic variables to the models did not improve but, on the contrary, reduced their prediction power. This result was confirmed by

an analysis of relative variable influence plots in the GB and RF models (Figure 3). A possible explanation could be that to maintain rating stability, international rating agencies use a through-the-cycle approach. This resulted in ratings that were nearly independent of cyclical changes and instead evaluated permanent components of credit risk. These results are inconsistent with the previous findings, such as those of [Karminsky \(2011\)](#). Therefore, further research is required.

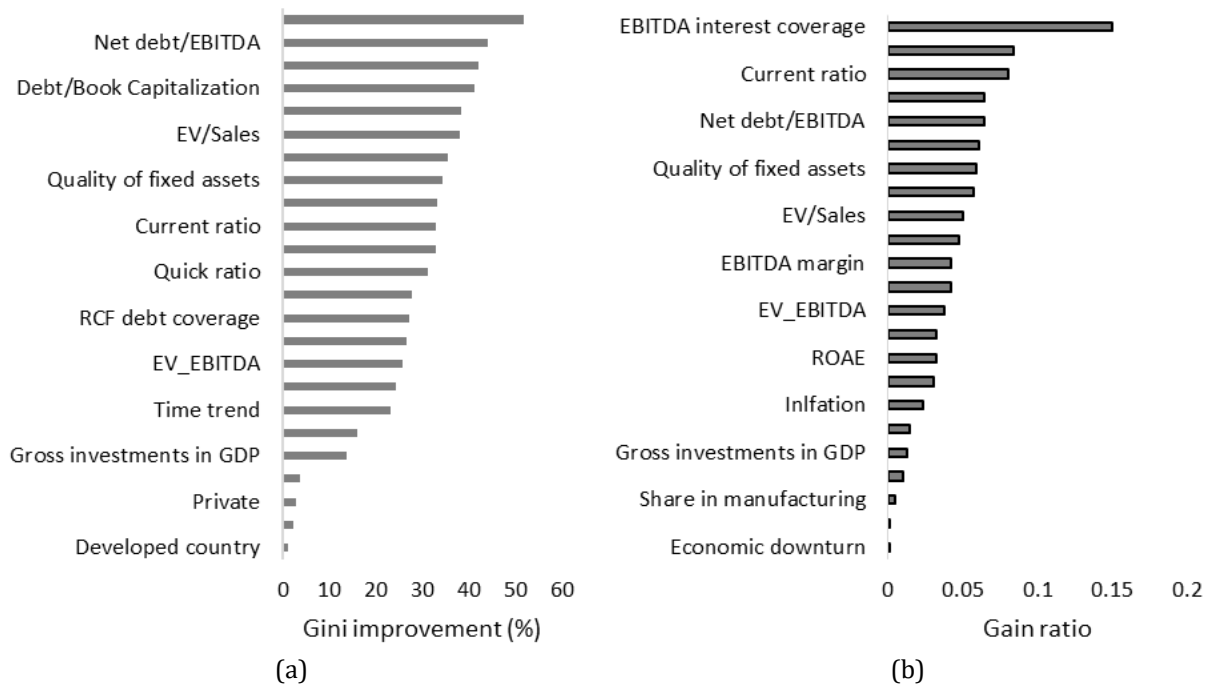


Figure 3 Contribution of macroeconomic variables into the models' accuracy: (a) Gini improvement by adding the variable in the model in random forest; and (b) error decrease by adding the variable in the model in gradient boosting. EBITDA: earnings before interest, taxes, depreciation, and amortization; enterprise value (EV); RCF: retained cash flow; GDP: gross domestic product; ROAE: return on average equity

H4 was rejected. The signs of several variables in the final specification of OLR did not match the expected signs (Table 4). This could be due to the elevated correlations (with pairwise cross-factor correlations above 0.3) between metrics with “wrong” signs. Further studies are required to build a better model specification.

Table 4 Signs obtained from the ordered logistic regression model estimation

Variable	Actual sign	Expected sign
EBITDA margin	-	-
Current ratio	+	-
Cash ratio	+	-
Debt/Book capitalization	+	+
Debt/Market capitalization	+	+
EBITDA interest coverage	-	-
Return on average equity	+	-
Interest paid	+	+
EV/EBITDA	-	-
EV/Sales	+	-
Net debt/EBITDA	+	+

Variable	Actual sign	Expected sign
Share in global manufacturing	–	–
Inflation	–	undefined
Control of corruption	–	–
Private company	+	+
Time trend	+	+

EBITDA: earnings before interest, taxes, depreciation, and amortization;
enterprise value (EV)

Based on our results, we recommend using ML techniques for reproducing or assessing the quality of global manufacturing companies' PCRs. This recommendation is based on their greater predictive power and discrimination ability, as well as the possibility of considering a large set of explanatory factors. However, these methods should be used with caution, given their low interpretability and susceptibility to overfitting. If the goal of reproducing a credit rating is to identify the exact impact of each explanatory factor on the rating, OLR should be preferred.

4. Conclusion

In this study, we compared the accuracy of OLR, RF, and GB in reproducing global manufacturing companies' PCRs. RF and GB outperformed OLR by a factor of almost 2. Random sampling yielded higher predictive power than time-dependent sampling. The inclusion of macroeconomic variables did not improve the models' predictive power. The predictive power of our models is consistent with the literature. We conclude that ML techniques can be effective in reproducing manufacturing companies' PCRs. Future research should examine how the addition of non-financial metrics can improve models' predictive power. Such metrics include indicators of market and operational performance, corporate governance and quality of management, and companies' intellectual capital. Future studies should also determine and compare sets of explanatory factors for reproducing credit ratings for diverse non-financial industries, such as oil and gas, metal and mining, chemical, automotive, and fast-moving consumer goods companies. Determining the best sets of explanatory factors for each industry will help improve the models' predictive power.

Acknowledgements

This research was supported by the Academic Excellence Project 5-100 proposed by Peter the Great St. Petersburg Polytechnic University.

References

- Abdi, H., Williams, L.J., 2010. Principal Component Analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, Volume 2(4), pp. 433–459
- Altman, E., Haldeman, R., Narayanan, P., 1977. ZETA™ Analysis. A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking & Finance*, Volume 1(1), pp. 29–54
- Amato, J., Furfine, C., 2004. Are Credit Ratings Procyclical?. *Journal of Banking & Finance*, Volume 28(11), pp. 2641–2677
- Basel Committee on Banking Supervision, 2017. *Basel III: Finalising Post-Crisis Reforms*. Bank for International Settlements. Available Online at <https://www.bis.org/bcbs/publ/d424.pdf>, Accessed on July 20, 2020

- Bhushan, S., Reddy, C., 2016. A Four-Level Linear Discriminant Analysis-based Service Selection in the Cloud Environment. *International Journal of Technology*, Volume 7(5), pp. 859–870
- Cuny, C., Dube, S., 2017. When Transparency Pays: The Moderating Effect of Disclosure Quality on Changes in the Cost of Debt. Hutchins Center, Working Paper No. 34
- Chopra, A., Bhilare, P., 2018. Application of Ensemble Models in Credit Scoring Models. *Business Perspective and Research*, Volume 6(2), pp. 129–141
- Demeshev, B., Tikhonova, A., 2014. Default Prediction for Russian Companies: Intersectoral Comparison. *Economic Journal of Higher School of Economics*, Volume 18(3), pp. 359–386
- Fachrurrazi, H.S., Munirwansyah, H., 2017. The Subcontractor Selection Practice using ANN-Multilayer. *International Journal of Technology*, Volume 8(4), pp. 761–772
- Hossin, M., Sulaiman, M.N., 2015. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining and Knowledge Management Process*, Volume 5(2), pp. 1–11
- Huang, Z., Chen, H., Hsu, C.J., Chen, W.H., Wu, S., 2004. Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study. *Decision Support Systems*, Volume 37(4), pp. 543–558
- Langohr, H., Langohr, P., 2008. *The Rating Agencies and Their Credit Ratings: What They Are, How They Work, and Why They are Relevant*. USA: John Wiley & Sons, Inc.
- Liao, Y., Loures, E., Deschamps, F., Ramos, L.F., 2017. Past, Present and Future of Industry 4.0 – A Systematic Literature Review and Research Agenda Proposal. *International Journal of Production Research*, Volume 55(12), pp. 3609–3639
- Lee, Y.C., 2007. Application of Support Vector Machines to Corporate Credit Rating Prediction. *Expert Systems with Applications*, Volume 33(1), pp. 67–74
- Kamstra, M., Kennedy, P., Suan, T., 2001. Combining Bond Rating Forecasts using Logit. *Financial Review*, Volume 36(2), pp. 75–96
- Karminsky, A., 2011. Corporate Rating Models for Emerging Markets. *Journal of Corporate Finance Research*, Volume 5(3), pp. 19–29
- Karminsky, A., Peresetsky, A., 2007. Rating Models of International Agencies. *Applied Econometrics*, Volume 1(5), pp. 3–19
- Karminsky, A., Peresetsky, A., 2009. Ratings as Measure of Financial Risk: Evolution, Function and Usage. *Journal of the New Economic Association*, Volumes 1–2, pp. 86–104
- Karminsky, A., Polozov, A., 2016. *Handbook of Ratings: Approaches to Ratings in the Economy, Sports, and Society*, Switzerland: Springer
- Kumar, K., Bhattacharya, S., 2006. Artificial Neural Network vs Linear Discriminant Analysis in Credit Ratings Forecast: A Comparative Study of Prediction Performances. *Review of Accounting and Finance*, Volume 5(3), pp. 216–227
- Moody's Investors Service, 2020. Manufacturing Methodology. Available Online at https://www.moody.com/research/Manufacturing-Methodology--PBC_1206079, Accessed on July 20, 2020
- Saitoh, F., 2016. Predictive Modeling of Corporate Credit Ratings using a Semi-Supervised Random Forest Regression. In: 2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Bali 2016, pp. 429–433
- Senaviratna, N.A.M.R., Cooray T.M.J.A., 2019. Diagnosing Multicollinearity of Logistic Regression Model. *Asian Journal of Probability and Statistics*, Volume 5(2), pp. 1–9
- Sermpinis, G., Tsoukas, S., Zhang, P., 2018. Modelling Market Implied Ratings using LASSO Variable Selection Techniques. *Journal of Empirical Finance*, Volume 48, pp. 19–35

- Tsai, C.F., Chen, M.L., 2010. Credit Rating by Hybrid Machine Learning Techniques. *Applied Soft Computing*, Volume 10(2), pp. 374–380
- Ye, Y., Liu, S., Li, J., 2008. A Multiclass Machine Learning Approach to Credit Rating Prediction. *In: 2008 International Symposiums on Information Processing, Moscow 2008*, pp. 57–61
- World Bank Open Data, 2020. Available Online at <https://data.worldbank.org/>, Accessed on July 20, 2020