DOI 10.15826/vestnik.2024.23.1.009 Original Paper

Default Prediction for Russian Food Service Firms: Contribution of Non-Financial Factors and Machine Learning

Egor O. Bukharin[1] ✉, *Sofia I. Mangileva*[2], *Vladislav V. Afanasev*[1]

[1]*Higher School of Economics, Saint-Petersburg, Russia*

[2]*Yakov & Partners, Moscow, Russia*

✉ *egorbukharin01@gmail.com*

Abstract. The food service industry's instability due to COVID-19 and sanctions has heightened the need for developing an efficient tool to assess default risks in this industry. Default prediction modelling relies heavily on how well a model fits the specific environment. Due to that, some adjustments have to take place in order to adapt the classical default prediction models to the Russian food service industry. We build hypotheses that adding non-financial factors and employing modern prediction methods can increase the accuracy of the models significantly. The aim of this study is to determine the effect of non-financial factors' inclusion and modern modelling methods on the accuracy of default prediction for the food service industry in Russia. Tests for a sample of 1241 firms for the period from 2017 to 2021 have shown that creating a prediction model with modern methods, such as Random Forest and XGBoost increases the accuracy of the prediction from 70 % to about 80 %, compared to standard Logit model. The addition of non-financial factors to the models also increases the accuracy slightly, which however, does not provide a significant effect. The most important metrics in predicting default turned out to be Current Liquidity Ratio and the ratio of Working Capital to Total Assets. The most important non-financial factors are Total Assets and Age. Our results correspond with existing research in this field and form a new knowledge layer due to being applied to a specific industry. The results can be used by banks or other counterparties that interact with food service industry firms in order to assess their credit risk.

Key words: default prediction; food service; non-financial factors; Machine Learning.

JEL G32, G33, G21, C58

1. Introduction

Recent years have been highly unstable for the enitre Russian business sphere due to COVID-19 and a heavier burden of sanctions. However, one of the strongest risks of sustainability was posed to the food service industry. During 2020–2021, due to restrictions limiting attendance at places where large groups congregate, demand for such services decreased or had changed to the "take-away" format. The following year, food services experienced issues with supplies of some products and equipment, cost growth as well while a decrease in demand due to a decline of disposable income of citizens.

Hence, the situation has heightened the need for the development of an efficient tool to assess default risks in food service industry.

Although the topic of default prediction is well-established in existing research literature, this investigation is still innovative for several reasons.

Firstly, a lot of time has passed since the first scientific works in the domain of

  ISSN 2712-7435

default prediction performed by Beaver [1], Altman [2], Taffler [3], Ohlson [4], Zmijevski [5] who have created the first bankruptcy prediction models, but there is still a lack of research exploring the experience of Russian private firms. Most of the foreign researchers have mentioned that their models are applicable to a certain type of firms, predominantly to American listed ones, but the specific Russian market conditions might differ significantly from those market conditions, which were examined in previous studies. Therefore, in terms of prediction accuracy, the Issue of conducting the research based on country specific data is highly relevant.

Secondly, although considerable research has been devoted to financial data usage for default prediction, relatively little attention has been paid to the integration of non-financial and industry specific parameters. We assume that this approach may significantly increase the prediction power of the model. Kwon & Lee [6] emphasized the importance of industry specific factors in the field of default prediction. In addition, authors claim that crises have different effects on the increase of unsustainability of firms in different industries.

Thirdly, the tools used for modelling have been developing over the years. Prediction modeling began with simple methods like multivariate discriminant analysis (MDA), like in the case of Altman [2] or linear Logit regressions as shown by Ohlson [4] and has now evolved into using such profound methods as General Regression Neural Network model, such as one created by Pan [7], and other Machine Learning tools, which make prediction for binary variables more precise. The examples include Random Forest, which was used by, for example, Brown & Mues [8] we set out to compare several techniques that can be used in the analysis of imbalanced credit scoring data sets. In a credit scoring context, imbalanced data sets frequently occur as the number of defaulting loans in a portfolio is usually much lower than the number of observations that do not default. As well as using traditional classification techniques such as logistic regression, neural networks and decision trees, this paper will also explore the suitability of gradient boosting, least square support vector machines and random forests for loan default prediction. Five real-world credit scoring data sets are used to build classifiers and test their performance. In our experiments, we progressively increase class imbalance in each of these data sets by randomly under-sampling the minority class of defaulters, so as to identify to what extent the predictive power of the respective techniques is adversely affected. The performance criterion chosen to measure this effect is the area under the receiver operating characteristic curve (AUC. Many of modern Machine Learning methods have never been applied to Russian market data or even for default risk analysis in general.

In this research we determine whether the addition of non-financial factors and application of modern modeling methods can increase the accuracy of default prediction for the Russian food service firms.

Our hypotheses are:

H1: Employing modern modelling techniques can significantly increase the accuracy of default prediction models in the food service industry. We suppose that the models, trained using Machine Learning algorithms show higher accuracy rates than conventional tools, like Logit regression, being able to capture for non-linear dependencies.

H2: Adding non-financial factors can significantly increase the accuracy of default prediction models in the food service industry. We hypothesize that expansion of predictors list by adding non-financial data will make the prediction power of the models significantly higher, because these new factors can enrich the information about the firm.

The aim of our research is to determine if non-financial factors like macroeconomic data, age of the company, number of owners, total assets, employee turnover etc. and modern modelling methods such as Random Forest and XGBoost increase the prediction power of default prediction models. And if so, create a model that can outperform the conventional ones in terms of accuracy.

This paper consists of several sections. The next section is devoted to the literature review. Then we describe the methods and data we used. Then we provide the results and discuss them.

2. Literature review

This section consists of 3 sub-sections. The first one is devoted to the overview of the history of default prediction, paying attention to the development of the approaches in terms of the default predictors chosen and modelling tools used. Then, we discuss in more details the benefits of new Machine Learning based modelling tools usage for default prediction. Finally, we give a picture of the previous research in the field of default prediction in food service industry.

2.1. The development of default prediction from 1960-s to the recent studies

Studies, which are considered to be the basis of default prediction modeling, are the works of Beaver [1], who used normative values for the financial ratios in order to assess credit risk, and Altman [2], who applied discriminant analysis to make a probabilistic model for US listed firms, followed by, for example, Jaki & Ćwięk [9], Xie et al. [10] and others.

The ability of current financial data to indicate future defaults has been the main assumption in the default prediction studies since the first research in this domain. For example, Jaki & Ćwięk [9] applied a vast variety of book ratios to create the models: liquidity, profitability, debt structure, debt, equity & assets coverage ratios. Other studies, like Boubaker et al. [11], use market indicators, which are not book ratios, but are still financial data. Also, some of the studies use dynamic ratios, like the change in revenue, net income or financial ratios, in particular Iwanicz-Drozdowska et al. [12], or Ohlson [4].

In the same time, a new trend in default prediction is developing nowadays — the use of non-financial variables. For example, Lugovskaya [13] reports an 11 percentage points increase in overall accuracy when using size and age variables along with financial ratios to predict the defaults. Bhimani et al. [14] report 0.21 units increase in area under ROC curve when using macroeconomic and non-financial (management, ownership and financial support related) variables along with financial ratios. This study also pays attention to the use of non-financial data to increase the accuracy of default prediction, using size and age, following for example Blanco-Oliver et al. [15], Altman et al. [16], Lugovskaya [13].

The methodology of default prediction has been developing since 1960-s. The first studies conducted were designed for public firms mostly and can be characterized by small samples in comparison to later papers. For example, Altman [2] used just 66 firms in his research. Also, Multiple Discriminant Analysis (MDA) approach for prediction models was the prevailing approach among the first default prediction studies, however it is sometimes characterized as biased e. g. Frank et al. [17] and inferior to other methods, e. g. Xie et al. [10], Wilson & Sharda [18].

The first to create a model based on the logistic regression, which is still one of the most popular ways of modelling credit risk, was Ohlson [4]. Apart from new statistical tool, the researcher used a big sample of firms to create the model — more than 2 thousand observations.

For a long time, Logit (and sometimes Probit, which is a similar algorithm) regression was the prevailing statistical tool for the purpose of default prediction modelling. For example, Zmijewski [5], Altman & Sabato [19], Gruszczyński [20], Hunter & Isachenkova [21], Lin & Piesse [22], Sirirattanaphonkun & Pattarathammas [23] followed the approach to create the models. One can find very recent works, which utilize logistic regression for the purpose of default prediction, such as Zhao & Lin [24] The studies report high accuracy, as it is shown by some examples in the Table 1.

Table 1. **Accuracy of Logit model accuracy in selected studies**

Study	Logit model accuracy, %
Zhao & Lin [24]	85
Altman & Sabato [19]	87
Gruszczyński [20]	up to 91
Hunter & Isachenkova [21]	up to 85

Source: prepared by the authors

However, as stated in, for example, Mselmi et al. [25], Barboza et al. [26], Machine Learning algorithms tend to be more precise in predicting defaults than discriminant analysis or Logit. The next section is devoted to a brief overview of such algorithms used in default prediction.

2.2. Modern Machine Learning algorithms in default prediction

Several modern modelling tools have been implemented in default prediction research and there is an increasing number of the studies conducted with use of new techniques. For example, the idea of new Machine Learning technique usage in default prediction was executed by Barboza et al. [26]. Using a sample of North American firms, they constructed models with support vector machine (SVM), boosting, bagging, Random Forest, and artificial neural networks (ANN) and achieved high accuracy (up to 0.93 in terms of the area under ROC curve).

Another example is the work by Brown & Mues [8], which showed the accuracy of classification of up to 0.95 (area under ROC curve) when using Machine Learning algorithms. The authors also show the Random Forest classifier, which is also used in this paper, as one of the best performing algorithms.

There are several research studies focused on different methods comparison, e. g. Mselmi et al. [25], Wu et al. [27]. In the latter study Altman's ratios and some new ones (EBIT to Sales, Total assets growth, Sales growth, Number of employees growth, ROE growth, Market price to Book price growth) was taken to build several models using different algorithms. As it was expected by the authors, Machine Learning approaches demonstrated better results on a testing sample in comparison with MDA and Logit (Random Forest 87.1 %, Boosting 86.7 %, Bagging 85.7 %, SVM-radial basis function 79.8 %, Logit 76.3 %, MDA 52.18 %).

Machine Learning algorithms can be applied separately, but an additional increase in performance can be achieved if one uses an ensemble. For example, Fedorova et al. [28] have applied Machine Learning techniques to predict bankruptcy in the sample of French, Italian, Russian and Spanish firms. The researchers have applied stacking ensemble technique for bankruptcy prediction and compared it with single classifier and bagging ensemble models. Results have shown that the stacking ensemble method is more accurate.

Overall, it seems that Machine Learning tools can increase the performance of default prediction significantly.

2.3. Default prediction for food service firms

Default prediction models have also been applied to the food service industry in, for example, the studies by Situm [29], Kim & Upneja [30], Gu [31].

The study by Situm [29] is a rare example of research, covering the value of non-financial data in default prediction for food service firms. Logistic regression was utilized to create the classification models. The significant impact of the size of the firm and the restaurant location was found. However, financial data is still stated as the strongest group of predictors.

The study by Kim & Upneja [30] utilized both Machine Learning algorithms (decision tree — based) and also non-financial factors, such as, for example, board holding ratio, and received a 74 % accurate model for US listed food service firms.

Gu [31] achieved a way higher (92 %) accuracy with an MDA model. The list of variables consisted of 12 financial ratios, without including any non-financial ones. The main interpretation of the results is the fact that firms with low EBIT and high total liabilities are the most susceptible to default, which corresponds with other studies, covering all industries.

Overall, it seems that there are some studies in the domain of default prediction for food service, however few of them utilize Machine Learning and, as shown in a summary of existing research, provided by Situm [29], most of the existing research covers USA or developed EU countries firms and few of the studies utilize non-financial data other than size and age. Moreover, we could not find any study, covering Russian food service industry.

3. Methods

In this research we make prediction models starting with the most basic ones, such as Logit, using the most popular predictors, then add more variables and use more advanced methods (Random Forest, XGBoost).

3.1. Binary linear regression (Logit)

Logit and Probit models are non-linear models which are constituted by linear combination of parameters. Logit is used more frequently in default prediction studies having several advantages over Probit models.

Firstly, the coefficients in Logit models are more easily interpreted as they represent the change in the log odds of the outcome for a one-unit change in the predictor variable, whereas in Probit models they represent the change in the standard deviation of the latent variable. Secondly, Logit models are more efficient than Probit models when errors are heteroskedastic, providing more precise estimates of coefficients. Thirdly, Logit models have a simpler likelihood function, making them easier to compute for large datasets. Fourthly, Logit models are more robust to outliers in data that may contain extreme values. Finally, Logit models are more commonly used in fields such as economics and political science, making them more familiar and easier to work with for researchers.

3.2. Random Forest

Nowadays Random Forest, described by Breiman [32], is a frequently used technique for classification tasks. It works by creating a large number of decision trees, each trained on a random subset of the full data sample. The trees are built using a random selection of features, which helps to reduce overfitting and improve generalization performance. The algorithm aggregates the predictions of all the trees to make the final prediction.

Random Forest has a high predictive power because it combines the predictions of multiple decision trees, which helps to reduce the risk of overfitting and improves the accuracy and robustness of the model. The algorithm also uses techniques such as bagging and feature importance measures to further improve performance and prevent overfitting.

3.3. XGBoost

XGBoost (Extreme Gradient Boosting), described by Chen & Guestrin [33], is a type of Machine Learning algorithm that is used for regression and classification problems. It is an extension of the gradient boosting algorithm that uses a regularized model to prevent overfitting and improve accuracy.

The XGBoost algorithm works by combining multiple weak models (decision trees) into a single strong model. Each decision tree is trained on a subset of the data, and the algorithm learns from the errors of each tree to improve its predictions. The trees are built iteratively, with each new tree attempting to correct the errors of the previous trees. XGBoost also includes several regularization techniques to prevent overfitting, such as L1 and L2 regularization, and early stopping. These techniques help to improve the generalization performance of the model and prevent it from memorizing the training data.

Overall, XGBoost is a powerful and flexible algorithm that has been applied successfully for many different tasks, including image classification, natural language processing, and predictive maintenance. However, this algorithm has not been widely used for default prediction yet.

3.4. Approaches comparison

All three methods have their benefits and drawbacks. Logit regression is a simple and interpretable method that works well for some cases, while Random Forest and XGBoost are more complex and suitable for large datasets with a large number of features. XGBoost is known for its high performance but can be sensitive to overfitting.

Depending on the size and complexity of the dataset, any of these methods could be suitable for default prediction. However, if computational efficiency is a concern, Logit might be the most suitable option. If the dataset is large and complex, and performance is a priority, Random Forest or XGBoost might be better options.

In case of this study, we are interested in the comparison between Logit and Machine Learning tools, that is why all three algorithms are used.

3.5. Data and Variables

There are many approaches for default definition and data selection to construct the model. Usually, the firm is identified as defaulted if the legal insolvency procedure has been launched or the firm is in process of liquidating voluntarily, for example in Kazakov & Kolyshkin [34], Karminsky & Burekhin [35], Afanasev & Tarasova [36].

For this study it was decided to define the date of a default event as the date of submission of a notice by the creditor, which is submitted to the court to start the legal case for insolvency. For the defaulted firms with no information about such notice it was decided to use the date of their insolvency declaration minus the average difference between the date of the creditor's notice submission and the date of insolvency declaration for the firms in a sample we have full information about, which turned out to be 315 days in this case.

The data used for the empirical study was collected from SPARK-Interfax database. In concordance with Russian Classification of Economic Activities we selected the following sub-industries, which are definitely parts of the food service industry:

- Restaurant activities and food delivery services;
- Full-service restaurants and cafes, cafeterias, fast food and self-service restaurants;
- Beverage service;
- Event catering activities;
- Other catering activities.

The final dataset consists of 1241 observations for the period from 2017 to 2021. Due to the huge disproportion of default and non-default firms we decided to use an over-sampling method for train sample construction.

The number of defaulted firms in the sample is 87. According to the data of the Federal Tax Service, the percent of default firms is stable over the years at the level of 1 %, therefore the share of default firms in our sample is comparable with the general population distribution.

Over-sampling has been done by the ROSE (Random Over-Sampling Examples) over-sampling method, which works by increasing the number of instances in the minority class by generating new synthetic observations. The process involves the following steps.

1. Random selection of instances.
2. Duplication of instances (the selected instances are then duplicated to create new synthetic instances. This results in an increase in the number of instances in the minority class).
3. Introduction of variations: To avoid creating exact replicas of the original instances, slight variations are introduced to create new synthetic instances.
4. Addition of synthetic instances: The synthetic instances are added to the original dataset, resulting in a balanced dataset.

The ROSE over-sampling method is effective in improving the accuracy of Machine Learning models when dealing with imbalanced datasets. We used the ROSE method only for the train sample after splitting the dataset for train and test.

We have estimated the accuracy test and the comparison of the models on a real test sample, which contains 30 % of observations. According to empirical evidence, provided by Menardi & Torelli [37], the usage of corrected data for tests, for example by ROSE method, leads to biased results which significantly differ from the outputs for the same tests taken on the real data.

We considered 19 financial and non-financial variables or ratios, obtained from the review of the previous research literature on default prediction. All the variables are considered quantitative and are represented in Table 2.

Table 2. **Variables used in the models**

Variable	Code
Efficiency	
Income/Non-Current Assets	E1
Income/Total Assets	E2
Revenue/Total Assets	E4

End of table 2

Variable	Code
Liquidity & Cash-Flow	
Current Assets/Current Liabilities	L1
EBIT/Current Liabilities	L2
Working Capital/Total Assets	L3
Profitability	
Net Profit/Income	P1
Net Profit/Total Assets (ROA)	P2
Gross profit margin	P4
Solvency	
Net Equity/Non-Current Liabilities	S3
Tendency	
Revenue growth rate	T1
Income growth rate	T2
Change in profitability	T3
Macroeconomics	
Disposable income change	M1
Non-Financial	
Total Assets (Log)	N1
Age (Log)	N2
Number of owners	N3
Ownership of patents	N4
Employee turnover	N5

Source: prepared by the authors

In order to model the prediction process, all the values for the variables for the defaulted firms were calculated based on the reporting, certainly available on the date of default. To ensure that the reporting is available, we used the reporting for the year, which is 2 years prior to the year of default, for the firms with the date of default from January to June, and the reporting for the year, which is 1 year prior to the year of default, for the firms with the date of default from July to December.

However, for two variables (Number of owners (N3), Ownership of patents (N4)) we used the last available information, making an assumption that this data does not change significantly from year to year.

The sample of non-default firms is distributed similar (in terms of the periods, which are used to calculate the vales for the independent variables) to the default sample. Also, we did not take non-default firms, which existed less than a half of the year before the reporting date, which is taken into account.

Table 3. **Descriptive statistics for the variables**

Variable	Min.	1st Qu.	Median	Mean	3rd Qu	Max.
E1	–3.13	0.44	2.37	181.53	18.31	92 986
E2	–80.885	–0.06	0.01	–5.53	0.15	47 895.9
E4	0	0.26	0.84	598.66	2.51	116 060.2
L1	–590.25	0.67	1.14	6.098	3.53	1 618
L2	–389	–0.005	0.09	1.05	0.62	1 436.29
L3	–2 620.94	–0.11	0.03	–0.65	0.36	133.4
P1	–2 276.95	–0.15	0.02	–1.2	0.14	588.31
P2	–80 885	–0.06	0.01	–5.53	0.15	47 895.9
P4	–12 955	0.016	0.21	–3.31	0.49	2
S3	–14 720	0.1	1	828	3.4	1 263.729
M1	–4.5	–2	1.2	0.53	3	3
T1	–146.7	–0.19	0.17	36.61	0.69	90 337.5
T2	–2 302.74	–1.27	–0.53	3.04	0.38	5 567.17
T3	–1 602.09	–1.17	–0.67	–0.55	0.26	2 835.65
N1	8	16.12	17.496	17.33	18.89	24.76
N2	0	1.95	2.49	2.36	2.89	3.47
N3	1	1	1	1.55	2	9
N5	–42 309	–3	0	–27.71	2	708
N4	0: 4 805 obs.			1: 50 obs.		

Source: prepared by the authors

4. Results

4.1. Altman's model

To determine the starting point, we conducted default prediction based on model by Altman [2] using our dataset. The main point of this task was to determine whether the conventional models are good enough for making predictions for the Russian food service industry firms.

We used the initial financial indicators and ratios of Altman's model with the corresponding coefficients, provided by Altman. The accuracy of predictions came out to be 67 % (69 % for non-default and 64 % for default companies). For this calculation we used a sample of 2422 (after oversampling) observations. We suppose that the reason for such a low prediction accuracy is that the initial coefficients of the model cannot be applied to a new dataset and should be estimated on our data.

Following that, we decided to fit Logit regression on our training sample. The results are shown in Table 4. The accuracy of prediction on the test sample is higher, but still not very high: 68 % (71 % for non-default and 65 % for default firms).

4.2. Models with extended list of variables

4.2.1 Logit model

To begin with, we constructed a linear Logit regression with intercept, heteroskedasticity corrected standard errors. We created and compared two versions of this model: one with financial variables only and the second with both financial and non-financial variables (see Table 5).

Table 4. **The results for Logit regression with Altman's ratios**

Variable	Coefficient	St. error	Significance*
Intercept	0.490	0.27	0.05
Working Capital / Total Assets	0.015	0.038	1
Retained Earnings / Total Assets	0.013	0.015	1
Earnings Before Interest and Tax / Total Assets	0.027	0.10	1
Equity / Total Liabilities	-1.27	0.38	0.001
Total Sales / Total Assets	0,019	0.02	1

Note: prepared by the authors; * If the p-value is higher than 0.1, we code Significance as "1", showing that the coefficient is insignificant. If the p-value is less or equal to 0.1, we code Significance according to the nearest significance level (0.1, 0.05, 0.01 or 0.001)

Table 5. **Logit model results**

Variable	Only financial data		Financial and non-financial data	
	Coefficient	Significance	Coefficient	Significance
Intercept	–0.097	1	4.06	0.05
E1	0.002	1	0.0013	1
E2	–0.62	0.05	–0.85	0.01
E4	0.05	0.05	0.07	0.05
L1	–0.05	1	–0.04	1
L2	0.11	0.1	0.11	0.1
L3	–0.15	1	–0.26	0.1
P1	–0.08	1	–0.11	1
P4	–0.27	1	–0.39	1
S3	–0.002	1	–0.0018	1
T1	–0.005	1	–0.003	1
T2	0.03	1	0.007	1
T3	–0.03	1	–0.02	1
M1	–0.11	0.1	–0.10	1
N1	—	—	–0.10	1

End of table 5

Variable	Only financial data		Financial and non-financial data	
	Coefficient	Significance	Coefficient	Significance
N2	—	—	–0.91	0.001
N3	—	—	–0.23	1
N4	—	—	6.50	1
N5	—	—	–0.003	1

Source: prepared by the authors

For these and all further models the main metrics, which determine how accurate is the model, are sensitivity (accuracy for default firms) and specificity (accuracy for non-default firms), estimated on the test samples. Even though overall accuracy is an important metric, these two give a clearer picture and make it possible to check both for Type 1 (defining a defaulted firm as a healthy one) and Type 2 (defining a healthy firm as a defaulted one) errors.

The first version of the model (the one without non-financial factors) provided poor results in terms of overall accuracy, sensitivity and specificity: 0.678, 0.600 and 0.679 respectively. The addition of non-financial factors in the second version of the model led to an insignificant increase in accuracy: 0.681, 0.550, 0.683 respectively. The only one important non-financial variable in Logit regression is N2 (Log Age of a company).

Both models produced low accuracy. Machine Learning was expected to provide much more accurate predictions.

4.2.2 Random Forest model

Next, we implemented some of the modern modeling methods, starting with Random Forest. We tested out three models using this method: without non-financial factors, with non-financial factors, and a variation with Altman's factors only.

The first two models produced overall accuracy, sensitivity, and specificity values of 0.77, 0.90, 0.76 and 0.79, 0.85, 0.78 respectively. Two things can be clearly seen in these cases. First, the addition of non-financial factors does not increase the predicting power of the model significantly. Second, the model itself is much more accurate than Logit.

Random Forest also provides us with the mean decrease of Gini coefficient. It is a measure of how each variable contributes to the homogeneity of the nodes and leaves in the resulting Random Forest. This means that the importance of each variable can be determined by how high its mean decrease Gini value is.

Five variables with the highest influence on Gini index are M1, L1, L3, E4 and N1 (Disposable income change, Current assets/Current liabilities, Working capital/Total assets, Revenue/Total assets, Total assets(log) respecitvely). Another non-financial factor N2 (Age (log)) is also present in the top 10 variables. Even though non-financial factors do not provide a significant effect on the accuracy, some of them are still among the most important variables.

These results can be explained by several factors. M1 (Disposable income change) could have been placed so high due to the fact that attending restaurants or other food services is usually a leisurely or luxurious activity, so if people do not possess enough disposable income, then the inflow of customers for food services is greatly decreased. L1 and L3 (Current assets/Current liabilities, Working capital/Total assets) are also important due to the fact that food services have a much higher asset turnover than many other industries, so these two factors play an especially big role.

The two most important non-financial factors are N1 and N2 (Total assets(log), Age (log)). N2 may be important due to the specific nature of the food service market. Since the average life cycle of food service companies is low, at around two years, the age turns out to be a decisive factor. The older is the firm, the higher is the probability to overcome financial distress.

The Random Forest model with Altman's variables also showed much better results than Logit model, providing values of 0.79, 0.80 and 0.77 for accuracy, sensitivity and specificity respectively, which is a significant improvement compared to Logit model.

The Gini coefficient shows that Equity/Total liabilities is the most important variable, followed by Working capital/Total assets and then by Retained earnings/Total assets, Total sales/Total assets and EBIT/Total assets.

4.2.3 XGBoost

XGBoost is one of the strongest modern prediction models, which has not yet been used much in default prediction research. This method is a type of Machine Learning which is based on decision trees and works well with unbalanced samples of observations.

The model created with this method provided better results than the previous ones. It resulted in the accuracy of 0.81, while sensitivity and specificity are 0.70 and 0.82 respectively.

This method also provided a list of the most important variables. The five most important ones are L1, L3, T1, N2 and E2 (Current assets/Current liabilities, Working capital/Total assets, Revenue growth rate, Age (log), Income/Total assets). N1 (Total assets (log)) is also included in the top 10.

It is interesting to compare this result to the result of the Random Forest model. Both models placed L1 and L3 among the most important variables and also N1 and N2 were the ones included in the top 10 (Table 6).

Table 6. **Results of Logit, Random Forest and XGBoost, %**

	Logit (financial data only)	Logit (financial and non-financial data)	Random Forest (financial data only)	Random Forest (financial and non-financial data)	XGBoost (financial and non-financial data)
Accuracy	68	68	77	79	81
Sensitivity	60	55	90	85	70
Specificity	68	68	76	78	82

Source: prepared by the authors

5. Discussion

5.1. Hypotheses confirmation

The first hypothesis of the study is *H1:* Employing modern modelling techniques can significantly increase the accuracy of default prediction models in the food service industry. The hypothesis is confirmed: we show that the accuracy of default prediction is significantly higher if Machine Learning algorithms are used to train the models, rather than in case of Logit regression.

Some researchers go deeper and use neural networks. For example, in a recent work by Becerra-Vicario et al. [38] Deep Recurrent Convolutional Neural Network (DRCNN) were constructed for default prediction on the Spanish restaurant industry, accuracy of the model exceeded logistic regression in predicting capacity. The DRCNN model predicted default on data one, two, and three years prior to default with 93.5 %, 89.6 %, 85.6 %, respectively. The researchers used financial ratios and quality certification results as variables.

The main Issue of this method is its "black-box" nature, that means that we actually do not know what kind of rules were developed, the only thing we know is the importance of a variable. Although we suppose that neural networks might have slightly higher accuracy rate, we decided not to use them in this study. Firstly, Neural Networks require huge data samples in order to be well trained. Secondly, "black-box" decisions are not demanded by users of default prediction models due to the lack of interpretability.

The second hypothesis of the study is *H2:* Adding non-financial factors can significantly increase the accuracy of default prediction models in the food service industry. We treat this hypothesis as unconfirmed. The accuracy, achieved with non-financial data, is indeed the highest, however, the increase is only few percentage points. Given that non-financial data usually requires more time and effort to collect, it is necessary to conduct cost-benefit analysis in every case of default prediction to estimate whether it is beneficial to invest in collection of non-financial data, which is analyzed in this study.

5.2. Limitations and potential for future research

Despite, the second hypothesis is not confirmed, due to existing literature, for example Altman et al. [16], Fernando et al. [39], Makeeva & Sinilshchikova [40], Lugovskaya [13], Bhimani et al. [14] which report increase in accuracy while using non-financial data, we still believe that there is a room for improvement in default prediction models through non-financial factors.

One of the limitations of this research is that we do not use some non-financial variables that may increase the prediction capacity for food service companies, because the collection of such data requires significant resources and cannot be conducted automatically. Some of these variables include:

1. Customer reviews and ratings. This can provide insights into customer satisfaction and loyalty, which are important indicators of a firm's success.
2. Social media engagement. Measuring the level of social media engagement can provide valuable information about a firm's brand awareness and reputation.
3. Employee satisfaction. Employee satisfaction can impact customer service and productivity, which are important factors for a firm's success.
4. Location and demographic information. Analyzing the geographic location and demographic information of a company's customer base can help identify areas for growth and target marketing efforts.
5. Menu variety and innovation. Offering unique and innovative menu items can set a company apart from its competitors and attract new customers.
6. Online ordering and delivery. The ability to order food online and have it delivered is becoming increasingly important in the food service industry, and companies that offer this service may have an advantage over those that do not.
7. Health and safety practices. Ensuring that food is prepared and served safely is critical for maintaining customer trust and preventing foodborne illnesses.

By incorporating these non-financial variables into a prediction model, it may be possible to improve the accuracy of default predictions for food service companies. This is considered as a potential path for the future research.

Another limitation of the research is that we have to treat the data for legal entity as the data reflecting the condition of the whole business. However, often Russian entrepreneurs disaggregate the business, and it may consist of several legal entities and sole proprietors. This limitation is common for many studies related to Russian firms.

It is also important to notice the prevalence of type I errors over type II errors in our tests. This can be a potential topic for research on whether banks and other counterparties would rather accept the risk of dealing with a company that is going to go bankrupt or missing out on profits from declining relationships with firms which turn out to be healthy business.

6. Conclusion

The aim of the research was to determine the difference in results between the classical default models and modern methodology by introducing new variables and advanced Machine Learning methods.

Tests performed with classical models showed results with low prediction accuracy at the levels of about 70 %. We attribute such a result to a number of factors such as the model having poor compatibility with non-listed companies, a small size of the test sample, and the food service industry being too specific for general models.

Modern methods like Random Forest and XGBoost showed much better results, producing accuracy of about 80 %. Thus, the implementation of modern algorithms greatly increased the accuracy and proved them to be stronger default prediction tools.

We discovered that Current Assets/Current Liabilities and Working capital/Total assets are two variables that both Machine Learning models found to be among the most important factors.

Also, the addition of non-financial factors into the models led to a slightly higher accuracy of prediction. We, however, did not record a significant increase. The most important non-financial factors are Total Assets and Age.

Thus, we tested how the addition of non-financial factors and modern modelling tools can impact the accuracy of default prediction models. This research contributes to both the field of default prediction research in general and to the research of food service companies in particular by examining the factors, which influence the accuracy of credit risk estimation. Our research and model can be practically useful for credit organizations or any other counterparties that could deal with food service firms in Russia.

References

1. Beaver, W. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research,* Vol. 4, 71–111. https://doi.org/10.2307/2490171

2. Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance,* Vol. 23, No. 4, 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

3. Taffler, R. (1982). Forecasting Company Failure in the UK Using Discriminant Analysis and Financial Ratio Data. *Journal of the Royal Statistical Society. Series A (General),* Vol. 145, No. 3, 342–358. https://doi.org/10.2307/2981867

4. Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research,* Vol. 18, No. 1, 109–131. https://doi.org/10.2307/2490395

5. Zmijewski, M. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research,* Vol. 22, 59–82. https://doi.org/10.2307/2490859

6. Kwon, T., Lee, Y. (2018). Industry specific defaults. *Journal of Empirical Finance,* Vol. 45, 45–58. https://doi.org/10.1016/j.jempfin.2017.10.002

7. Pan, W.-T. (2012). A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example. *Knowledge Based Systems — KBS,* Vol. 26, 69–74. https://doi.org/10.1016/j.knosys.2011.07.001

8. Brown, I., Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications,* Vol. 39, Issue 3, 3446–3453. https://doi.org/10.1016/j.eswa.2011.09.033

9. Jaki, A., Ćwięk, W. (2021). Bankruptcy Prediction Models Based on Value Measures. *Journal of Risk and Financial Management,* Vol. 14, Issue 1, 6. https://doi.org/10.3390/jrfm14010006

10. Xie, C., Luo, C., Yu, X. (2011). Financial distress prediction based on SVM and MDA methods: The case of Chinese listed companies. *Quality & Quantity,* Vol. 45, No. 3, 671–686. https://doi.org/10.1007/s11135-010-9376-y

11. Boubaker, S., Cellier, A., Manita, R., Saeed, A. (2020). Does corporate social responsibility reduce financial distress risk? *Economic Modelling,* Vol. 91, 835–851. https://doi.org/10.1016/j.econmod.2020.05.012

12. Iwanicz-Drozdowska, M., Laitinen, E.K., Suvas, A., Altman, E. (2016). Financial and non-financial variables as long-horizon predictors of bankruptcy. The *Journal of Credit Risk,* Vol. 12, No. 4, 49–78. https://doi.org/10.21314/JCR.2016.216

13. Lugovskaya, L. (2010). Predicting default of Russian SMEs on the basis of financial and non-financial variables. *Journal of Financial Services Marketing,* Vol. 14, Issue 4, 301–313. https://doi.org/10.1057/fsm.2009.28

14. Bhimani, A., Gulamhussen, M., Lopes, S. da R. (2013). The Role of Financial, Macroeconomic, and Non-financial Information in Bank Loan Default Timing Prediction. *European Accounting Review,* Vol. 22, Issue 4, 739–763. https://doi.org/10.1080/09638180.2013.770967

15. Blanco-Oliver, A., Irimia-Dieguez, A., Oliver-Alfonso, M.D., Vázquez-Cueto, M.J. (2016). Hybrid model using Logit and nonparametric methods for predicting micro-entity failure. *Investment Management and Financial Innovations,* Vol. 13, 35–46. http://dx.doi.org/10.21511/imfi.13(3).2016.03

16. Altman, E., Sabato, G., Wilson, N. (2010). The value of non-financial information in SME risk management. *Journal of Credit Risk,* Vol. 6, 95–127. http://doi.org/10.21314/JCR.2010.110

17. Frank, R., Massy, W., Morrison, D. (1965). Bias in Multiple Discriminant Analysis. *Journal of Marketing Research,* Vol. 2, No. 3, 250–258. https://doi.org/10.2307/3150183

18. Wilson, R., Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems,* Vol. 11, No. 5, 545–557. https://doi.org/10.1016/0167-9236(94)90024-8

19. Altman, E., Sabato, G. (2007). Modelling Credit Risk for SMEs: Evidence from the U. S. Market. *Abacus,* Vol. 43, Issue 3, 332–357. https://doi.org/10.1111/j.1467-6281.2007.00234.x

20. Gruszczyński, M. (2004). Financial Distress of Companies in Poland. International *Advances in Economic Research,* Vol. 10, No. 4, 249–256. https://doi.org/10.1007/BF02295137

21. Hunter, J., Isachenkova, N. (2001). Failure risk: A comparative study of UK and Russian firms. *Journal of Policy Modeling,* Vol. 23, Issue 5, 511–521. https://doi.org/10.1016/S0161-8938(01)00064-3

22. Lin, L., Piesse, J. (2004). The identification of corporate distress in UK industrials: a conditional probability analysis approach. *Applied Financial Economics,* Vol. 14, Issue 2, 73–82. https://doi.org/10.1080/0960310042000176344

23. Sirirattanaphonkun, W., Pattarathammas, S. (2012). Default Prediction for Small-Medium Enterprises in Emerging Market: Evidence from Thailand. *Seoul Journal of Business,* Vol. 18, No. 2, 25–54. http://dx.doi.org/10.35152/snusjb.2012.18.2.002

24. Zhao, Y., Lin, D. (2023). Prediction of Micro- and Small-Sized Enterprise Default Risk Based on a Logistic Model: Evidence from a Bank of China. *Sustainability,* Vol. 15, Issue 5, 4097. https://doi.org/10.3390/su15054097

25. Mselmi, N. Lahiani, A., Hamza, T. (2017). Financial distress prediction: The case of French small and medium-sized firms. *International Review of Financial Analysis,* Vol. 50, 67–80. https://doi.org/10.1016/j.irfa.2017.02.004

26. Barboza, F., Kimura, H., Altman, E. (2017). Machine Learning models and bankruptcy prediction. *Expert Systems with Applications,* Vol. 83, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

27. Wu, Y., Gaunt, C., Gray, S. (2010). A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting & Economics,* Vol. 6, Issue 1, 34–45. https://doi.org/10.1016/j.jcae.2010.04.002

28. Fedorova, E., Ledyaeva, S., Drogovoz, P., Nevredinov, A. (2022). Economic policy uncertainty and bankruptcy filings. *International Review of Financial Analysis,* Vol. 82, 102174. https://doi.org/10.1016/j.irfa.2022.102174

29. Situm, M. (2023). Financial distress in the Austrian tourism industry: hotels and restaurants analysis. *European Journal of Tourism Research,* Vol. 34, 3411. https://doi.org/10.54055/ejtr.v34i.2992

30. Kim, S., Upneja, A. (2014). Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models. *Economic Modelling,* Vol. 36, 354–362. https://doi.org/10.1016/j.econmod.2013.10.005

31. Gu, Z. (2002). Analyzing bankruptcy in the restaurant industry: A multiple discriminant model. *International Journal of Hospitality Management,* Vol. 21, Issue 1, 25–42. https://doi.org/10.1016/S0278-4319(01)00013-5

32. Breiman, L. (2001). Random Forests. *Machine Learning,* Vol. 45, No. 1, 5–32. https://doi.org/10.1023/A:1010933404324

33. Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* San Francisco, 785–794. https://doi.org/10.1145/2939672.2939785

34. Kazakov, A., Kolyshkin, A. (2018). Bankruptcy prediction models development in modern Russian conditions. *St Petersburg University Journal of Economic Studies,* Vol. 34, No. 2, 241–266. (In Russ.). https://doi.org/10.21638/11701/spbu05.2018.203

35. Karminsky, A., Burekhin, R. (2019). Comparative analysis of methods for forecasting bankruptcies of Russian construction companies. *Business Informatics,* Vol. 13, No. 3, 52–66. https://doi.org/10.17323/1998-0663.2019.3.52.66

36. Afanasev, V., Tarasova, Y. (2022). Default Prediction for Housing and Utilities Management Firms Using Non-Financial Data. *Financial Journal,* Vol. 14, No. 6, 91–110. https://doi.org/10.31107/2075-1990-2022-6-91-110

37. Menardi, G., Torelli, N. (2012). Training and assessing classification rules with unbalanced data. *Data Mining and Knowledge Discovery,* Vol. 28, 92–122. https://doi.org/10.1007/s10618-012-0295-5

38. Becerra-Vicario, R., Alaminos, D., Aranda, E., Fernández-Gámez, M.A. (2020). Deep Recurrent Convolutional Neural Network for Bankruptcy Prediction: A Case of the Restaurant Industry. *Sustainability,* Vol. 12, Issue 12, 5180. https://doi.org/10.3390/su12125180

39. Fernando, J., Li, L., Hou, G. (2020). Financial versus Non-Financial Information for Default Prediction: Evidence from Sri Lanka and the USA. *Emerging Markets Finance and Trade,* Vol. 56, Issue 3, 673–692. https://doi.org/10.1080/1540496X.2018.1545644

40. Makeeva, E., Sinilshchikova, M. (2020). News Sentiment in Bankruptcy Prediction Models: Evidence from Russian Retail Companies. *Journal of Corporate Finance Research,* Vol. 14, No. 4, 7–18. https://doi.org/10.17323/j.jcfr.2073-0438.14.4.2020.7-18

INFORMATION ABOUT AUTHORS

Egor Olegovich Bukharin
Master Student, Saint-Petersburg School of Economics and Management, National Research University Higher School of Economics, Saint-Petersburg, Russia (190121, Saint-Petersburg, Soyuza Pechatnikov street, 16); ORCID https://orcid.org/0009-0004-3135-5469 e-mail: egor-bukharin01@gmail.com

Sofia Igorevna Mangileva
Senior Research Analyst, «Yakov & Partners», Moscow, Russia (125047, Moscow, Lesnaya street, 5, Building C); ORCID https://orcid.org/0009-0001-9574-1834 e-mail: sofiamangileva@gmail.com

Vladislav Viktorovich Afanasev
Lecturer, Department of Finance, Saint-Petersburg School of Economics and Management, National Research University Higher School of Economics, Saint-Petersburg, Russia (190121, Saint-Petersburg, Soyuza Pechatnikov street, 16); ORCID https://orcid.org/0000-0002-4041-4465 e-mail: vvafanasev@hse.ru

FOR CITATION

Bukharin, E.O., Mangileva, S.I., Afanasev, V.V. (2024). Default Prediction for Russian Food Service Firms: Contribution of Non-Financial Factors and Machine Learning. *Journal of Applied Economic Research,* Vol. 23, No. 1, 206–226. https://doi.org/10.15826/vestnik.2024.23.1.009

ARTICLE INFO

Received June 30, 2023; Revised December 5, 2023; Accepted January 12, 2024.

УДК 336.6

Прогноз дефолта для российских предприятий общественного питания: вклад нефинансовых факторов и машинного обучения

Е. О. Бухарин[1] ✉, *С. И. Мангилева*[2], *В. В. Афанасьев*[1]

[1]*Национальный исследовательский университет «Высшая школа экономики», г. Санкт-Петербург, Россия*

[2]*Компания «Яков и Партнеры», г. Москва, Россия*

✉ *egorbukharin01@gmail.com*

Аннотация. Нестабильность на рынке общественного питания в связи с пандемией COVID-19 и санкциями обострила потребность в разработке эффективного инструмента оценки рисков дефолта в этой отрасли. Качество прогнозирования дефолта в значительной степени зависит от того, насколько хорошо модель соответствует конкретной среде. В связи с этим необходимо внести некоторые коррективы, чтобы адаптировать классические модели прогнозирования дефолтов к российскому сектору общественного питания. В статье выдвинута гипотеза о том, что добавление нефинансовых факторов и использование современных методов прогнозирования может существенно повысить точность моделей. Целью данного исследования является определение влияния включения нефинансовых факторов и современных методов моделирования на точность прогнозирования дефолтов для предприятий общественного питания в России. Тесты на выборке из 1241 фирмы за период с 2017 по 2021 г. показали, что создание модели прогнозирования с помощью современных методов, таких как Random Forest и XGBoost, повышает точность прогнозирования с 70 % до примерно 80 %, по сравнению со стандартной логит-моделью. Добавление в модели нефинансовых факторов также несколько повышает точность, однако не дает существенного эффекта. Важнейшими метриками в прогнозировании дефолта оказались коэффициент текущей ликвидности и отношение оборотного капитала к совокупным активам. Наиболее важными нефинансовыми факторами являются совокупные активы и возраст. Наши результаты согласуются с уже существующими исследованиями в этой области и формируют новый пласт знаний за счет применения в конкретной отрасли. Результаты могут быть использованы банками или другими контрагентами, которые взаимодействуют с предприятиями общественного питания, для оценки их кредитного риска.

Ключевые слова: прогнозирование дефолта; общественное питание; нефинансовые факторы; машинное обучение.

Список использованных источников

1. *Beaver W.* Financial Ratios as Predictors of Failure // Journal of Accounting Research. 1966. Vol. 4. Pp. 71–111. https://doi.org/10.2307/2490171

2. *Altman E.* Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy // Journal of Finance. 1968. Vol. 23, No. 4. Pp. 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

3. *Taffler R.* Forecasting Company Failure in the UK Using Discriminant Analysis and Financial Ratio Data // Journal of the Royal Statistical Society. Series A (General). 1982. Vol. 145, No. 3. Pp. 342–358. https://doi.org/10.2307/2981867

4. *Ohlson J.* Financial Ratios and the Probabilistic Prediction of Bankruptcy // Journal of Accounting Research. 1980. Vol. 18, No. 1. Pp. 109–131. https://doi.org/10.2307/2490395

5. *Zmijewski M.* Methodological Issues Related to the Estimation of Financial Distress Prediction Models // Journal of Accounting Research. 1984. Vol. 22. Pp. 59–82. https://doi.org/10.2307/2490859

6. *Kwon T., Lee Y.* Industry specific defaults // Journal of Empirical Finance. 2018. Vol. 45. Pp. 45–58. https://doi.org/10.1016/j.jempfin.2017.10.002

7. *Pan W.-T.* A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example // Knowledge Based Systems — KBS. 2012. Vol. 26. Pp. 69–74. https://doi.org/10.1016/j.knosys.2011.07.001

8. *Brown I., Mues C.* An experimental comparison of classification algorithms for imbalanced credit scoring data sets // Expert Systems with Applications. 2012. Vol. 39, Issue 3. Pp. 3446–3453. https://doi.org/10.1016/j.eswa.2011.09.033

9. *Jaki A., Ćwięk W.* Bankruptcy Prediction Models Based on Value Measures // Journal of Risk and Financial Management. 2021. Vol. 14, Issue 1. 6. https://doi.org/10.3390/jrfm14010006

10. *Xie C., Luo C., Yu X.* Financial distress prediction based on SVM and MDA methods: The case of Chinese listed companies // Quality & Quantity. 2011. Vol. 45, No. 3. Pp. 671–686. https://doi.org/10.1007/s11135-010-9376-y

11. *Boubaker S., Cellier A., Manita R., Saeed A.* Does corporate social responsibility reduce financial distress risk? // Economic Modelling. 2020. Vol. 91. Pp. 835–851. https://doi.org/10.1016/j.econmod.2020.05.012

12. *Iwanicz-Drozdowska M., Laitinen E. K., Suvas A., Altman E.* Financial and nonfinancial variables as long-horizon predictors of bankruptcy // The Journal of Credit Risk. 2016. Vol. 12, No. 4. Pp. 49–78. https://doi.org/10.21314/JCR.2016.216

13. *Lugovskaya L.* Predicting default of Russian SMEs on the basis of financial and non-financial variables // Journal of Financial Services Marketing. 2010. Vol. 14, Issue 4. Pp. 301–313. https://doi.org/10.1057/fsm.2009.28

14. *Bhimani A., Gulamhussen M., Lopes S. da R.* The Role of Financial, Macroeconomic, and Non-financial Information in Bank Loan Default Timing Prediction // European Accounting Review. 2013. Vol. 22, Issue 4. Pp. 739–763. https://doi.org/10.1080/09638180.2013.770967

15. *Blanco-Oliver A., Irimia-Dieguez A., Oliver-Alfonso M. D., Vázquez-Cueto M. J.* Hybrid model using Logit and nonparametric methods for predicting micro-entity failure // Investment Management and Financial Innovations. 2016. Vol. 13. Pp. 35–46. http://dx.doi.org/10.21511/imfi.13(3).2016.03

16. *Altman E., Sabato G., Wilson N.* The value of non-financial information in SME risk management // Journal of Credit Risk. 2010. Vol. 6. Pp. 95–127. http://doi.org/10.21314/JCR.2010.110

17. *Frank R., Massy W., Morrison D.* Bias in Multiple Discriminant Analysis // Journal of Marketing Research. 1965. Vol. 2, No. 3. Pp. 250–258. https://doi.org/10.2307/3150183

18. *Wilson R., Sharda R.* Bankruptcy prediction using neural networks // Decision Support Systems. 1994. Vol. 11, No. 5. Pp. 545–557. https://doi.org/10.1016/0167-9236(94)90024-8

19. *Altman E., Sabato G.* Modelling Credit Risk for SMEs: Evidence from the U. S. Market // Abacus. 2007. Vol. 43, Issue 3. Pp. 332–357. https://doi.org/10.1111/j.1467-6281.2007.00234.x

20. *Gruszczyński M.* Financial Distress of Companies in Poland // International Advances in Economic Research. 2004. Vol. 10, No. 4. Pp. 249–256. https://doi.org/10.1007/BF02295137

21. *Hunter J., Isachenkova N.* Failure risk: A comparative study of UK and Russian firms // Journal of Policy Modeling. 2001. Vol. 23, Issue 5. Pp. 511–521. https://doi.org/10.1016/S0161-8938(01)00064-3

22. *Lin L., Piesse J.* The identification of corporate distress in UK industrials: a conditional probability analysis approach // Applied Financial Economics. 2004. Vol. 14, Issue 2. Pp. 73–82. https://doi.org/10.1080/0960310042000176344

23. *Sirirattanaphonkun W., Pattarathammas S.* Default Prediction for Small-Medium Enterprises in Emerging Market: Evidence from Thailand // Seoul Journal of Business. 2012. Vol. 18, No. 2. Pp. 25–54. http://dx.doi.org/10.35152/snusjb.2012.18.2.002

24. *Zhao Y., Lin D.* Prediction of Micro- and Small-Sized Enterprise Default Risk Based on a Logistic Model: Evidence from a Bank of China // Sustainability. 2023. Vol. 15, Issue 5. 4097. https://doi.org/10.3390/su15054097

25. *Mselmi N., Lahiani A., Hamza T.* Financial distress prediction: The case of French small and medium-sized firms // International Review of Financial Analysis. 2017. Vol. 50. Pp. 67–80. https://doi.org/10.1016/j.irfa.2017.02.004

26. *Barboza F., Kimura H., Altman E.* Machine Learning models and bankruptcy prediction // Expert Systems with Applications. 2017. Vol. 83. Pp. 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

27. *Wu Y., Gaunt C., Gray S.* A comparison of alternative bankruptcy prediction models // Journal of Contemporary Accounting & Economics. 2010. Vol. 6, Issue 1. Pp. 34–45. https://doi.org/10.1016/j.jcae.2010.04.002

28. *Fedorova E., Ledyaeva S., Drogovoz P., Nevredinov A.* Economic policy uncertainty and bankruptcy filings // International Review of Financial Analysis. 2022. Vol. 82. 102174. https://doi.org/10.1016/j.irfa.2022.102174

29. *Situm M.* Financial distress in the Austrian tourism industry: hotels and restaurants analysis // European Journal of Tourism Research. 2023. Vol. 34. 3411. https://doi.org/10.54055/ejtr.v34i.2992

30. *Kim S., Upneja A.* Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models // Economic Modelling. 2014. Vol. 36. Pp. 354–362. https://doi.org/10.1016/j.econmod.2013.10.005

31. *Gu Z.* Analyzing bankruptcy in the restaurant industry: A multiple discriminant model // International Journal of Hospitality Management. 2002. Vol. 21, Issue 1. Pp. 25–42. https://doi.org/10.1016/S0278-4319(01)00013-5

32. *Breiman L.* Random Forests // Machine Learning. 2001. Vol. 45, No. 1. Pp. 5–32. https://doi.org/10.1023/A:1010933404324

33. *Chen T., Guestrin C.* XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, 2016. Pp. 785–794. https://doi.org/10.1145/2939672.2939785

34. *Казаков А., Колышкин А.* Разработка моделей прогнозирования банкротства в современных российских условиях // Вестник Санкт-Петербургского университета. Экономика. 2018. Т. 34, № 2. С. 241–266. https://doi.org/10.21638/11701/spbu05.2018.203

35. *Karminsky A., Burekhin R.* Comparative analysis of methods for forecasting bankruptcies of Russian construction companies // Business Informatics. 2019. Vol. 13, No. 3. Pp. 52–66. https://doi.org/10.17323/1998-0663.2019.3.52.66

36. *Afanasev V., Tarasova Y.* Default Prediction for Housing and Utilities Management Firms Using Non-Financial Data // Financial Journal. 2022. Vol. 14, No. 6. Pp. 91–110. https://doi.org/10.31107/2075-1990-2022-6-91-110

37. *Menardi G., Torelli N.* Training and assessing classification rules with unbalanced data // Data Mining and Knowledge Discovery. 2012. Vol. 28. Pp. 92–122. https://doi.org/10.1007/s10618-012-0295-5

38. *Becerra-Vicario R., Alaminos D., Aranda E., Fernández-Gámez M. A.* Deep Recurrent Convolutional Neural Network for Bankruptcy Prediction: A Case of the Restaurant Industry // Sustainability. 2020. Vol. 12, Issue 12. 5180. https://doi.org/10.3390/su12125180

39. *Fernando J., Li L., Hou G.* Financial versus Non-Financial Information for Default Prediction: Evidence from Sri Lanka and the USA // Emerging Markets Finance and Trade. 2020. Vol. 56, Issue 3. Pp. 673–692. https://doi.org/10.1080/1540496X.2018.1545644

40. *Makeeva E., Sinilshchikova M.* News Sentiment in Bankruptcy Prediction Models: Evidence from Russian Retail Companies // Journal of Corporate Finance Research. 2020. Vol. 14, No. 4. Pp. 7–18. https://doi.org/10.17323/j.jcfr.2073-0438.14.4.2020.7-18

ИНФОРМАЦИЯ ОБ АВТОРАХ

Бухарин Егор Олегович
Студент магистратуры Санкт-Петербургской школы экономики и менеджмента Санкт-Петербургского филиала, Национальный исследовательский университет «Высшая школа экономики», г. Санкт-Петербург, Россия (190121, г. Санкт-Петербург, ул. Союза Печатников, 16); ORCID https://orcid.org/0009-0004-3135-5469 e-mail: egorbukharin01@gmail.com

Мангилева Софья Игоревна
Старший аналитик, Компания «Яков и Партнеры», г. Москва, Россия (125047, Москва, ул. Лесная 5, строение С); ORCID https://orcid.org/0009-0001-9574-1834 e-mail: sofiamangileva@gmail.com

Афанасьев Владислав Викторович
Преподаватель департамента финансов Санкт-Петербургской школы экономики и менеджмента Санкт-Петербургского филиала, Национальный исследовательский университет «Высшая школа экономики», г. Санкт-Петербург, Россия (190121, г. Санкт-Петербург, ул. Союза Печатников, 16); ORCID https://orcid.org/0000-0002-4041-4465 e-mail: vvafanasev@hse.ru

ДЛЯ ЦИТИРОВАНИЯ

Бухарин Е. О., Мангилева С. И., Афанасьев В. В. Прогноз дефолта для российских предприятий общественного питания: вклад нефинансовых факторов и машинного обучения // Journal of Applied Economic Research. 2024. Т. 23, № 1. С. 206–226. https://doi.org/10.15826/vestnik.2024.23.1.009

ИНФОРМАЦИЯ О СТАТЬЕ

Дата поступления 30 июня 2023 г.; дата поступления после рецензирования 5 декабря 2023 г.; дата принятия к печати 12 января 2024 г.