

DOI: <https://doi.org/10.17323/j.jcfr.2073-0438.16.1.2022.99-112>

JEL classification: C23, G17, G23, G32



Comparative Analysis of the Predictive Power of Machine Learning Models for Forecasting the Credit Ratings of Machine-Building Companies

Sergei Grishunin

Candidate of Sciences (PhD), CFA, Senior Lecturer, Graduate School of Industrial Economics, Institute of Industrial Management, Economics and Trade,
Peter the Great St. Petersburg Polytechnic University, Saint Petersburg,
Russia, sergei.v.grishunin@gmail.com, [ORCID](#)

Alexandra Egorova

Manager, Deloitte and Touche, Moscow, Russia,
alxegorova@gmail.com, [ORCID](#)

Abstract

The purpose of this study is to compare the predictive power of different machine learning models to reproduce Moody's credit ratings assigned to machine-building companies. The study closes several gaps found in the literature related to the choice of explanatory variables and the formation of a data sample for modeling. The task to be solved is highly relevant. There is a growing need for high-precision and low-cost models for reproducing the credit ratings of machine-building companies (internal credit ratings). This is due to the ongoing growth of credit risks of companies in the industry, as well as the limited number of assigned public ratings to these companies from international rating agencies due to the high cost of the rating process. The study compares the predictive power of three machine learning models: ordered logistic regression, random forest, and gradient boosting. The sample of companies includes 109 machine-building enterprises from 18 countries between 2005 and 2016. The financial indicators of companies that correspond to Moody's industry methodology and the macroeconomic indicators of the companies' home countries are used as explanatory variables. The results show that artificial intelligence models have the greatest predictive ability among the models studied. The random forest model demonstrated a prediction accuracy of 50%, the gradient boosting model - 47%. Their predictive power is almost twice as high as the accuracy of ordered logistic regression (25%). In addition, the article tested two different ways of forming a sample: the random method and one that accounts for the time factor. The result showed that the use of random sampling increases the predictive power of the models. The incorporation of macroeconomic variables into the models does not improve their predictive power. The explanation is that rating agencies follow a "through the cycle" rating approach to ensure rating stability. The results of the study may be useful for researchers who are engaged in assessing the accuracy of empirical methods for modeling credit ratings, as well as banking industry practitioners who use such models directly to assess the creditworthiness of machine-building companies.

Keywords: credit ratings, internal credit ratings, machine-building companies, machine learning models, rating agencies

For citation: Grishunin, S., and Egorova A. Comparative Analysis of the Predictive Power of Machine Learning Models for Forecasting the Credit Ratings of Machine-Building Companies. *Journal of Corporate Finance Research*. 2022;16(1): 99-112. <https://doi.org/10.17323/j.jcfr.2073-0438.16.1.2022.99-112>

The journal is an open access journal which means that everybody can read, download, copy, distribute, print, search, or link to the full texts of these articles in accordance with CC Licence type: Attribution 4.0 International (CC BY 4.0 <http://creativecommons.org/licenses/by/4.0/>).

Introduction

In the past few years the fourth industrial revolution has fundamentally changed the business environment and business models of machine-building companies (MBC). It provides new opportunities for profit and increases company value in this industry, but exposes them to elevated risks. The dangers are as follows: 1) uncertainty in regard to key suppliers and delivery prices; 2) reduction of the product life cycle; 3) discontinuity of operations caused by technology breakdowns, information failures and outer interference; 4) shortage of qualified staff at all levels; 5) increased competition created by manufacturers from emerging markets, as well as by companies from other industries; 6) other internal and external risks [1]. Growing uncertainty, volatility and variability of the external and internal environment increase the probability of default of MBC. This makes relevant the task of constructing high-precision models of MBC credit risk assessment. Investors need these models to evaluate MBC creditworthiness within the planning horizon and the landscape of making decisions on provision of financing.

In order to assess MBC creditworthiness, investors use credit ratings (CR) assigned by expert international rating agencies, such as Moody's Investor Service, Fitch Ratings or Standard and Poor's [2]. They provide an opportunity to thoroughly examine MBC's financial and business profiles, evaluate their advantages and disadvantages and predict the likelihood of MBC's timely settlement of their financial obligations. CR also helps to compare the credit quality of companies from various countries and markets [3]. The credit rating is a kind of MBC's "seal of excellence". It enables MBC to appeal to more investors as well as to increase the amounts and periods of financing, reduce the cost of capital and gradually increase the probability of cooperating with investors when their credit profile is improved [4]. The high cost of assigning and maintaining a CR, as well as the demanding requirements of international rating agencies for the minimal company size and quality of corporate governance are among the drawbacks of a CR [3]. Therefore, the scope of a CR use is limited to large multi-industry manufacturers, mainly from developed markets. Thus, credit ratings do not cover small and medium-size MBC or firms from emerging markets because they lack the financial and organizational resources to maintain a CR. Another disadvantage of a CR is big update intervals, typically, one year long [4].

In order to eliminate these blind spots, investors evaluate internal credit ratings (ICR) of companies, including MBC. The approach, which has proved to be efficient, implies a reproduction of the missing credit ratings using empirical models based on public financial and non-financial company data [3]. The obtained ICR are unbiased and uncostly assessments of companies' creditworthiness. However, the predictive power of ICR (i.e. the ability to reproduce CR accurately) varies greatly depending on the models at the basis of the ICR [5]. In its turn, the literature review demonstrated that the majority of studies in this sphere use

companies from numerous industries (as a rule, from developed countries) as a sample, thus leaving out the specific nature of MBC's operations and special features of their work in developed markets. Some other drawbacks were also revealed: a small observation period in samples and inconsistency of explanatory variables in the models with the factors used by international rating agencies.

Our research fills the abovementioned gaps in literature. Its purpose is to 1) compare the predictive power of different machine learning models in order to reproduce Moody's credit ratings focused on MBC; and 2) to define the optimum model in terms of data availability, forecast accuracy and result interpretability. For modelling we selected the creditworthiness factors which explicitly examine the special aspects of MBC operations and correspond to Moody's credit rating methodologies. The MBC sample comprises companies from both developed and emerging markets. We have also verified whether the addition of macroeconomic factors enhances the accuracy of CR prediction, as demonstrated in literature [6]. We use the 2005–2016 period in this paper. Research results may be useful to theorists who evaluate the accuracy of empirical CR modelling methods and practitioners who use such models to assess MBC's creditworthiness.

Setting the Objective and Description of the Research Model

Literature review

There is a range of models aimed to assess and predict credit ratings. They differ in their assumptions. The majority of studies use linear regression, logistic regressions or the discriminant analysis method. These are standard approaches to credit rating modelling. Besides, some studies use neural networks or duration and hazard models to predict rating transitions.

Econometric Methods

Early studies [7] use the univariate parameter method to predict the probability of default. Later Altman [8] used linear discriminant analysis in his paper to predict credit quality. At the close of the XX century logit and probit models were first applied because they have a greater predictive power than the models that use the discriminant and quadratic discriminant analysis. Martin [9] and Ohlson [10] were the first ones to use logit regression to construct a model of bank bankruptcy probability. Empirical studies [11] revealed that ordered logistic regression models yield more results and have a greater predictive power than the least squares and discriminant analysis methods. The ordered logistic regression method is used in many new studies dedicated to business and economics issues [12–14]. This method is superior in defining credit ratings because of its ordered structure. Apart from that, it was noted that those methods had the greatest predictive power in comparison to linear regression, linear discriminant analysis, quadratic discriminant analysis and discriminant analysis of the mixture of distributions.

At present a lot of studies are dedicated to the use of the LAS-SO model [16] in order to search for the parameters that are most significant for the prediction of corporate credit rating.

Machine Learning Methods

The issue of assigning a credit rating may be considered a classification objective as well. In the XXI century machine learning methods which were used to forecast the probability of default and corporate credit quality have gained popularity. Machine learning models may be “trained” using the sample of ratings and corresponding data. For example, in neural networks training is defined as a search for weights in order to obtain the most accurate result [17]. However, the majority of such studies are conducted beyond the scope of economic analysis, as part of development and use of alternative methods in informatics.

Support Vector Machines (SVM) [18] were proposed as a method characterized by a great predictive power, however, its formation requires numerous financial and non-financial indicators. Apart from Support Vector Machines, classification trees [19–21] and neural networks [22–24] gained popularity in terms of rating prediction and probability of bankruptcy. Thus, in some studies Support Vector Machines and neural networks method demonstrate the same predictive accuracy of about 80% [25]. Comparison of the predictive power of the neural network model to linear discriminant analysis when forecasting Moody's ratings for different companies [26] showed that the use of a neural network delivers accuracy of 79%, which exceeds the result of discriminant analysis (33%).

Gradient boosting is another alternative method of credit rating forecasting. Paper [27] proves that gradient boosting outperforms the decision tree method from the viewpoint of the credit scoring models' predictive power. Another paper [28] notes that the gradient boosting algorithm demonstrates the greatest predictive power in the random forest, decision trees and neural networks models.

Each of the above methods of credit rating forecasting has its advantages and disadvantages. For instance, econometric methods are easy to use and interpret. However, these methods have low predictive power and amounts to 40–50% on average [11]. Apart from that, it is necessary to select data before using it in econometric methods. Machine learning models have a great predictive power, however, the majority of them are uninterpretable and may be subject to data overfitting [29].

Explanatory Variables

Literature defines three groups of factors that explain CR. The first category comprises financial ratios and financial

data [11]. The second category consists of corporate management and risk management factors [14; 30]. The third category includes macroeconomic factors. Studies [5; 13] reveal that in case of CR prediction for financial organizations, the introduction of macroeconomic variables in the models significantly improves the quality of model fitting and enhances its predictive power. However, when CRs were modelled for non-financial companies, some of the macroeconomic indicators (i.e., GDP growth) turned out to be insignificant or their signs failed to meet expectations [6]. A major issue in the selection of variables for analysis is multicollinearity between dependent variables [13], therefore, the choice of the model specification and variable selection assume a great significance.

Absence of focus on a certain industry (in our case it's machine building) is a gap in CR modelling because in the majority of studies CR modelling is performed using a sample of companies from various industries (in most cases the industries are identified by introducing dummy variables into the models). This makes it impossible to clearly define the explanatory variables characteristic of a certain industry. Also, companies from certain countries (Taiwan, USA, Korea, China) are examined, preventing one from generalizing the results of modelling of a wide range of such companies. Besides, studies are limited by the following: 1) a short time interval applied in the samples; 2) use of explanatory variables other than the ones utilized by rating agencies. The purpose of this paper is to fill the above gaps in studies.

Research Methodology

We have built an MBC credit quality assessment model that emulates Moody's rating. For this purpose, we applied the following methods: ordered logistic regression (OLR), random forest (RF) and gradient boosting (GB).

For an MBC, the model predicting CR may be expressed as follows:

$$Y_t = f(X_{1t} \dots X_{nt}), \quad (1)$$

where Y_t is a dependent variable, MBC's credit rating assigned by Moody's at the time t . The agency assigned a rating expressed as a literal notation in accordance with its own scale [34]. We transferred the rating to a qualitative scale, where whole numbers correspond to literal notations of the rating, they are presented in ascending order: the lower the rating, the bigger the number (Table 1);

X_{1t}, \dots, X_{nt} is a set of n explanatory variables defined at the time t .

$Y_t = \tau_t$ is a numerical value of rating from Table 1.

Table 1. Numerical scale of dependent variable (transfer of the Moody's rating literal notation into an order scale)

Moody's rating	AAA	Aa1	Aa2	Aa3	A1	A2	A3	Baa1	Baa2
Numerical rating value (τ)	1	2	3	4	5	6	7	8	9
Moody's rating	Baa3	Ba1	Ba2	Ba3	B1	B2	B3	Caa1–Caa3	C–Ca
Numerical rating value (τ)	10	11	12	13	14	15	16	17	18

Source: [34].

Ordered logistic regression. As long as the dependent variable Y_i is an ordered one and accepts k values of the rating levels $k \in [1; 18]$, we applied ordered logistic regression (OLR) [6]. We introduce the latent variable z related to the rating value and dependent variables as follows:

$$\begin{cases} z_i = 1, \text{ if } z_i = x_i' \theta + e_i \leq \tau_1; \\ z_i = r, \text{ if } \tau_{r-1} \leq z_i = x_i' \theta + e_i \leq \tau_r, 2 \leq r \leq k-1, \\ z_i = k, \text{ if } z_i = x_i' \theta + e_i \geq \tau_{k-1}, \end{cases} \quad (2)$$

where i is the observation sequential number; τ_r are threshold values of the rating level cut-off; e_i – errors which are supposed to be estimated, normally distributed and have a zero mathematical expectation.

By using this model we expect to obtain an assessment of the coefficient vector θ , as well as a set of threshold values of cut-offs for each rating level (τ_1, τ_{k-1}) by applying the maximum likelihood method for the system of the following equations:

$$\begin{cases} P(y_1 = 0) = F(\tau_1 - x_i' \theta); \\ P(y_i = r) = F(\tau_{r-1} - x_i' \theta) - F(\tau_r - x_i' \theta), 2 \leq r \leq k-1; \\ P(y_i = k) = F(\tau_{k-1} - x_i' \theta), \end{cases} \quad (3)$$

where $F(x)$ is a logistic function [6]; $P(y_i = r)$ is the probability of assigning MBC with the set of values x_i to the rating grade r .

In equation (3) standard errors are specified in the White-Huber form, thus reducing their heteroscedasticity. After obtaining θ and τ scores, predictive probabilities \hat{P}_j from equation (3) are calculated. MBC is assigned the rating j , for which the value of \hat{P}_j is the biggest. We will use

McFadden R^2 criterion [6] as a measure of quality of the model approximation to actual data, which is a variation of criterion R^2 widely used in econometrics. Other indicators presented in section 2 will also be quality criteria.

Random forest. Unlike OLR, random forest (RF) is a machine learning algorithm, which results in building of a multitude of decision trees models during training [32]. Output data is obtained on the basis of voting results of individual tree classes for the classification model and as an average response (averaging) – for the regression model [35]. The result of the rating forecasting objective is an average value of multiple regression trees

$$Y = f(x_1, \dots, x_n) = \frac{1}{G} \sum_{g=1}^G h(x_i; T_g), \quad (4)$$

where G is the number of trees; h is the regression tree function obtained at the input T_g .

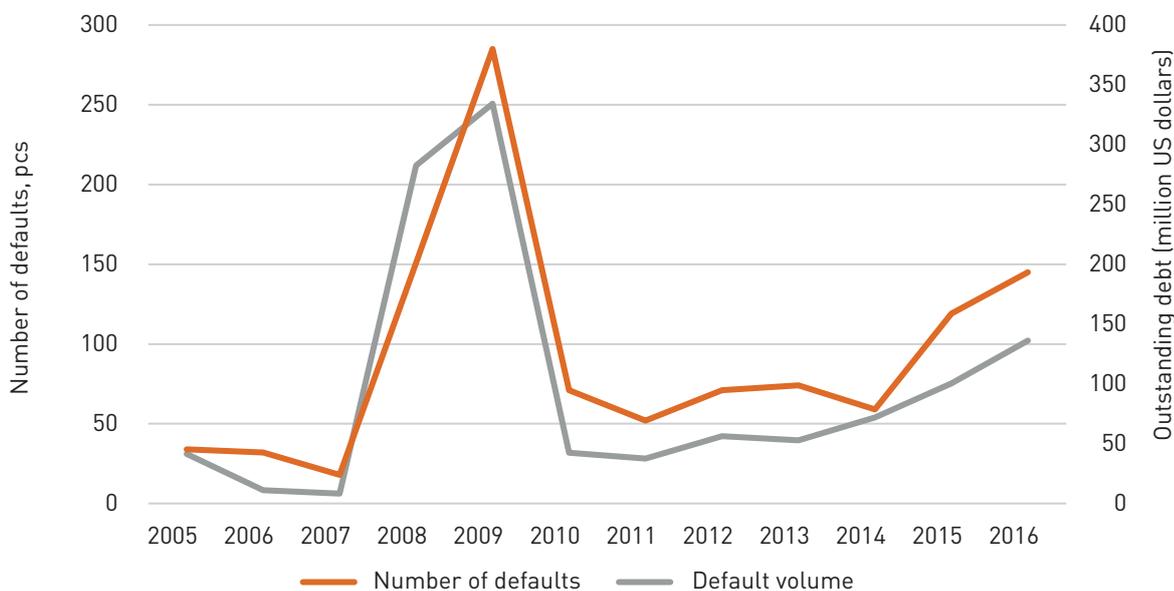
Gradient boosting (GB). This method is also an ensemble learning method, but it applies another ensemble formation strategy. The algorithm trains weak models consistently, in many iterations, taking into consideration the error of the whole ensemble defined at the moment in order to provide a more accurate assessment of the corporate credit rating. A gradient descent is used for optimization [36]

$$y = f(x_1, \dots, x_n, \hat{\theta}); \hat{\theta} = \underset{\theta}{\operatorname{argmin}} E_x [E_y(\varphi[y, f(x, \theta)]) | X], \quad (5)$$

where θ – parameters for evaluation; $\varphi(y, f(x))$ – the target function.

Data and Explanatory Variables

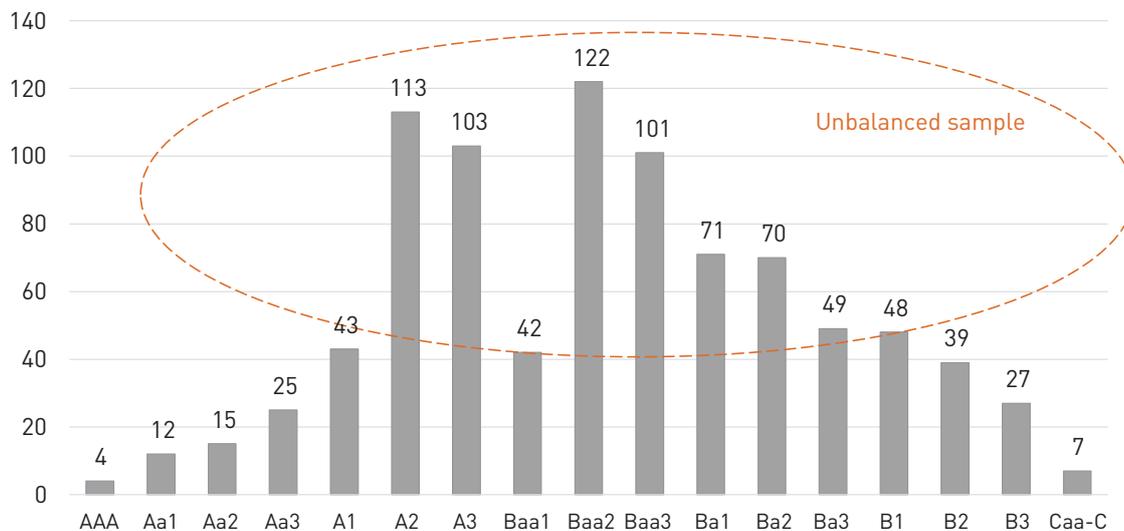
Figure 1. Credit cycle in financial markets in 2005–2016



Source: [34].

However, the sample is not balanced according to rating categories (Figure 2).

Figure 2. Distribution of MBC credit ratings in the sample



Source: [34].

In this paper we use financial and non-financial data of 109 companies engaged in different machine-building sectors of 18 countries. We present observations for each company for the period of 2005 to 2016. There are 891 observations in total. The data comprises observations of MBCs that manufacture machines and equipment for metalworking and mineral industry, power generating, medical industry, agriculture and construction industry. Motor manufacturers and manufacturers of machines and equipment for aerospace and defense industry are left out of the sample because Moody's uses another set of factors for these companies to explain their creditworthiness, which is described in separate methodologies.

The sample consists of 62 companies from the USA, 13 – from Japan, 8 – from Germany, 3 – from Sweden, 3 – from Great Britain, 3 – from France, 2 – from Finland, 2 – from Ireland and China each. Canada, Greece, Netherlands, Peru, Russia, Turkey, Mexico and Indonesia are represent-

ed by 1 MBC each. The temporal pattern of the dataset covered the entire credit cycle (Figure 1).

It is related to MBCs' high business risks caused by the industry's significant capital intensity, cyclical nature of demand, dependence on large customers, duration of the manufacturing cycle. These risks have a significant positive correlation with MBCs' credit risks. This limits the MBCs' capability to get high ratings (the average machine-building companies' rating assigned by Moody's is Baa3) [34].

Explanatory variables comprise financial indicators that represent MBCs' performance results, as well as macroeconomic variables in their countries of business. We used Moody's methodology for manufacturing companies [34] to make a list of financial indicators. Financial indicators and ratings data were obtained from Thomson Reuters Eikon, macroeconomic variable data – from the World Bank network. Table 2 contains the list of variables, their descriptive statistics and expected signs of influence on the rating.

Table 2. List of explanatory variables

Explanatory variable	Description	UOM	Formula	Expected sign	Mean value	Standard deviation
Indicators that define the nature of a company's business						
Share of gross investments in GDP	Amount of gross investments in fixed capital	%	Gross investments /GDP in the country of operation	“-”	20.8	20.4
Share in global manufacturing	Share in the global industry	%	Company's proceeds/ share of the added cost in the industry	“-”	0.14	0.21
Time trend	Time indicator	year	Number of years since the first observation	“+”	5	3

Explanatory variable	Description	UOM	Formula	Expected sign	Mean value	Standard deviation
Economic downturn flag	Indicator of economic depression	1/0	Dummy variable equals 1 if there is an economic downturn in the year of observation, 0 – otherwise	“+”	—	—
Private company flag	Indicator of a private company	1/0	Dummy variable equals 1 if the company is private, 0 – if it is government-owned	“+”	—	—
Resident in developed country flag	Indicator of operations in developed economies	1/0	Dummy variable equals 1 if the company operates in developed markets, 0 – otherwise	“–”	—	—
Quality of fixed assets	Quality of assets	%	Amortization/Assets	“+”	8.1	3.5
Market value to sales multiple	Ratio of the market value to proceeds	multiplier	$\frac{EV}{Sales} = \frac{MC + D - CC}{Annual\ Revenue}$, where <i>MC</i> – market capitalization; <i>D</i> – liabilities; <i>CC</i> – cash and cash equivalents	“–”	1.76	1.54
Market value to <i>EBITDA</i> multiple	Ratio of the market value to <i>EBITDA</i>	multiplier	$\frac{EV}{EBITDA} = \frac{MC + D - CC}{EBIT + DA + FI}$, where <i>EBITDA</i> – earnings before interest, taxes, depreciation and amortization; <i>EBIT</i> – earnings before interest and taxes; <i>DA</i> – depreciation and amortization; <i>FI</i> – other financial income	“–”	7.3	5.8
Interest paid	Interest which has been paid	multiplier	Company’s annual interest costs	“+”	4.25	1.16
Profit Indicators						
Return on average equity (<i>ROAE</i>)	Return on average equity	%	$ROAE = \frac{NPATBUI}{Average\ Equity} \cdot 100\%$, where <i>NPATBUI</i> – Net Profit After Taxes Before Unusual Items	“–”	13.7	30.4
<i>EBITDA</i> margin	Cost-effectiveness of <i>EBITDA</i>	%	$EBITDA\ margin = \frac{EBITDA}{Revenue} \cdot 100\%$	“–”	15.0	6.0
Indicators of Debt						
<i>Net debt/EBITDA</i>	Debt load ratio	multiplier	$\frac{Net\ debt}{EBITDA} = \frac{Debt - CC}{EBITDA}$	“–”	2.9	4.5

Explanatory variable	Description	UOM	Formula	Expected sign	Mean value	Standard deviation
Debt/Book Capitalization (BC)	Ratio between liabilities and book value of capitalization	%	$\frac{Debt}{BC} = \frac{Debt}{Book\ Value\ of\ Equity} \cdot 100\%$	“+”	61.8	12.7
Debt/Market Capitalization (MC)	Ratio between liabilities and company's market capitalization	%	$\frac{Debt}{MC} \cdot 100\%$	“+”	27.32	15.3
Cash ratio	Cash ratio	%	$Cash\ ratio = \frac{CC}{Debt} \cdot 100\%$	“-”	47.4	79.5
Retained cash flow (RCF) to net debt	Ratio of retained cash flow to net liabilities	%	$\frac{RCF}{Net\ debt} = \frac{CFO - \Delta WC - Div}{Debt - CC} \cdot 100\%$, where <i>CFO</i> – cash flow from operations; <i>ΔWC</i> – changes in working capital; <i>Div</i> – paid dividends;	“-”	75.2	23.2
Available RCF debt coverage	Retained cash flow available for settlement of debt	%	$\frac{Available\ RCF}{Debt} = \frac{RCF - Capex}{Debt} \cdot 100\%$, where <i>Capex</i> – capital expenditure	“-”	16.4	22.0
EBITDA interest coverage	Ratio of EBITDA coverage	multiplier	$\frac{(EBITDA - Capex)}{Interest}$, where <i>Interest</i> is paid interest	“-”	8.6	11.5
Liquidity Indicators						
Current ratio (CR)	Current liquidity ratio	multiplier	$CR = \frac{Current\ Assets}{Current\ Liabilities}$	“-”	1.9	0.7
Quick ratio (QR)	Acid test ratio	multiplier	$QR = \frac{CC + AR}{Current\ liabilities}$, where <i>AR</i> – accounts receivable	“-”	1.1	0.5
Macroeconomic Variables						
Real GDP growth	GDP growth rate	%	Annual growth rate of real GDP in the country of operations	“-““	1.6	2.1
Inflation	Inflation	%	Annual consumer price index	“?”	1.7	1.4

Explanatory variable	Description	UOM	Formula	Expected sign	Mean value	Standard deviation
Rule of law	Supremacy of law	multiplier	World Bank Index (WB)*, which measures efficiency of the legislative system, crime rate and citizens' attitude to crime in the country of business	"_"	—	—
Government effectiveness	Governmental authorities' efficiency	multiplier	WB index, which measures the quality of internal state policy, confidence in the government, the quality of the government mechanism operation in the country of business	"_"	—	—
Control of corruption	Corruption	multiplier	WB index, which measures perception of corruption in the society, existence of corruption at a high political level, influence of corruption on economic development in the country of business	"_"	—	—

Note. Numerical values of dependent variable scores are adjusted in such a way that a bigger value corresponds to the lowest score. Consequently, a positive sign denotes a negative influence of the explanatory variable on the dependent variable and vice versa.

* The methodology of World Bank's corporate governance indicators is described. URL: <http://info.worldbank.org/governance/wgi/>

Source: developed by the authors.

Data Preparation

We built a correlation matrix and excluded the most correlated variables (with paired correlation coefficients exceeding 0.8) in order to solve the multicollinearity problem in the OLR model. For other variables we evaluated the variance inflation factors (VIF) [37] and eliminated all variables with the VIF exceeding 5 from the sample. In order to evaluate the predictive power of explanatory variables, we also applied principal component analysis (PCA) [38]. When modelling ratings using machine learning methods, we applied the entire set of independent variables with no regard for the abovementioned selection. Machine learning methods are not susceptible to multicollinearity problem, while a large set of variables in ML allows to find the optimum combination of factors. In order to build models, in this paper we used the data not included in the set intended for verification of model quality (out of sample) at the ratio of 70% (training set) and 30% (test set).

Research Hypotheses

Hypothesis H1. Use of the gradient boosting model will provide an opportunity to get the greatest predictive power of the rating model. In other words, this model will demonstrate the greatest probability of concordance of the predicted and observed rating ($P(|\Delta| = 0)$). Consequently, the random forest model will be the second in predictive accuracy after gradient boosting. OLR will have the lowest predictive power among the three considered models. This corresponds with the evidence presented in paper [27]. A

nother reason against the high predictive power of the OLR model is that coefficients are assessed using the maximum likelihood function, and as long as the sample is unbalanced its results may be biased towards the most frequent rating values.

Hypothesis H2. Random data separation into the training and test samples will provide a greater predictive power for the model than data separation, which takes into consideration the time factor where the training set (70% of the sample) comprises data on the earliest observations and the test sample (30% of the sample) consists of the data on new observations. As long as the sample is unbalanced, we presume that a random separation into the training and test samples may provide a more accurate rating prediction.

Hypothesis H3. Addition of macroeconomic variables to the model will improve its predictive power. This is consistent with the data from [5; 31] which demonstrated that macroeconomic variables were statistically significant and their addition to the model enhanced its predictive power. In order to validate this hypothesis, we evaluated specifications of models with macroeconomic explanatory variables and without them.

Hypothesis H4. The gradient boosting model has the lowest probability of deviation of the predicted rating from the observed one by more than one step ($P(|\Delta| \geq 1)$). Among the considered models OLR will demonstrate the highest probability of deviation by more than one step. This corresponds to the evidence presented in the paper [27].

The smaller the dispersion of deviations of the predicted rating from the observed one, the ampler the possibilities of using the ICR model in order to assess the level of interest rates an MBC can expect to receive. It is related to the fact that interest rates may change significantly along with the rating change of more than one step [6].

Results and Discussion

Table 3 presents the results of forecasting MBC credit ratings by applying the abovementioned models. For the purpose of comparability, we submit the results of credit rating prediction using the “naive model”, i.e. a randomly obtained value of an MBC credit rating using a random number generator. In order to evaluate the predictive power, we applied multiclass classification models assessment metrics [39]. The predictive power metric (Accuracy) evaluates the correlation between the correct forecasts of

the rating and the general number of assessed ratings. The modified accuracy evaluates the correlation between the number of forecasts with the maximum error of one rating and the general number of observations. The completeness metric (Recall) evaluates the model’s capability to select the correct rating, and the Precision metric measures the positive results defined accurately from the total number of predicted results in the positive grade and assesses the model capability to distinguish a correct rating from other ratings. The F1 Score metric evaluates the harmonic mean value of predictive accuracy. The Kappa Accuracy metric indicates the ratio of the difference between the probability of the correct model classification and the probability of a random correct classification to the probability of a random wrong classification. Finally, the Akaike information criterion (AIC) indicates a relative order of the compared models: the smaller the indicator, the better the model from the point of view of its predictive power.

Table 3. Results of the models’ evaluation

The model that accounts for the time factor (70%/30%) and macrovariables								
Model	Accuracy, %	Modified Accuracy, %	Kappa Accuracy, %	McFadden R ²	AIC	Precision, %	Recall, %	F1 Score, %
Random forecast	7.63	12.70	-1.57	-	-	5.88	5.53	14.96
OLR	22.88	41.52	14.92	22.33	3174.00	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	36.230	40.26
The model that does not account for the time factor, but macrovariables								
Random forecast	9.00	16.85	-0.20	-	-	4.16	4.66	12.37
OLR	26.97	39.32	18.23	22.45	2924.00	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
The model accounts for the time factor (70%/30%), but does not account for macrovariables								
Random forecast	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
The model that does not account for the time factor or macrovariables								
Random forecast	9.00	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

H1 was partially confirmed. The GB and RF models demonstrated a higher quality than the OLR model by all accu-

racy indicators. Apart from that, all models significantly surpassed the random (naïve) forecast. However, the GB model was not better than the RF model in terms of several accuracy indicators. It may be due to the fact that when an ensemble is formed, each model uses different techniques (see section 1.2). In our unbalanced sample with the observations from different countries over 11 years the expected model error should be unpredictable and the GB model should agree with the RF model results. However, further research is necessary to analyze the obtained differences.

H2 was confirmed. A random division of data into the training and test samples ensured higher model accuracy (according to all indicators except the Modified Accuracy metric for OLR). A random division into the training and test samples had a similar distribution into rating grades, resulting in more accurate forecasts. On the contrary, separation of data on the basis of the time factor increased the imbalance in the rating distribution by scores which had been initially present in the sample (Figures 3 and 4).

H3 was not confirmed. Addition of macroeconomic variables did not enhance the predictive power of the models.

On the contrary, it made the results worse. This conclusion was confirmed by analysis of diagrams of variable information significance in the GB and RF models (Figures 5 and 6). This may be due to the fact that international rating agencies trying to provide consistency of rating scores used the “skip-cycle” approach and evaluated the constant component of MBC’s credit risk. However, as long as our conclusion disagrees with conclusions of other research papers [5; 31], it is necessary to study the obtained result further.

H4 was confirmed partially. In the GB model, modified accuracy is the highest indicator in all model specifications except for the model that does not account for the time factor or macrovariables. In its turn, in the OLR model the modified accuracy indicator is the lowest one in all model specifications. Analysis of obtained differences in modified accuracy for the GB and RF models when applying various sample creation methods requires further research. Nevertheless, in our opinion, the gradient boosting model is more promising for building the ICR model in order to evaluate the level of interest rates an MBC may count on.

Figure 3. Distribution of predicted ratings averaged among the models according to levels for the sample with regard to the time factor (leaving out macrovariables)

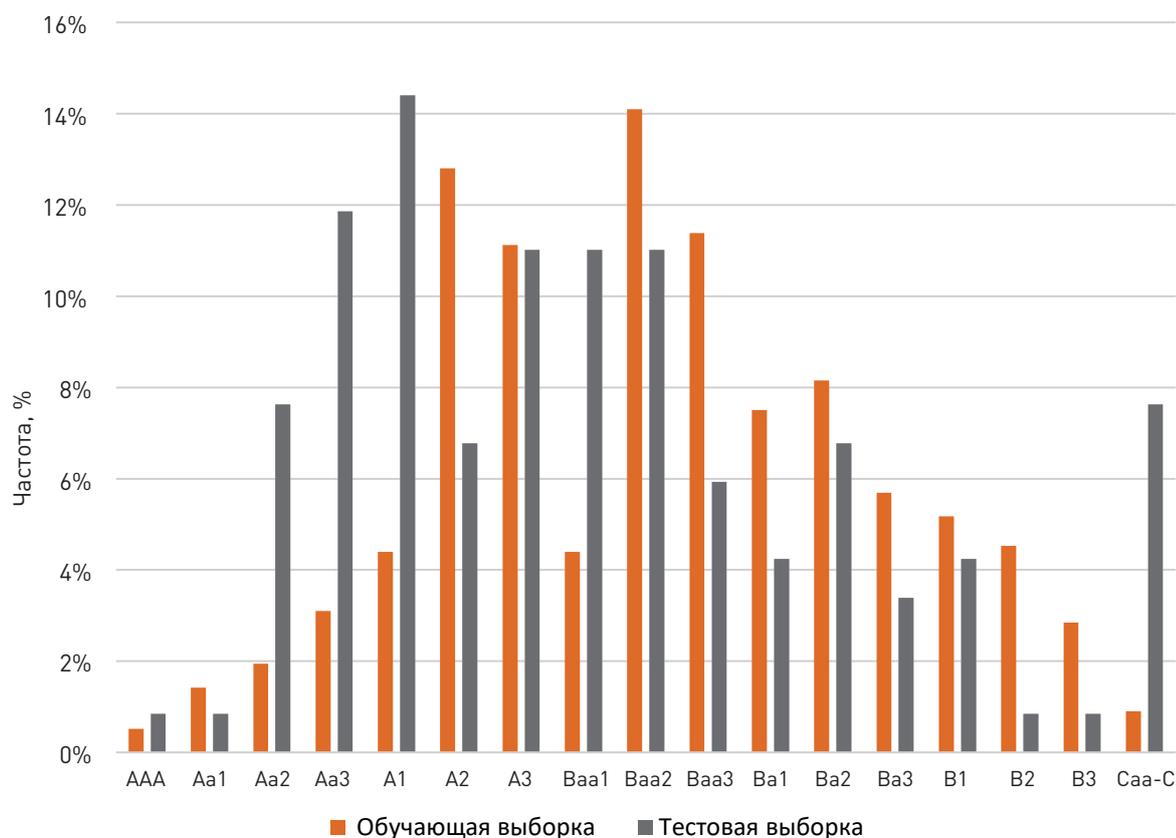


Figure 4. Distribution of predicted ratings averaged among the models according to levels for the random sample that accounted for the time factor (leaving out macrovariables)

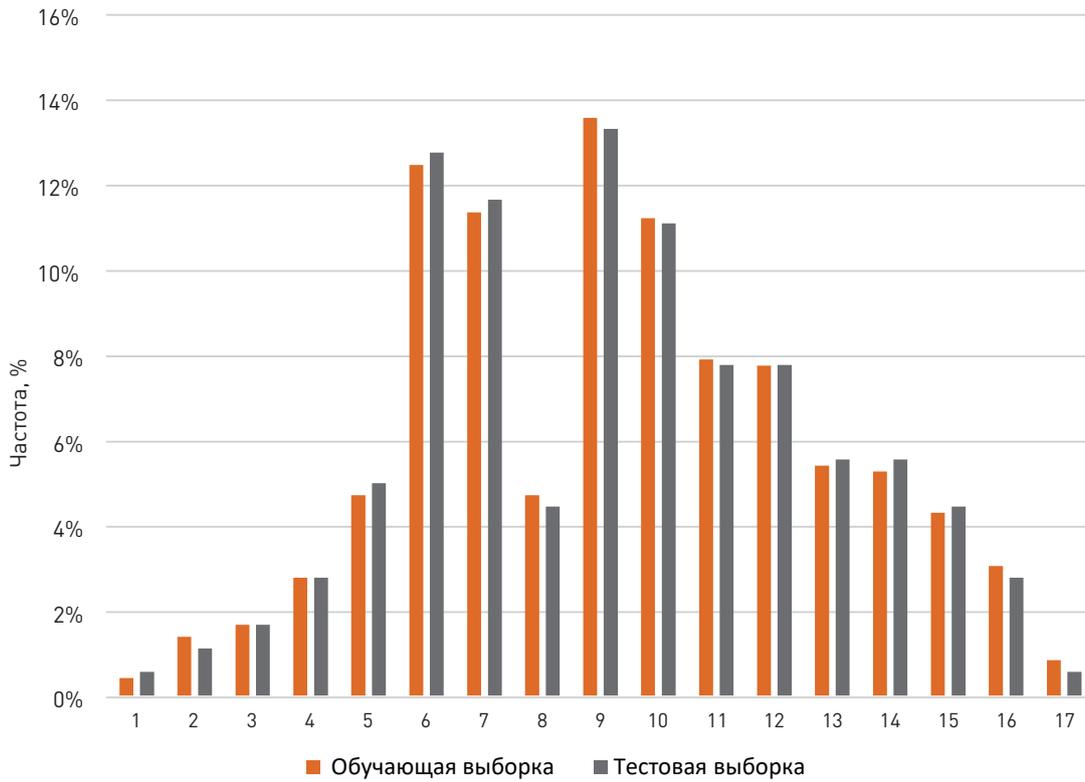
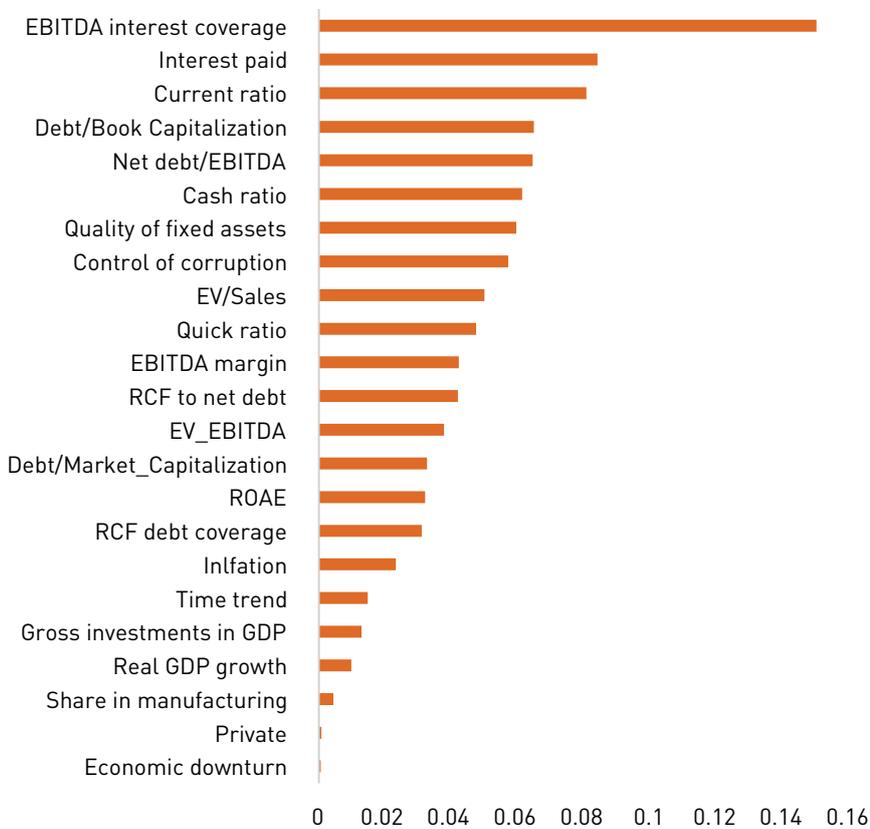
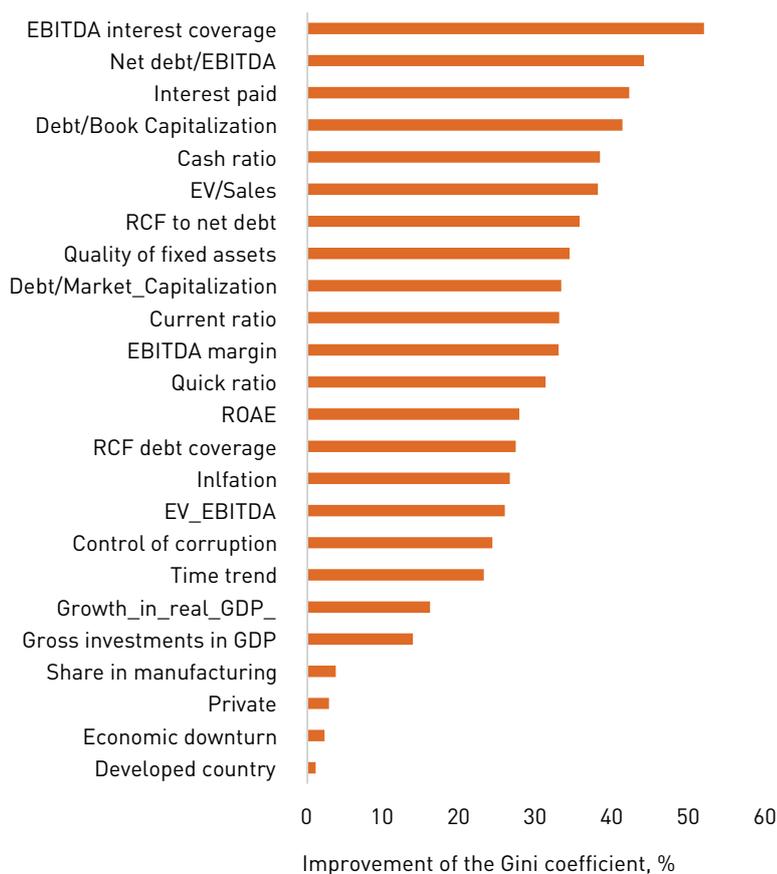


Figure 5. Information significance of explanatory variables in the GB model



Informaton criterion of increase (Information Gain Ratio) for the GB model

Source: [36].

Figure 6. Influence of individual explanatory variables on the Gini coefficient increase in the random forest model (RF)

Conclusion

In this paper we compared the predictive power of empirical models of logistic regression and machine learning models for modelling the internal credit ratings of machine-building companies. Random forest and gradient boosting were used as machine learning models. The objective is of relevance because, on the one hand, MBCs' credit risks are still increasing and, on the other hand, just a few MBCs have a public credit rating. The paper filled the gaps in literature in the following ways: 1) use of explanatory indicators that take into consideration the specific character of the machine-building industry to the greatest extent; 2) use of the sample for a significant period of time that covers the whole credit cycle; 3) adding companies from the developed and emerging economies to the sample. The results showed that the predictive power of machine learning models is almost twice as high as the predictive power of ordered logistic regression and the share of predicted ratings, which deviate from the actual ones by more than one step is low. Therefore, use of machine learning models may have a wide practical application for building internal credit ratings of machine-building companies. Apart from that, we've discovered that a random division into the training and test samples enhanced the models' predictive power when compared to a division according to the time factor.

However, we failed to prove that addition of macroeconomic indicators to the model as explanatory variables enhances its predictive power. Therefore, in future studies it is necessary to perform additional testing of the effect of adding macroeconomic factors. Another line of research is the evaluation of the influence produced by the addition of non-financial indicators to model specification on its predictive power. The non-financial factors comprise the factors which define MBCs' competitive advantages in the target markets, operational performance indicators, knowledge capital efficiency indicators and MBC corporate governance efficiency indicators. Finally, a separate line of research may be represented by comparison of various sets of explanatory variables in order to improve the predictive power of CR assessment models from different industries, such as: oil and gas industry, metalworking and mineral industry, chemical industry, automobile construction etc.

Acknowledgements

The paper was written as a part of the work performed by the Group on Research of the School of Finance of the Faculty of Economic Sciences of HSE Banking Sector Innovations, its Financial Soundness and Prudential Regulation. The authors are grateful to Stepan Barkhatov, Elina Agaeva and Vladimir Lozovoy, students of the Faculty of Economic Sciences of HSE for help in collecting data and active involvement in building models and analyzing results.

References

1. Liao Y., Loures, E., Deschamps, F., Ramos, L.F. Past, Present and Future of Industry 4.0 – a Systematic Literature Review and Research Agenda Proposal. *International Journal of Production Research*. 2017; 55 (12)
2. Karminsky A.M., Peresetsky A.A. Rejtingi kak mera finansovyh riskov. Evolyuciya, naznachenie, primenenie. *Zhurnal Novoj ekonomicheskoy asociacii*. 2009; 1-2
3. Karminsky A.M., Polozov A.A. Enciklopediya rejtingov: ekonomika, obshchestvo, sport. Forum; 2016.
4. Langohr H., Langohr P. *The Rating Agencies and their Credit Ratings: What They Are, How They Work and Why They Are Relevant*, John Wiley & Sons, Inc., Hoboken, New Jersey; 2008.
5. Karminsky, A.M. Peresetsky A.A. Modeli rejtingov mezhdunarodnyh agentstv. *Prikladnaya ekonometrika*. 2007; 1(5)
6. Karminsky, A. M. Kreditnye rejtingi i ih modelirovanie. *Izd. dom Vysshej shkoly ekonomiki*; 2015
7. Beaver, W. Financial Ratios as Predictors of Failure. *Journal of Accounting Research*. 1966; 4
8. Altman E.I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*. 1968; 23
9. Martin, D. Early Warning of Bank Failure: A Logit Regression Approach. *Journal of Banking and Finance*. 1977; 1, 249-276
10. Ohlson, J. A. Financial Ratios and Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*. 1980; 18, 109-131
11. Ederington, L. Classification models and bond ratings. *The financial review*. 1985; 20
12. Blume, M., Lim F., MacKinlay A. C. The declining quality of US corporate debt: Myth or reality? *Journal of Finance*. 1998; 53
13. Amato, J., Furfine C. Are credit ratings procyclical? *Journal of Banking & Finance*. 2004; 28
14. Ashbaugh-Skaife, H., Collins D., LaFond R. The Effects of Corporate Governance on Firms Credit Ratings. *Journal of Accounting and Economics*. 2006; 42
15. Demeshev B. B., Tihonova A. S. Dinamika prognoznoj sily modelej bankrotstva dlya srednih i malyh rossijskih kompanij optovoj i roznicnoj trgovli. *Korporativnye finansy*. 2014. T. 31. № 3. S. 4-22
16. Sermpinis G. Tsoukas S., Zhang P. Modelling market implied ratings using LASSO variable selection techniques. *Journal of Empirical Finance* 2018; 48
17. Kwon Y., Han I., Lee K. Ordinal pairwise partitioning (OPP) approach to neural networks training in bond rating. *Intelligent Systems in Accounting, Finance & Management*. 1997; 6: 23–40
18. Bellotti T., Crook J., Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications*. 2009; 36 (2), p. 3302-3308
19. Davis, R. H., Edelman, D. B., & Gamberman, A. J. Machine-learning algorithms for credit-card applications. *IMA Journal of Management Mathematics*. 1992; 4(1), 43–51
20. Zhou, S. R., & Zhang, D. Y. A nearly neutral model of biodiversity. *Ecology*. 2008; 89(1), 248–258
21. Frydman, H., Altman, E. I., & Kao, D. L. Introducing recursive partitioning for financial classification: The case of financial distress. *The Journal of Finance*. 1985; 40(1), 269–291
22. Jensen, H. L. Using neural networks for credit scoring. *Managerial Finance*. 1992; 18(6), 15–26
23. West, D. Neural network credit scoring models. *Computers & Operations Research*. 2000; 27(1), 1131–1152
24. West, D., Dellana, S., & Qian, J. X. Neural network ensemble strategies for financial decision applications. *Computers & Operations Research*. 2005; 32(10), 2543–2559
25. Huang Z., Chen H., Hsu C., Chen W., Wu S. Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems*. 2004; 37
26. Kumar, K., Bhattacharya, S. Artificial neural network vs linear discriminant analysis in credit ratings forecast: a comparative study of prediction performances. *Review of Accounting and Finance*. 2006; 5, 216–227
27. Chopra F., Bhilare P. Application of Ensemble Models in Credit Scoring Models. *Business Perspectives and Research*. 2018; 6 (4)
28. Wang, G., & Ma, J. Study of corporate credit risk prediction based on integrating boosting and random subspace. *Expert Systems with Applications*. 2011; 38(4), 13871–13878
29. Balios, D., Thomadakis, S., Tsipouri, L. Credit rating model development: An ordered analysis based on accounting data. *Research in International Business and Finance*. 2016

30. Bhojraj, S., P. Sengupta. Effect of Corporate Governance on Bond Ratings and Yields: The Role of Institutional Investors and Outside Directors. *Journal of Business*. 2003; 76(3), 455 – 475
31. Karminsky A. M. Metodicheskie voprosy postroeniya konstruktora dinamicheskikh rejtingov. *Vestnik mashinostroeniya*. 2008
32. Saitoh, F. Predictive modeling of corporate credit ratings using a semi-supervised random forest regression. *IEEE International Conference on Industrial Engineering and Engineering Management*. 2016; 429-433
33. Grilli, L., Rampichini, C. Ordered Logit Model. *Encyclopedia of Quality of Life and Well-Being Research*. 2014; p. 4510-4513
34. Internet-resurs www.moodys.com
35. Biau, G. Analysis of a Random Forests Model. *The Journal of Machine Learning Research*. 2012; 98888, 1063–1095
36. Natekin A., Knoll A. Gradient Boosting Machines. *Frontiers in Neuroinformatics*. 2013; 7 (21)
37. Senaviratna, N. A. M. R., Cooray A., T. M. J. Diagnosing Multicollinearity of Logistic Regression Model. *Asian Journal of Probability and Statistics*. 2019; 5(2), 1-9
38. Abdi, H. and Williams, L.J. Principal Component Analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*. 2010; 2, 433-459
39. Hossin M. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*. 2015; 5(2):01-11

Contribution of the authors: the authors contributed equally to this article.

The authors declare no conflicts of interests.

The article **was submitted** 16.01.2022; **approved after reviewing** 18.02.2022; **accepted for publication** 20.03.2022.