

Article

Predicting the Performance of Retail Market Firms: Regression and Machine Learning Methods

Darko B. Vukovic ^{1,2}, Lubov Spitsina ³, Ekaterina Griбанова ³, Vladislav Spitsin ^{4,*} and Ivan Lyzin ⁵

¹ Graduate School of Management, Saint Petersburg State University, Volkhovskiy Pereulok 3, 199004 Saint Petersburg, Russia; darko.vukovich@spbu.ru or vdarko@hotmail.rs

² Geographical Institute "Jovan Cvijic" SASA, Djure Jaksica 9, 11000 Belgrade, Serbia

³ Division for Social Sciences and Humanities, School of Engineering Education, National Research Tomsk Polytechnic University, Lenina Avenue, 30, 634050 Tomsk, Russia; spicyna@tpu.ru (L.S.); katag@yandex.ru (E.G.)

⁴ School of Engineering Entrepreneurship, National Research Tomsk Polytechnic University, Lenina Avenue, 30, 634050 Tomsk, Russia

⁵ School of Information Technology and Robotics Engineering, National Research Tomsk Polytechnical University, Lenina Avenue, 30, 634050 Tomsk, Russia; lyzin@tpu.ru

* Correspondence: spitsin_vv@mail.ru

Abstract: The problem of predicting profitability is exceptionally relevant for investors and company owners. This paper examines the factors affecting firm performance and tests and compares various methods based on linear and non-linear dependencies between variables for predicting firm performance. In this study, the methods include random effects regression, individual machine learning algorithms with optimizers (DNN, LSTM, and Random Forest), and advanced machine learning methods consisting of sets of algorithms (portfolios and ensembles). The training sample includes 551 retail-oriented companies and data for 2017–2019 (panel data, 1653 observations). The test sample contains data for these companies for 2020. This study combines two approaches (stages): an econometric analysis of the influence of factors on the company's profitability and machine learning methods to predict the company's profitability. To compare forecasting methods, we used parametric and non-parametric predictive measures and ANOVA. The paper shows that previous profitability has a strong positive impact on a firm's performance. We also find a non-linear positive effect of sales growth and web traffic on firm profitability. These variables significantly improve the prediction accuracy. Regression is inferior in forecast accuracy to machine learning methods. Advanced methods (portfolios and ensembles) demonstrate better and more steady results compared with individual machine learning methods.

Keywords: firm performance; non-linear models of panel data forecasting; retail market companies; profitability prediction; random effects regression; machine learning methods; Random Forest; long short-term memory; deep neural network; portfolio algorithm; ensemble algorithm

MSC: 91B42



Citation: Vukovic, D.B.; Spitsina, L.; Griбанова, E.; Spitsin, V.; Lyzin, I. Predicting the Performance of Retail Market Firms: Regression and Machine Learning Methods. *Mathematics* **2023**, *11*, 1916. <https://doi.org/10.3390/math11081916>

Academic Editor: Alfonso Mateos Caballero

Received: 22 March 2023

Revised: 13 April 2023

Accepted: 16 April 2023

Published: 18 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Profitability is one of the most important indicators for evaluating a company's performance [1,2]. The literature explores a wide range of factors that affect the profitability of a firm, such as leverage and firm size [1–4], working capital management and global crises [5,6], customer relationship management, and innovation [7]. Traditionally, economists have used linear methods (regression) to evaluate and forecast indicators. However, linear methods may not correctly reflect the real relationships between economic indicators. A modern trend is to use machine learning methods (neural networks, Random Forest, and so on), which assume the presence of non-linear hidden dependencies [8].

Several studies have solved the problem of profitability forecasting using machine learning methods, but most of them predict profitability as a binary variable; that is, positive and negative profitability [9], increase or decrease profitability [10], and so on. At the same time, the problem of predicting profitability as an interval variable remains relevant and needs further studies. Only some works carry out calculations based on panel data [11].

The main purpose of this study is to empirically test and compare methods of predicting a firm's profitability and identify the best ones. This paper combines two approaches (stages) to solve the problem of predicting the firms' performance:

- Econometric analysis of the influence of factors on the company profitability. The results of this analysis justify the addition of these factors to the models (sets of variables) of profitability prediction.
- Prediction of the company profitability using machine learning methods. The problem of predicting profitability is exceptionally relevant for investors and company owners. Profitability and sales growth are two main indicators that characterize the success of a business [1,2]. However, sales growth is difficult to model and forecast. As a rule, the proportion of explained variation (R^2) in regression models is low [12]. On the contrary, many studies are devoted to identifying the factors that affect the profitability of the company and model this influence [7,13–15]. This study predicts the firm's profitability as an interval variable for three models (sets of variables) and the following prediction methods: regression with random effects (Regr), individual methods (or algorithms) of machine learning (Random Forest (RF), deep neural network (DNN), long short-term memory (LSTM)), and advanced methods (or sets of algorithms) of machine learning (portfolios (Port1 and Port2) and ensembles (Ens)). We compare these methods and determine the best of them by applying parametric and non-parametric prediction measures, as well as analysis of variance to identify differences in absolute forecast errors. As a novelty of this study, we indicate tuning of hyper parameters with an Adam optimizer for DNN and LSTM and create two portfolios: Portfolio 1 minimizes the mean absolute errors for the training set and Portfolio 2 is balanced by a set of individual methods, minimizing and adjusting the mean absolute errors in applications of different forecasting methods. Such an approach has a managerial implication in forecasting company performance (like in the retail market) and comparing potential portfolios in their decision-making.

A detailed description of these stages and hypotheses of the study is presented below (Section 2).

Calculations are carried out on a sample of 551 companies focused on retail sales of their products. The training sample contains data for 2017–2019 (panel data, 1653 observations). The test sample contains data for these companies for 2020. The dependent (predictable) variable is the net return on assets.

The novelty and distinctive features of this study include the following:

- The study combines the methods of econometric analysis and machine learning. Econometric analysis is used to identify variables that affect the profitability of the company and justify their addition to the prediction model;
- Machine learning methods assume the presence of non-linear latent dependencies. This method analyzes panel data and predicts the economic indicator firm's profitability, which is an interval variable. We build and apply both individual machine learning methods (DNN, Random Forest, LSTM) and advanced methods (portfolios and ensembles) to identify the best ones;
- The methods were tested on three models that differ in the proportion of variance explained (R^2). Accordingly, this paper defines not only the best methods for one model, but also methods that give consistently high accuracy for all models. The study shows that the proposed advanced machine learning methods including algorithm sets (portfolios and ensembles) give better forecasting results in low-predictive-accuracy models. They outperform high-accuracy forecasts and can be recommended for similar

tasks. The study proposes a technique for creating a balanced portfolio adapted for panel data;

- The accuracy of predictive methods is assessed by parametric and non-parametric (rank) prediction measures. The study uses ANOVA to identify differences between absolute forecast errors and to determine the best models and methods for predicting firm's profitability.

2. Research Stages and Hypothesis Development

2.1. Econometric Analysis of the Influence of Factors on the Company Profitability

The literature explores a wide range of factors that affect firm profitability, such as leverage and firm size [1–3], working capital management and global crises [5,6], customer relationship management, and innovation [7]. However, researchers rarely consider last year's profitability as a test variable. Perhaps this variable is not of interest from the standpoint of managing the current profitability of the firm. Indeed, last year's profitability is already a fait accompli, and it is impossible to manage it this year. However, this study is going to show that this variable is exceedingly important in the problems of profitability forecasting and that this variable significantly increases the accuracy of the forecast.

Scholars explain this effect with the following reasons. First, firm profitability converges at a certain level as a result of market competition, which is referred to as persistence of profit (POP). POP studies argue that firm entry and exit are completely free, so any abnormal profit quickly disappears, and that the profitability of all firms tends to converge towards the long-run average value [16,17]. Second, firms try to manipulate profit to achieve the average level in the industry [18,19].

The positive impact of last year's profitability on the firm's current performance is confirmed in the works [11,16]. Accordingly, we test the following hypothesis:

Hypothesis 1. *Last year's profitability has a positive impact on the firm's performance in the current year.*

An important factor affecting the company profitability is the dynamics of sales. Scholars obtain conflicting results when modeling the impact of sales growth on profitability; that is, positive impact [20–22], no impact, or negative impact [23–25]. We believe that this factor leads to increased profitability if the firm does not have to cut prices to increase sales. Therefore, not only sales growth is important, but also the simultaneous development of sales channels and marketing communications with customers. The combination of these factors will enhance the positive impact on profitability. Along with the growth of sales, we studied (modern) digital sales and communication channels with customers (traffic on the company's website) [26–28], and considered the interaction of these factors.

Hypothesis 2. *Sales growth and web traffic have a positive effect on the firm's profitability, which is enhanced by the moderation effect of these variables' interaction.*

If this hypothesis is confirmed, the study will reveal the non-linear influence of these factors on profitability, and we will perform its 3D visualization.

2.2. Predicting the Companies' Profitability Using Machine Learning Methods

This paper predicts firm profitability for three models (three sets of variables) and for the following prediction methods: regression with random effects, individual methods of machine learning (Random Forest, DNN, LSTM), and advanced methods (or sets of algorithms) of machine learning (portfolios and ensembles).

An important issue in forecasting the performance of a firm determines the set of variables that will be included in the forecasting models. Scholars use several approaches:

- Initially, to fix the small set of variables for the predictive model and change only the forecasting methods (regression, machine learning methods, and so on) [8,9,29,30];

- Initially, to include many factors in the model and use all of them in forecasting [10,31].

This study takes a different approach. We have added new factors to the model with control variables if they significantly affect the profitability according to the regression analysis. Next, we evaluate whether the quality of the prediction has improved with the addition of new factors. According to this, we formulate the following hypotheses:

Hypothesis 3.1. *The addition of the “last year’s profitability” variable to the model with control variables improves the accuracy of forecasting the firms’ performance in the current year.*

Hypothesis 3.2. *The addition of the variables “sales growth”, “traffic”, and their interactions (moderation effects, non-linear dependence) to the model improves the accuracy of the firms’ performance forecast.*

Researchers compare the quality of predictive models and methods to determine the best ones and improve the accuracy of the forecast. Some forecasting approaches are associated with machine learning methods and include methods such as DNN [11,32], LSTM [33], Random Forest [34], and so on, which assume the presence of non-linear latent dependencies. The prediction accuracy of these methods is compared to traditional regression methods, which are based on linear dependencies [35]. Moreover, these methods are compared to each other in terms of prediction accuracy. The researchers obtained conflicting results:

- Most works confirm the advantage of machine learning methods (LSTM [36], neural networks [10,37,38], and so on) over regression methods;
- Some works do not reveal significant differences between machine learning methods and regression methods [39] or find that regression methods are better than machine learning methods [40].

We were guided by the prevailing point of view and have formulated the following hypothesis:

Hypothesis 4.1. *Machine learning methods, which assume the presence of non-linear latent dependencies, provide greater accuracy in predicting the profitability of firms compared with traditional regression models.*

Further development of that trend led to the emergence of advanced machine learning methods and approaches: combinations of methods [41,42], ensembles of methods [43–45], and portfolio methods [46,47]. As a rule, such methods include sets of methods or algorithms. This work has tested two advanced methods: the portfolio method and the ensemble method. Portfolios and ensembles were used to improve the accuracy of predictive models. Scholars have obtained conflicting results:

- Most articles show that these methods lead to an increase in the accuracy of forecasts [48–50];
- Some studies do not reveal the benefits of these methods compared with individual machine learning methods [51].

However, these methods have been studied little in relation to panel data. We have explored these approaches for panel data and three models and tested the following hypotheses:

Hypothesis 4.2. *Portfolios improve the accuracy of firm profitability forecasts compared with individual machine learning methods.*

Hypothesis 4.3. *Ensemble improves the accuracy of firm profitability forecasts compared with individual machine learning methods.*

To test hypotheses 3.1, 3.2, 4.1, 4.2, and 4.3, we additionally applied non-parametric (rank) prediction measures (median, 25–75% quartile range of absolute errors and squared errors) and analysis of variance to identify significant differences in absolute profitability prediction errors.

3. Materials and Methods

3.1. Data

The sample of companies consists of 551 industrial and service firms focused on selling their products to retail consumers. The sample includes firms in retail (grocery and electronics supermarkets), the food industry, the IT sector, the automotive industry, residential construction, and other industries. The criteria for inclusion of a firm in the sample are as follows:

- Sales of products of more than 100 million rubles annually during the period from 2016 to 2020;
- Availability of the company's website;
- The firm is focused on the retail consumers.

All firms that met these criteria were included in the sample. Firms with missing values of indicators or exhibiting major outliers were excluded from the study. Companies' financial indicators were sourced from the Spark Information Systems [52]. Web traffic metrics were obtained using the Seranking service [53].

The scope of the study is from 2017 to 2020. According to forecasting techniques, it is divided into two periods:

- The training period (2017–2019). It was used to train models and identify relationships between variables. It is panel data and includes 1653 observations (551 firms \times 3 years). We lost one year (2016) of observations as we calculated the growth rates of sales and used the "last year's profitability" variable;
- The test period (2020). It was used to predict the profitability of firms. It includes 551 observations (551 firms \times 1 year).

3.2. Econometric Analysis of the Influence of Factors on the Profitability of the Company

Dependent Variable. We consider the company's net return on assets (ROA) as a dependent variable that characterizes the efficiency of the enterprise. This approach to measuring firm performance is widely used in modern economic research [54–57]. ROA is calculated as the ratio of net profit to the firm's assets, multiplied by 100%.

Independent Variables. In accordance with the purpose of the work and the formulated hypotheses, we examined the effect of three independent variables on firm profitability:

- Last year's profitability (ROA $t - 1$) [11,16];
- Sales growth (growth) measured as the ratio of difference in revenue between years t and $(t - 1)$ to revenue in year $(t - 1)$ [20–22];
- Company website traffic (traffic) obtained from the Seranking service [53]. Given the nature of the data, we utilized the natural logarithm of the variable in our modelling [58,59].

Control Variables. In accordance with the generally accepted methodology of econometric calculations, we included a wide range of control variables, which can affect the dependent variable in our regression models, to control alternative explanations:

- Share of fixed assets in total assets (FATA) [56,60];
- Current liquidity ratio (CACL) measured as the ratio of current assets to current liabilities and controls for company's ability to launch and sustain capital-intensive initiatives [60];
- Leverage (leverage) calculated as the share of borrowed funds in the assets of the company [3,61];
- Asset turnover (turnover) measured as the ratio of revenue to the company's assets serves as a control for company's efficiency to generate sales [62];

- Firm’s age (age) measured as a number of years since company’s establishment according to SPARK database [61,63];
- Mean_ind variable, which reflects differences in firm performance across industries and years.

Descriptive statistics and correlations between variables are presented in Table 1.

Table 1. Descriptive statistics and correlations between variables.

N	Variables	Mean	St. Dev.	1	2	3	4	5	6	7	8
1	FATA	18.90	18.66	1.00							
2	CACL	3.39	10.42	−0.06 *	1.00						
3	Leverage	56.03	27.58	−0.10 ***	−0.30 ***	1.00					
4	Turnover	166.64	112.29	−0.11 ***	−0.12 ***	0.17 ***	1.00				
5	Age	18.05	6.73	0.12 ***	0.03	−0.16 ***	−0.20 ***	1.00			
6	Mean_ind	8.21	3.48	−0.10 ***	0.06 *	−0.20 ***	−0.02	0.04 ^λ	1.00		
7	ROA t − 1	8.11	11.64	−0.05 *	0.18 ***	−0.35 ***	0.05 *	−0.09 ***	0.25 ***	1.00	
8	Growth	0.12	0.37	−0.05 *	−0.02	0.09 ***	0.07 **	−0.12 ***	0.05 *	−0.04	1.00
9	Traffic	8.26	3.37	0.07 **	−0.03	0.01	0.00	0.08 ***	−0.03	0.03	−0.08 **

Source: calculated by the authors. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; ^λ $p < 0.10$.

The data in Table 1 show that there is no strong correlation between the predictor variables ($r \ll 0.70$); therefore, we can use them in the regression analysis. The correlation between last year’s and current year’s profitability is also less than 70%, so we include last year’s profitability in our models (a similar approach was used in [16], in which the same variable is included in the OLS and fixed-effects models).

3.3. Models

This research applies regression analysis to the panel data. The regression model based on the least squares method (OLS model) is evaluated as inadequate. For panel data analysis, either fixed-effects or random-effects models are commonly used. In this study, we choose random-effects models.

The general formula for a regression model with random effects is as follows:

$$Y_{it} = \text{Intercept} + X_{it} \times \beta + \mu_i + \epsilon_{it}, \tag{1}$$

where

- Intercept—a constant;
- X_{it} —variables and β —coefficients for variables;
- μ_i —a random error invariant in time for each object;
- ϵ_{it} —the model regression residual.

The models are presented in Table 2. In all models, ROA is the dependent variable.

Model 1 and Model 2 are linear, while Model 3, which includes the multiplication of variables, is non-linear.

According to Table 2, the formula of the regression model with random effects for Model 1 is as follows:

$$\text{ROA} = \text{Intercept} + \beta_1 \times \text{FATA} + \beta_2 \times \text{CACL} + \beta_3 \times \text{Leverage} + \beta_4 \times \text{Turnover} + \beta_5 \times \text{Age} + \beta_6 \times \text{Mean_ind} + \mu_i + \epsilon_{it} \tag{2}$$

Formulae for Models 2 and 3 are constructed similarly according to the data in Table 2.

Model 1 includes only control variables. Model 2 adds last year’s profitability to estimate its impact on the current year’s firm performance and test hypothesis 1. Model 3 adds sales growth, web traffic, and their interaction to test hypothesis 2.

Table 2. Regression models and their variables.

N	Variables	Model 1	Model 2	Model 3
1	FATA	+	+	+
2	CACL	+	+	+
3	Leverage	+	+	+
4	Turnover	+	+	+
5	Age	+	+	+
6	Mean_ind	+	+	+
8	ROA $t - 1$		+	+
9	Growth			+
10	Traffic			+
11	Growth \times Traffic			+

To minimize the problems of multicollinearity, all independent and control variables of regression models were standardized according to [64]. The regression analysis period was from 2017 to 2019.

3.4. Predicting the Companies’ Profitability Using Machine Learning Methods

The study uses the same data to predict profitability. We built predictive models on three sets of variables, which are shown in Table 2. For each set of variables, we applied the following prediction methods:

- regression with random effects (Regr);
- individual methods (or algorithms) of machine learning (Random Forest (RF), deep neural network (DNN), long short-term memory (LSTM));
- advanced methods (or sets of algorithms) of machine learning (portfolios (Port1 and Port2) and ensembles (Ens)).

These methods were trained on 2017–2019 data and then predicted profitability in 2020. Unlike regression, the machine learning methods consider the presence of non-linear latent dependencies when predicting a firm’s profitability.

A deep neural network (DNN) is a neural network model that sequentially calculates an output layer based on an input layer, using the outputs of the current layer as the inputs of the next one [65]. Hidden layers allow this method to non-linearly transform input data and approximate complex relationships between inputs and outputs. Thus, neural networks allow us to approximate arbitrary non-linear dependencies, the form of which is not known in advance. The structure of the neural network includes an input layer represented by the model input data (size, FATA, and so on) described earlier; two identical hidden layers, each of which contains 64 neurons; and an output layer that includes one neuron predicting profitability (Figure 1).

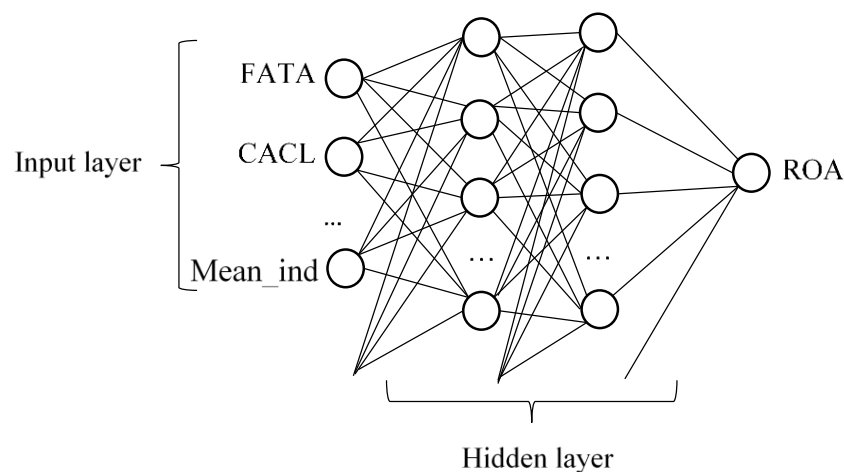


Figure 1. Structure of the deep neural network (DNN).

The calculation of profitability is carried out according to the following formula:

$$ROA = f \left(\sum_{l=1}^{n2} \left(f_l \left(\sum_{j=1}^{n1} \left(f_j \left(\sum_{i=1}^n (x_i w_{ji} + b_j) \right) \right) \right) \right) w_{lj} + b_l \right) w_l + b \tag{3}$$

where x are normalized values of input variables;

w are weight coefficients;

b is bias;

n is the number of input variables;

$n1$ is the number of neurons in the first intermediate layer;

$n2$ is the number of neurons in the second intermediate layer.

ReLU was used as the activation function of the intermediate layers:

$$f(u) = \begin{cases} 0, & \text{if } x < 0 \\ u, & \text{otherwise} \end{cases}$$

where $f(\cdot)$ is an activation function and u is the value calculated by summing the bias and multiplying the input values of the neuron by weight coefficients.

The activation function of the output neuron is linear as the regression problem is being solved.

In the process of training a neural network, the parameters (weights and biases) are determined by minimizing the loss function, which is usually expressed by characteristics such as MSE and MAE.

The recurrent neural network LSTM (long short-term memory) has gained wide popularity for time-series forecasting [8,33,38]. The advantages of the LSTM model include the ability to effectively build non-linear dependencies that describe the source data with high accuracy. Unlike DNN, the LSTM block does not consist of a neuron, but of a memory cell that stores information updated by three gates: input, forgetting, and output. The advantages of such a structure include the ability to process long-term information. To work with LSTM, the input data were converted to a many-to-many sequence, where the timestep is three, the number of features is equal to the number of variables in the model, and the number of samples is equal to the number of enterprises. The network structure also includes two intermediate layers (the number of neurons in the inner layer is 100) as well as one output layer. The algorithm applies a ReLU activation function for the inner layers and a linear activation function for the output layer.

To predict the profitability of firms, we use DNN and LSTM with the advanced and efficient Adam optimizer. The optimizer in DNN or LSTM is an algorithm that allows the model to find the optimal values of weights and biases during the training process. To train DNN and LSTM, we considered optimization algorithms such as Adam, AdaGrad, and RMSProp. The best performance was obtained using the Adam algorithm, as it is one of the most efficient optimization algorithms for training [66–69]. It combines the ideas of RMSProp and AdaGrad. When using this algorithm, the weight coefficients are updated according to the following formulae [70]:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t, \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \\ w_t &= w_{t-1} - \frac{\eta}{\sqrt{v_t + \epsilon}} m_t, \end{aligned}$$

where:

t —timestep;

η —learning rate ($\eta = 0.001$);

β_1, β_2 —exponential decay rates for the moment estimates ($\beta_1 = 0.9, \beta_2 = 0.999$);

g —objective function gradient;

ϵ —smoothing parameter to avoid division by zero ($\epsilon = 10^{-7}$);

w —weight coefficients;
 m —first moment estimate;
 v —second raw moment estimate.

We used MSE as a loss function, as it is the most popular and has several advantages, including sensitivity to outliers, which ensures the exclusion of large deviations of the predicted value from the real one [71,72]. In addition, this function is used when building a regression.

Random Forest is a machine learning algorithm that uses an ensemble of decision trees to solve a problem [73]. Like neural networks, this tool can detect non-linear relationships between a dependent variable and a set of independent variables. A decision tree is a way of representing if-then decision rules in a specific hierarchy. Thus, the resulting structure includes elements of two types: nodes (decision rules) and leaves (decisions). To build a tree, it is necessary to specify the minimum number of samples to split the node and the maximum depth of the tree. To solve the problem, values 2 and 7 were chosen, respectively. Random Forest includes 100 decision trees. The MSE function was used to evaluate the separation quality.

Portfolio methods are sets of individual methods or algorithms. As a rule, if there are several methods under study, one is chosen for forecasting by generalizing metrics (MAE, MSE, R^2) calculated on the training set [41]. However, this choice is not always effective, because the method that was the best on the training set may show moderate results or be the worst on the test set. For panel data, the same method may provide the smallest error for one object, but the largest error for another. To solve this problem, scholars apply the portfolio method, where a different forecasting method is chosen for each object [47,74,75].

We modified the portfolio approach presented in the literature taking into account panel data. Two portfolios were formed:

1. Portfolio with error minimization (Port1). This portfolio is aimed at minimizing the mean absolute errors for the training set. The algorithm for constructing this portfolio is shown in Figure 2.

Here and after, i is the firm number, n is the number of years in the training sample, k is the method number, j is the year number, e is the mean absolute error, h is the model number, ROA' is the predicted value of profitability based on the training samples, and ROA'' is the predicted value of profitability based on the test sample.

To predict profitability, this technique selects the method that provides the smallest value of the average absolute error from 2017 to 2019 for each firm.

The advantage of this approach is that it focuses on minimizing absolute forecast errors. The disadvantage is that an unbalanced portfolio can be obtained with the predominance of one of the methods. The MAE value can vary greatly between methods on the training set. At the same time, there were no such differences observed in the test sample. Inadequate differences in MAE on the training sample lead to the fact that the method with the smallest MAE is chosen for most objects and the method with the largest MAE is chosen for a minority of firms.

This paper offers a second portfolio option that solves this problem on panel data.

2. Portfolio balanced by a set of individual methods (Port2). This portfolio is aimed at minimizing the mean absolute errors adjusted for MAE differences between methods on the training set. To predict profitability, we chose the method that provides the smallest value of the average absolute error from 2017 to 2019 for a given company, divided by the MAE of the method, calculated from the training sample. In this case, the average absolute error is calculated by the following formula:

$$e_{ik} = \frac{\sum_{j=1}^n |ROA_{ij} - ROA'_{ijk}|}{n \cdot mae_k}.$$

Such a portfolio provides a greater variety of methods in the case when the accuracy of the methods on the training set is different.

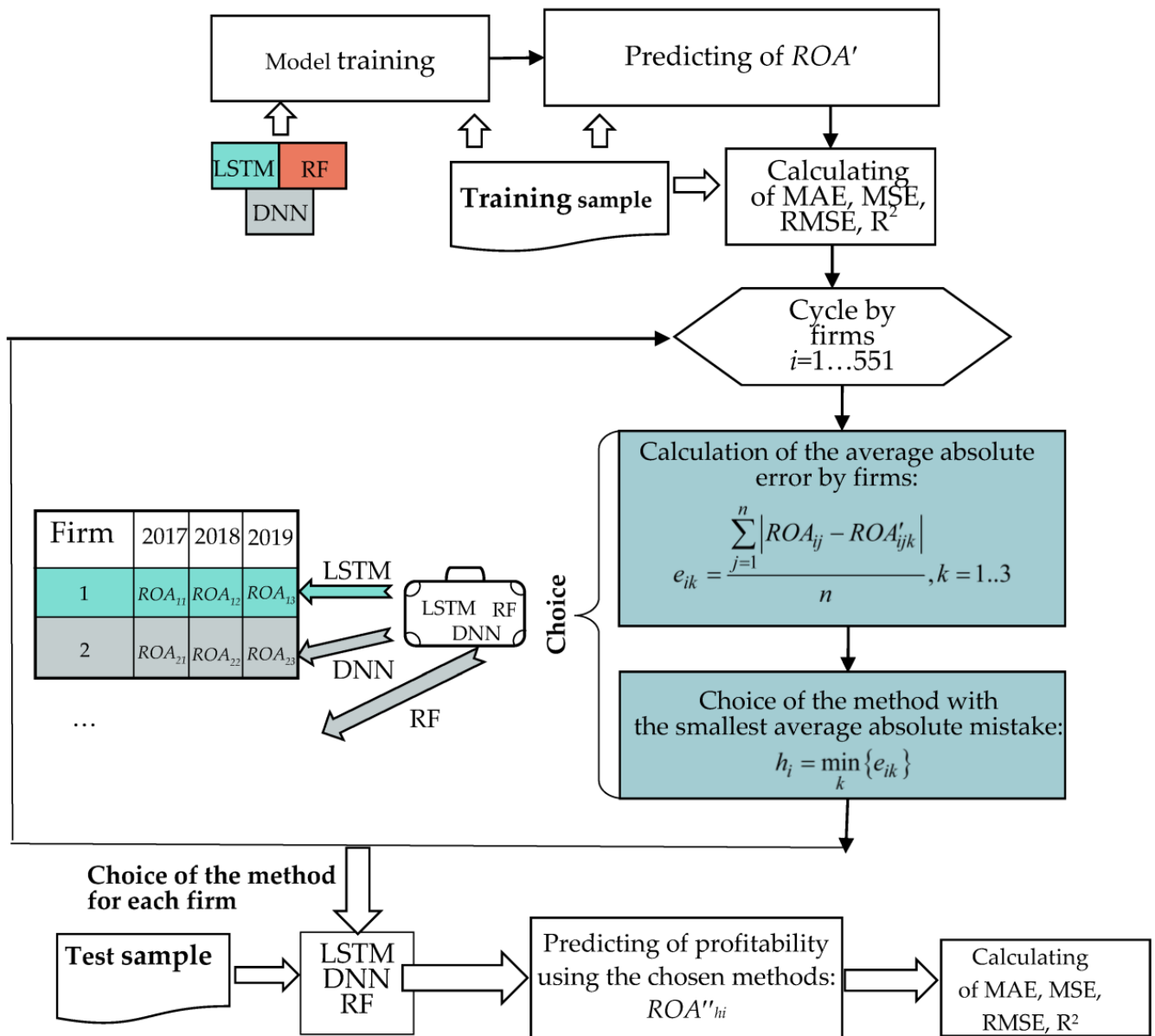


Figure 2. Forecast using a portfolio with minimization of training sample errors.

Ensembles are sets of individual methods or algorithms that are grouped at several levels [76]. Ensembles of algorithms can reduce the prediction error compared with single models by combining the predictions of several models, as well as increase resistance to outliers, noise, and data errors and provide better generalization ability [77]. It allows us to improve the quality metrics of individual algorithms.

This article discusses the construction of an ensemble of algorithms for predicting firm profitability. We propose the following technique for constructing the ensemble of algorithms structure. The method represents a two-level ensemble. The first level consists of gradient boosting methods, which are formed according to [78]. Feature selection is based on the significance of the features estimated for these gradient boosting methods in accordance with the methodology [79]. The study applies the hyperparameter optimization technique to tune the parameters of the methods at the first level. The second level consists of a linear regression that is applied to the features identified at the first level. Learning occurs through cross-validation [80]. A graphical representation of the proposed ensemble is shown in Figure 3.

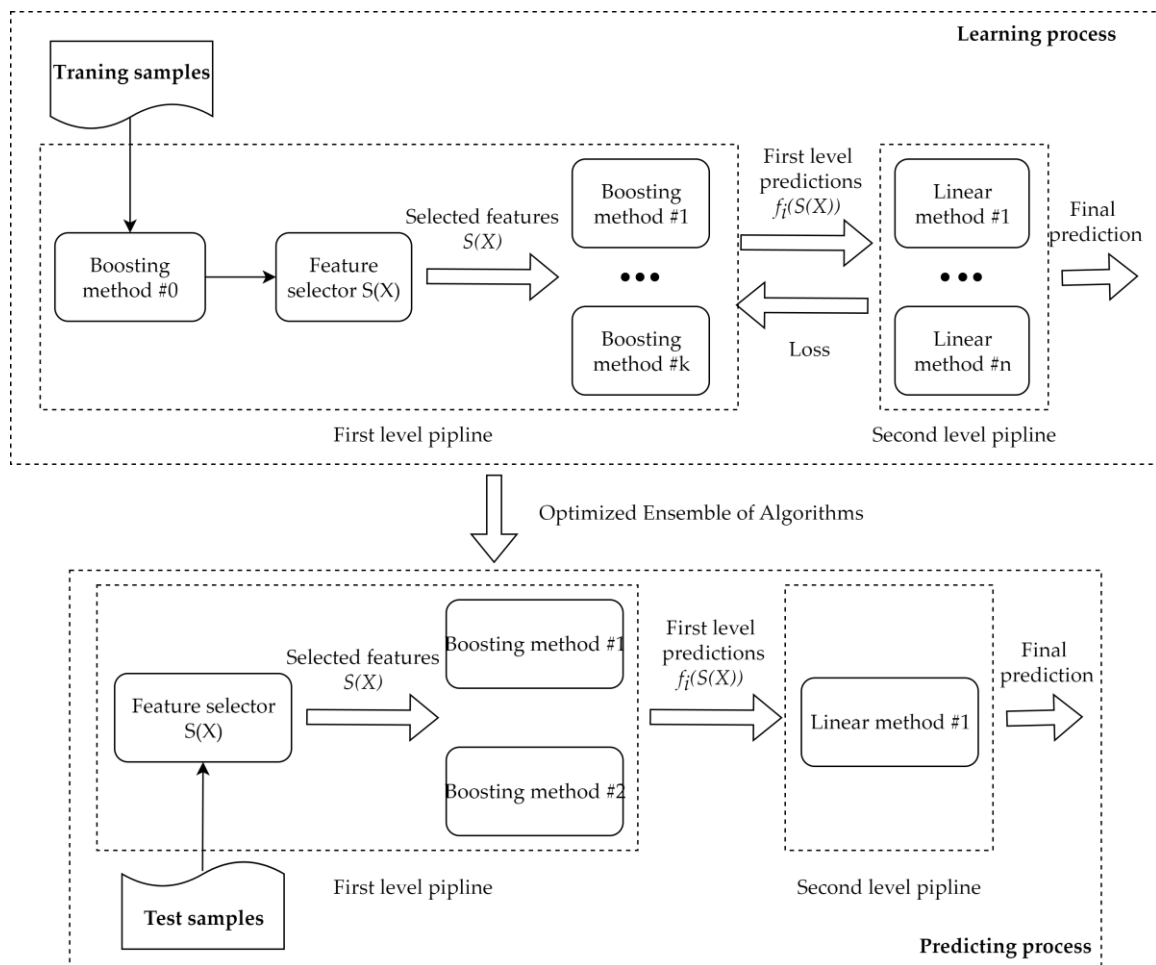


Figure 3. Ensemble of gradient boosting and linear regression methods to predict firm profitability.

The presented ensemble can be described as follows:

X is a feature matrix of dimension $N \times M$, where N is the number of observations and M is the number of features.

The pipeline used at the first level of the model consists of the following elements:

A feature selector $S(X)$ that selects the most important features from a matrix X based on a given threshold and a gradient boosting method.

Gradient boosting methods at the first level are denoted as $f_i(S(X))$, where $i = 1, \dots, k$ and k are numbers of gradient boosting methods ($k = 2$ in our case), which accept selected features and return predictions y_i for each observation.

The second-level pipeline of the model is as follows:

Second-level linear regression methods, denoted as $g_j(y_1, \dots, y_k)$, where $j = 1, \dots, n$ and n are numbers linear regression methods ($n = 1$ in our case), which accept the predictions of the first-level models and return the final predictions y_j for each observation.

A loss function $L(y_j, y_{true})$ measures the error of the model based on the final predictions y_j and the true values y_{true} .

The final model is defined as an ensemble of second-level methods trained on cross-validation and minimizing the error function:

$$\operatorname{argmin}_{g_1, \dots, g_n} \left(\frac{1}{k} \sum_{i=1}^k L \left(\sum_{j=1}^n w_j g_j (f_i(S(X_i)), \dots, f_k(S(X_i)), y_i) \right) \right)$$

where X_i is the feature matrix for the i -th observation, y_i is the true value for the i -th observation, and w_j is the weight of the j -th method of the second level.

To compare the quality of the above predictive models and methods, we calculate a standard set of parametric indicators (MAE, MSE, RMSE, and R²), according to [81,82]:

$$\begin{aligned}
 \text{MSE} &= \frac{\sum_{i=1}^m (ROA_i - ROA_i'')^2}{m}, \\
 \text{MAE} &= \frac{\sum_{i=1}^m |ROA_i - ROA_i''|}{m}, \\
 \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^m (ROA_i - ROA_i'')^2}{m}}, \\
 \text{R}^2 &= 1 - \frac{\sum_{i=1}^m (ROA_i - ROA_i'')^2}{\sum_{i=1}^m (ROA_i - RO\bar{A})^2},
 \end{aligned}$$

where *m*—number of firms in the sample;

ROA—average value of profitability (ROA) for the sample.

We do not calculate MAPE and SMAPE, as profitability can take values close to zero and these indicators are inadequate [83–85]. Moreover, we use non-parametric (rank) indicators (median, 25–75% quartile range of absolute errors and squared errors) and analysis of variance (Wilcoxon test) to assess the significance of differences in the absolute error of the profitability forecast.

4. Results

4.1. Econometric Analysis of the Influence of Factors on the Profitability of the Company

The results of regression modeling are presented in Table 3.

Based on the calculations (Table 3), Model 1 is described by the following formula:

$$\begin{aligned}
 \text{ROA} = & 8.21 - 0.99 \times \text{FATA} + 0.09 \times \text{CACL} - 5.43 \times \text{Leverage} + 2.21 \times \\
 & \text{Turnover} - 1.49 \times \text{Age} + 2.50 \times \text{Mean_ind} + \mu_i + \epsilon_{it}
 \end{aligned} \tag{4}$$

Formulae for Models 2, 3, and 4 are constructed similarly according to the data in Table 3.

Model 1, which includes only control variables, is highly significant, but it only explains a small percentage of the variation in the dependent variable (R² is 19.8%). Turnover and mean_ind have a stronger positive effect on the dependent variable. Leverage and age highly significantly negative impact on profitability.

The second model adds the ROA *t* – 1 variable, which has a highly significant positive effect on profitability. Model 2 is highly significant and better explains the variation in the dependent variable. R² increases to 41.8%. Hypothesis #1 is confirmed. As R² increases strongly, we expect the addition of the ROA *t* – 1 variable to significantly improve the accuracy of the 2020 profitability forecast.

The third model adds the growth and traffic variables, and their interaction (growth × traffic). These three variables have a significant positive effect on the firm’s profitability. Model 3 is highly significant, with R² rising to 45.2%. Hypothesis #2 is confirmed. The influence of growth and traffic variables on profitability is non-linear and its 3D visualization is required. In model 3, all variables (except the dependent variable) are standardized. Therefore, their average values are equal to zero. Considering the case where all variables take on averages, we obtain the following non-linear function for Model 3:

$$\text{ROA} = 8.26 + 2.43 \times \text{Growth} + 0.37 \times \text{Traffic} + 0.63 \times \text{Growth} \times \text{Traffic} \tag{5}$$

The 3D visualization of the non-linear function (5) is shown in Figure 4.

Table 3. Regression results (training sample, dependent variable—ROA, standard errors are shown in parentheses).

Variables	Model 1	Model 2	Model 3
Intercept	8.21 *** (0.36)	8.21 *** (0.22)	8.26 *** (0.21)
FATA	−0.99 ** (0.34)	−0.32 (0.23)	−0.26 (0.22)
CACL	0.09 (0.31)	0.15 (0.23)	0.14 (0.23)
Leverage	−5.43 *** (0.35)	−2.50 *** (0.25)	−2.62 *** (0.25)
Turnover	2.21 *** (0.33)	1.44 *** (0.23)	1.34 *** (0.22)
Age	−1.49 *** (0.36)	−0.76 *** (0.23)	−0.58 ** (0.22)
Mean_ind	2.50 *** (0.34)	1.65 *** (0.23)	1.47 *** (0.22)
ROA t − 1		5.42 *** (0.24)	5.53 *** (0.24)
Growth			2.43 *** (0.24)
Traffic			0.37 ^λ (0.22)
Growth × Traffic			0.63 ** (0.21)
Adj. R ²	0.198	0.418	0.452
Δ Adj.R ²	-	0.220	0.034
p	<0.001	<0.001	<0.001

Source: calculated by the authors according to SPARK and Seranking data. *** $p < 0.001$; ** $p < 0.01$; ^λ $p < 0.10$.

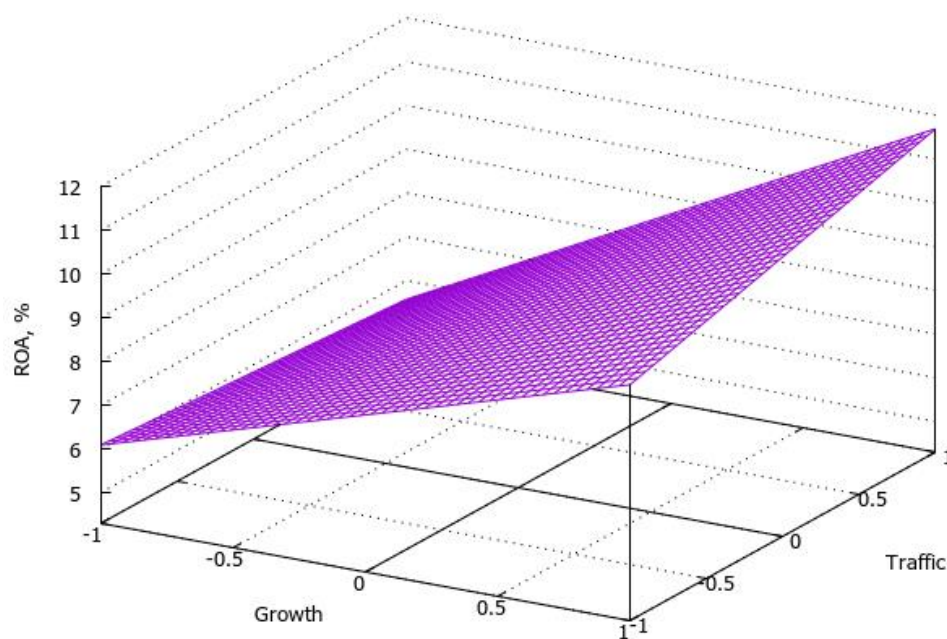


Figure 4. Combined impact of growth and traffic on profitability in model 3 (training sample). ROA is measured as a percentage, while growth and traffic are standardized values.

Figure 4 shows that firms maximize their profitability when two conditions are met: high sales growth and high website traffic. Fulfillment of only one of the conditions leads to a decrease in profitability. The worst option is the case of falling sales with high traffic to the company’s website. A detailed discussion of these results is provided in the “Discussion” section.

4.2. Predicting the Companies’ Profitability Using Machine Learning Methods

Figure 5a,b show the actual and predicted profitability for 551 firms for Models 1 and 3, respectively. These figures demonstrate the prediction accuracy for the advanced machine learning methods (Portfolio 1 and Portfolio 2 in our case) that assume non-linear relationships between variables.

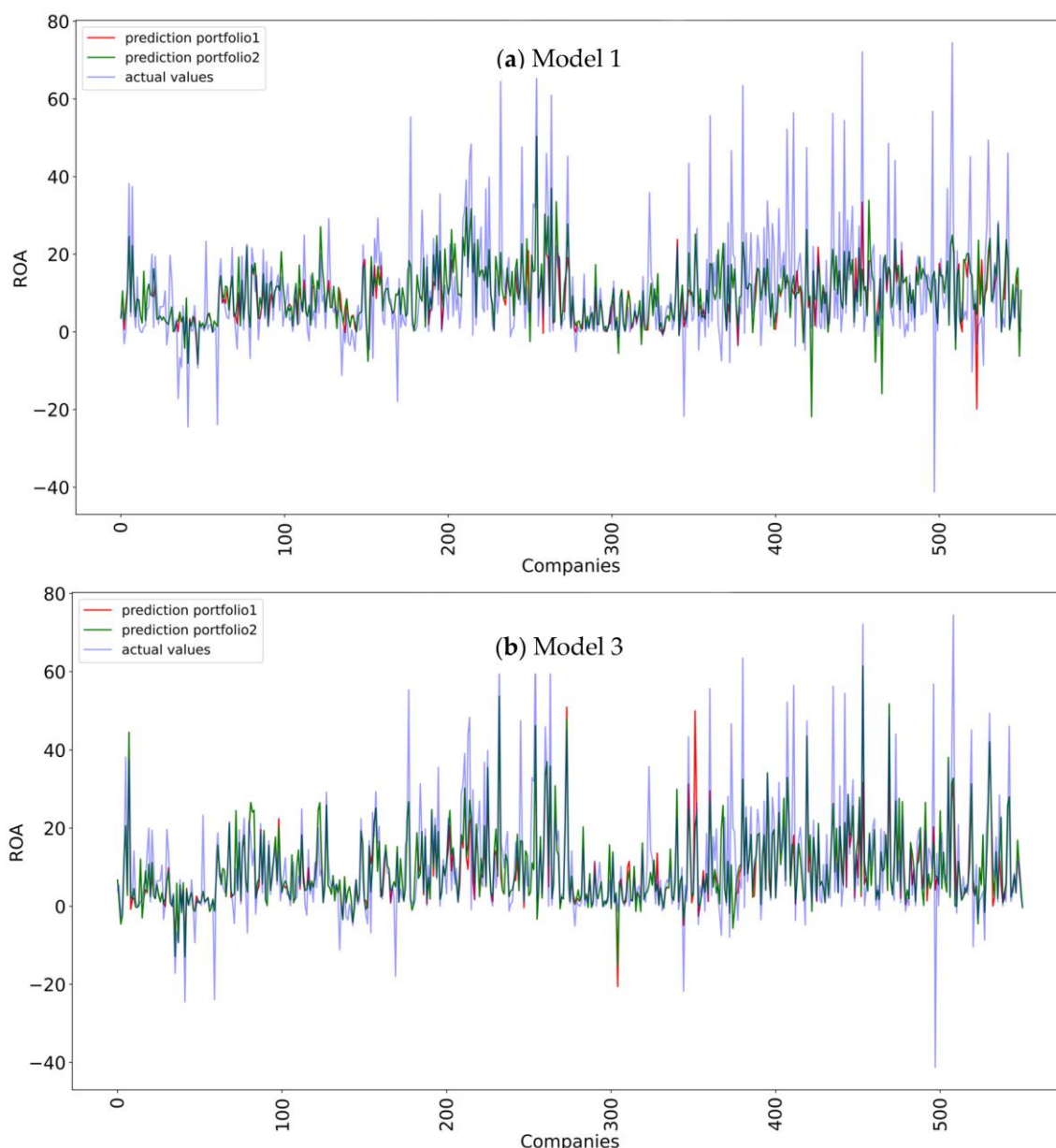


Figure 5. Actual and predicted profitability values using Portfolio 1 and Portfolio 2 for the test sample: (a) Model 1 and (b) Model 3.

Visually, Model 3 achieves a higher prediction accuracy than Model 1. To prove this, it is necessary to compare the prediction measures. The features’ importance of factors influencing the profitability of firms are presented in Figure 6.

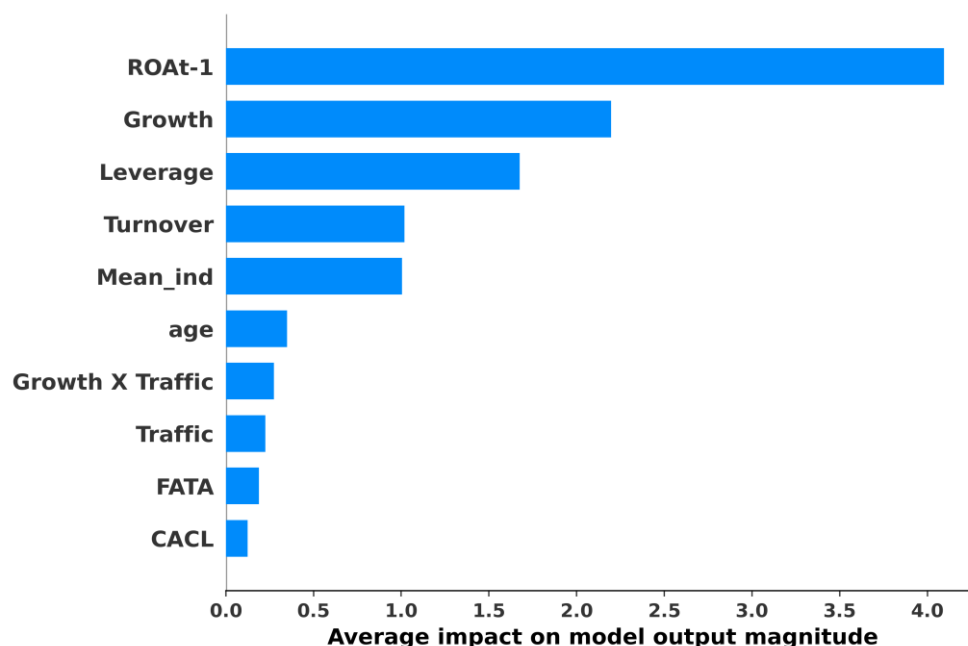


Figure 6. Features’ importance of factors affecting the profitability of firms for the DNN method and Model 3 (test sample).

ROA $t - 1$, leverage, growth, mean_ind, and turnover are the most significant factors to predict a firm’s performance for the case of using a neural network and a full set of variables (Model 3).

We calculated parametric indicators of forecast accuracy to compare the quality of models and methods for predicting profitability (Table 4).

Table 4. Comparison of prediction measures of different models for the test sample.

Method	MAE	MSE	RMSE	R ²
Model 1				
Regression (Regr)	7.90	138.19	11.76	0.28
LSTM	8.67	148.66	12.19	0.23
DNN	7.67	128.51	11.34	0.33
Random Forest (RF)	7.70	132.93	11.53	0.30
Portfolio 1 (Port1)	7.17	116.71	10.80	0.41
Portfolio 2 (Port2)	7.21	116.32	10.79	0.40
Ensemble (Ens)	7.34	126.09	11.23	0.34
Model 2				
Regression (Regr)	6.43	98.30	9.91	0.49
LSTM	6.16	101.73	10.09	0.47
DNN	6.20	97.94	9.90	0.49
Random Forest (RF)	6.29	100.92	10.05	0.47
Portfolio 1 (Port1)	6.11	98.98	9.95	0.52
Portfolio 2 (Port2)	6.12	100.49	10.02	0.50
Ensemble (Ens)	6.09	93.22	9.65	0.52
Model 3				
Regression (Regr)	6.13	86.62	9.31	0.55
LSTM	5.71	80.10	8.95	0.58
DNN	5.52	70.56	8.40	0.63
Random Forest (RF)	5.58	82.59	9.09	0.57
Portfolio 1 (Port1)	5.59	80.66	8.98	0.60
Portfolio 2 (Port2)	5.53	77.14	8.78	0.61
Ensemble (Ens)	5.46	77.57	8.81	0.60

Visually, these data show the following:

- Strong differences in the values of prediction measures are observed between Models 1, 2, and 3. The first model has a low quality of the forecast. The third model demonstrates the highest forecast accuracy.
- Differences between prediction methods are small. The best methods differ between models. Advanced methods (portfolios and ensembles) demonstrate consistently high prediction accuracy. Regression achieves good results and is not far behind the best machine learning models.

However, are these conclusions, based on a visual comparison of the means of prediction measures, reliable? Can we mathematically prove, for example, that Model 2 is better than Model 1 and that the DNN is better than regression? To solve this problem, we applied analysis of variance. We compared the absolute error values for the sample of firms under study for various models and methods using the Wilcoxon test.

Visualization of absolute errors for various models and methods for predicting profitability is shown in Figure 7.

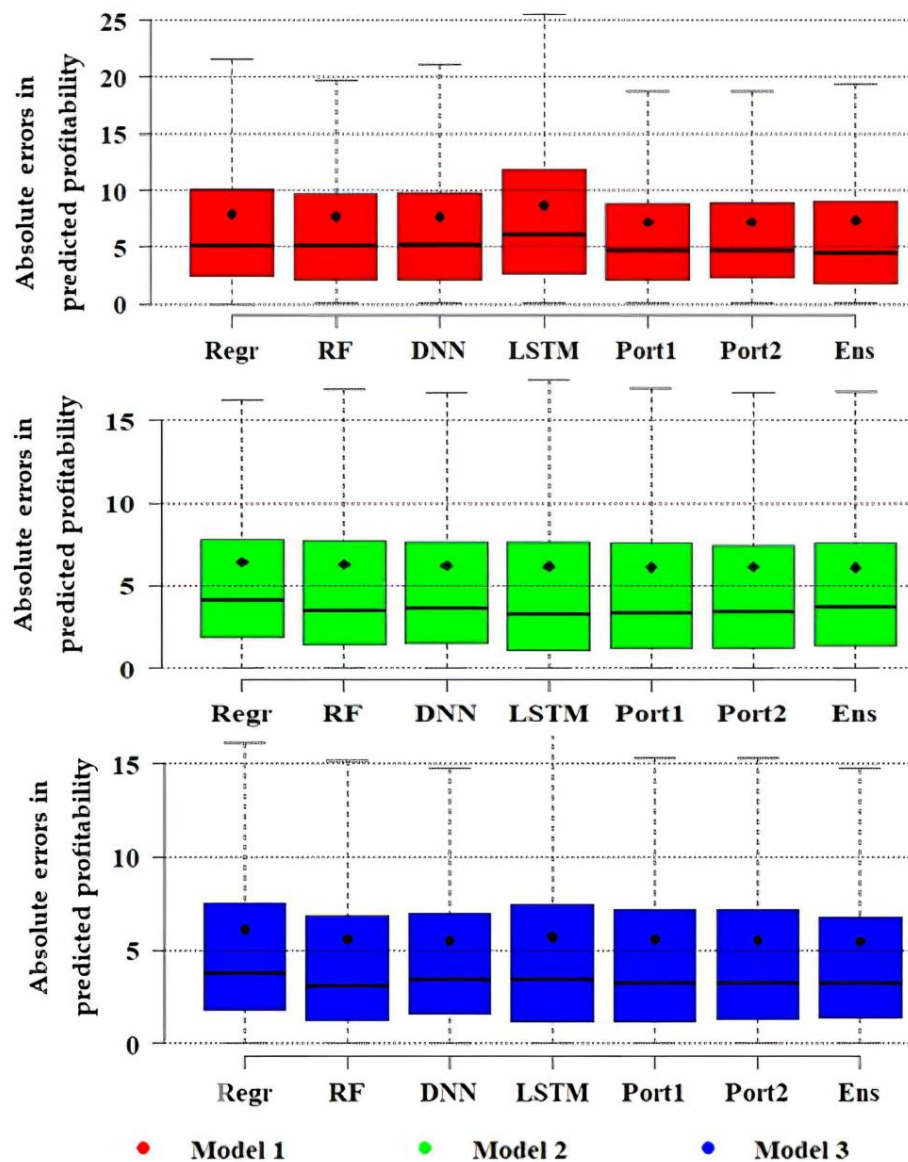


Figure 7. Boxplots of absolute errors for various models and methods for predicting profitability on the test sample. Hereinafter, point—mean value, line—median, rectangle—25–75% quartile range, whiskers—minimum and maximum values or 1.5 interquartile range.

Model 3 with a full set of variables achieves a good profitability forecast accuracy, especially for the best machine learning methods based on non-linear dependencies. MAE drops to 5.46 for ensemble. The median of absolute errors is even lower—3.07 for Random Forest, i.e., the error in predicting the profitability of 50% of firms does not exceed 3.07. Moreover, the absolute error of the lower quartile is 1.20 for Portfolio 1. This means that the error in predicting the ROA of 25% of firms is less than 1.20.

Large differences in the mean values (parametric characteristic of the sample) and medians (nonparametric characteristic of the sample) of the absolute errors are revealed in Figure 7. These differences raise the question of what researchers should prefer when assessing the accuracy of the forecast: parametric criteria or non-parametric criteria.

Analysis of variance (Wilcoxon test) of absolute errors showed the following results:

- Highly significant ($p < 0.001$ ***) differences between Models 1 and 2 for all forecasting methods. Hypothesis 3.1 is confirmed.
- Highly significant ($p < 0.001$ ***) differences between Models 2 and 3 for all forecasting methods, except for DNN. In the case of DNN, the differences are statistically significant ($p = 0.02$ *). Hypothesis 3.2 is confirmed.
- Regression as a prediction method, based mainly on linear dependencies, is inferior to the best individual methods of machine learning, which consider the presence of nonlinear hidden dependencies. For Model 1, Random Forest and DNN are slightly significantly ($p < 0.1$ λ) better than regression. For Model 3, Random Forest, DNN and LSTM are highly significantly ($p < 0.001$ ***) better than regression. Hypothesis 4.1 is confirmed, except for the case of LSTM in Model 1.
- Advanced methods that include sets of algorithms show better and more constant results than individual methods. For Model 1, portfolios and ensembles are better than DNN, RF, and LSTM (most of the differences are highly significant). For Model 2, they are better than DNN and RF (the differences are significant or slightly significant). For Model 3, they are slightly better than LSTM. Portfolios and ensembles achieve constant results for all three models. The differences between them are insignificant for all three models. Hypotheses 4.2 and 4.3 were fully confirmed for Model 1 and were partially confirmed for Models 2 and 3.

Portfolios showed good results, while two techniques for their construction were used in the study. Figure 8 shows the differences in portfolio structure for these techniques.

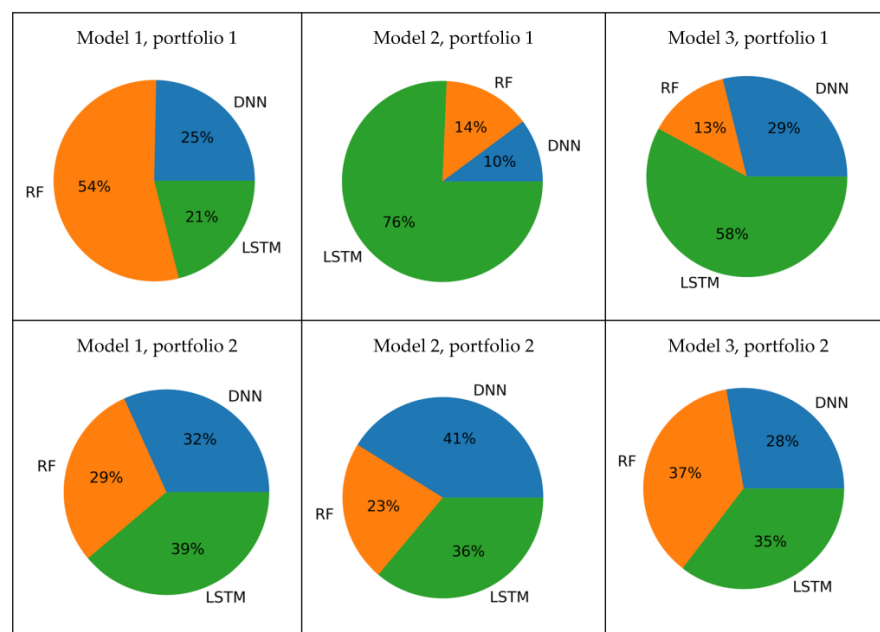


Figure 8. Portfolio structure depending on the technique of their construction. Portfolio structure is the same for the training and test sets.

The first portfolio, which is aimed at minimizing the absolute errors on the training sample, is unbalanced. One of the methods (RF or LSTM) strongly dominates in this portfolio, as its MAE on the training set is much lower than that of other methods. However, the RF and LSTM methods do not show the best results on the test set. Accordingly, such a portfolio structure is risky, although this portfolio shows a high accuracy of forecasts.

The technique of constructing the second portfolio provides a balanced portfolio for the structure of the machine learning methods included in it. This portfolio demonstrates high forecast accuracy and appears to be more reliable than Portfolio 1, which is heavily dominated by the RF and LSTM methods.

A detailed discussion of these results is provided in the next section.

5. Discussion and Practical Implementation

The results of the study are consistent with several studies and their findings, and significantly complement and refine them. The strong positive impact of last year's profitability on the current year's firm's performance is consistent with [11,16]. Scholars explain this effect with the following reasons. Persistence of profit studies argue that firm entry and exit are completely free, so any abnormal profit quickly disappears, and that the profitability of all firms tends to converge towards the long-run average value [16,17]. According to the papers [18,19], firms can also manipulate profit to achieve average level at the industry. Our study has confirmed these results and revealed a strong positive impact of the profitability of past years on the profitability of the current year, which must be considered for the forecasting.

The positive impact of sales growth on profitability was confirmed in some works [20–22], but was disputed in others [23–25]. This study solves the issue of conflicting results when it models the combined impact of sales growth and digital customer communications on a firm's profitability. We have found that firms achieve maximum profitability gains when two conditions are met: high sales growth and the development of modern digital sales channels and customer communications. The resulting dependence is non-linear. This study performs its 3D visualization. According to 3D visualization, fulfillment of only one of the conditions leads to a decrease in profitability. The worst situation is the case of falling sales with high website traffic.

Adding new variables and refining the model improve profitability prediction accuracy. Adding the variable "last year's profitability" to the forecast model significantly improves the accuracy of the forecast. We believe that this variable should be included in all firm profitability forecasting models. A similar approach is used in [11], where scholars also obtain high prediction accuracy. Moreover, we as well as other researchers believe that lagged profitability is the most significant factor in predictive models. The addition of the growth and traffic variables also improves prediction measures for all forecasting methods. Growth is one of the significant factors affecting profitability according to the calculated features' importance.

Regarding the other variables used in the study, we noted the following. Two variables (leverage and turnover) that managers of firms can influence have a pronounced effect on profitability according to the regression model and on the accuracy of predicting profitability according to the calculated features importance. This is consistent with the works of [3,61,62]. Other variables (age, FATA, and CALC) have less influence on the accuracy of the forecast according to the calculated features' importance. In future research, we are going to explore whether they should be included in profitability prediction models. As mentioned before in the study, scholars investigated the impact of a wide range of variables on the profitability of the firm [1–7]. We did not strive to explore a wide range of variables, but sought to find variables that significantly improve the accuracy of the forecast. In particular, the study does not include the firm size variable in the model, as it does not significantly affect profitability in Model 3. However, we are going to consider it when modeling the profitability of firms in other industries. Variables such as export earnings, innovative activity, ownership structure, and so on are of undoubted interest

for researchers. Nevertheless, we have limitations on the available data. Spark provides information primarily on the financial indicators of firms and allows testing of only a limited range of variables.

Many studies apply machine learning methods in forecasting, including profitability prediction. Furthermore, most of these works use a binary dependent variable [9,10]. Only a few papers predict profit or profitability as an interval variable. Scholars are achieving conflicting results on the prediction accuracy of different methods. Some research confirms the benefits of DNN [86,87] and Random Forest [88]. In some works, regression models demonstrate greater accuracy [89]. The distinctive feature of this study is that machine learning methods were tested on three models, which differ significantly in the accuracy of the forecast and the share of explanations of variation (R^2). We have evaluated not only the success of the method on a single model, but also the stability of the results for different models.

This paper confirms the advantage of individual machine learning methods over regression models. Because a lower prediction accuracy was obtained using regression, this allows us to conclude that there are hidden non-linear dependencies and complex relationships between variables that are better modeled by machine learning methods. We found that the DNN and Random Forest methods achieved the best results in predicting the profitability of firms (interval variable) compared with regression for all three models, which is consistent with the works of [86–88]. However, we have observed the instability of the results of individual machine learning methods on different models. In particular, LSTM performs poorly on Model 1, where it is worse than regression. At the same time, it shows good results on Models 2 and 3, which are better than those of the regression. Further, the DNN and Random Forest methods did not predict profitability very well in case of Model 1. They were slightly significantly better than regression.

To solve this problem of individual methods and improve the accuracy of prediction, the paper uses advanced machine learning methods, including sets of algorithms (portfolios and ensembles). These methods achieved the best results in models with low prediction accuracy and were among the best in models with high prediction accuracy, which is consistent with the works of [48–50]. Moreover, the study confirmed consistently high results of these methods for different panel data models.

To compare forecasting methods, we used traditional parametric indicators of forecast accuracy, proposed and tested non-parametric indicators, and used the analysis of variance method to assess forecast accuracy. We confirm that it is desirable to use machine learning methods with non-linear dependencies between variables to predict the profitability of firms (in the case of interval variable and panel data analysis).

6. Conclusions

Econometric analysis provided the following results of the influence of factors on the profitability of the company. The study has found that last year's profitability highly significantly affected the firm's performance in the current year. We believe that this fact reduces the ability to manage profitability, because part of the variation in current profitability is due to last year's results, which is difficult to influence. However, this pattern must be considered in profitability forecasting models.

The study finds that firms maximize profitability when two conditions are met: high sales growth and the development of modern digital sales channels and communications with customers. The resulting dependence is non-linear. Fulfillment of only one of the conditions leads to a decrease in profitability. The worst situation is the case of falling sales with high traffic to the company's website.

Predicting the companies' profitability using machine learning methods shows that adding independent variables (last year's profitability, growth, traffic, and their interaction) to the forecast model significantly increases the accuracy of the forecast and improves all prediction measures (MAE, MSE, RMSE, and R^2). Analysis of variance confirms that the absolute errors in this case are highly significantly lower.

The model with a full set of variables achieves good profitability forecast accuracy for the best machine learning methods based on non-linear dependencies. MAE drops to 5.46 for ensemble. The median of absolute errors is even lower—3.07 for Random Forest, i.e., the error in predicting the profitability of 50% of firms does not exceed 3.07. Random-effects regression is inferior in the accuracy of predictions to all machine learning methods according to the analysis of variance. This allows us to conclude that there are hidden non-linear dependencies and complex relationships between variables that are better modeled by machine learning methods.

This paper reveals that advanced machine learning methods including algorithm sets (portfolios and ensembles) achieve better results in low predictive accuracy models (compared with individual methods: DNN, Random Forest, and LSTM). They are also among the best in models with high accuracy forecast. Portfolios and ensembles are comparable in terms of forecast accuracy. They show consistently high results in predicting firm profitability for various panel data models and can be recommended for similar tasks. This study proposes a technique for building a balanced portfolio, which is adapted for panel data.

Finally, we can point the small sample of firms and considered period, covering 2016 to 2020, as a limitation of this study. The study did not have access to data for subsequent years and to additional firms with digital customer communications data. We believe that the accuracy of profitability prediction methods can be improved and the differences between prediction methods can become more pronounced as the number of observations increases. The future studies could test profitability forecasting by implementing Monte Carlo simulation in the LSTM model (after hyperparameters tuning) and to test more optimizations in LSTM, such as Fruit Fly, and more.

Author Contributions: Conceptualization, D.B.V.; methodology, L.S., E.G., I.L. and V.S.; software, E.G., I.L. and V.S.; validation, D.B.V. and L.S.; formal analysis, E.G., I.L. and V.S.; investigation, D.B.V. and L.S.; resources, D.B.V. and L.S.; data curation, V.S., L.S. and E.G.; writing—original draft preparation, D.B.V., L.S. and V.S.; writing—review and editing, V.S., E.G. and I.L.; visualization, E.G., I.L. and V.S.; supervision, D.B.V. and L.S.; project administration, L.S.; funding acquisition, L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Russian Science Foundation, grant number 22-28-01795 «Digital capital and its impact on enterprise development under sanctions and pan-demic: Econometric modeling», project registration link: <https://rscf.ru/project/22-28-01795/> (accessed on 10 March 2023).

Data Availability Statement: The data that support the findings of this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Spitsin, V.; Vukovic, D.; Mikhalechuk, A.; Spitsina, L.; Novoseltseva, D. High-tech gazelle firms at various stages of evolution: Performance and distinctive features. *J. Econ. Stud.* **2022**, *ahead-of-print*. [[CrossRef](#)]
2. Spitsin, V.V.; Mikhalechuk, A.; Vukovic, D.B.; Spitsina, L.Y. Technical Efficiency of High-Technology Industries in the Crisis: Evidence from Russia. *J. Knowl. Econ.* **2022**, 1–26. [[CrossRef](#)]
3. Ibhagui, O.W.; Olokoyo, F.O. Leverage and firm performance: New evidence on the role of firm size. *N. Am. J. Econ. Financ.* **2018**, *45*, 57–82. [[CrossRef](#)]
4. Spitsin, V.; Vukovic, D.B.; Spitsina, L.; Özer, M. The impact of high-tech companies' performance and growth on capital structure. *Compet. Rev.* **2021**, *32*, 975–994. [[CrossRef](#)]
5. Abuzayed, B. Working capital management and firms' performance in emerging markets: The case of Jordan. *Int. J. Manag. Financ.* **2012**, *8*, 155–179. [[CrossRef](#)]
6. Akbar, M.; Akbar, A.; Draz, M.U. Global financial crisis, working capital management, and firm performance: Evidence from an Islamic market index. *SAGE Open* **2021**, *11*, 1–14. [[CrossRef](#)]
7. Guerola-Navarro, V.; Oltra-Badenes, R.; Gil-Gomez, H.; Gil-Gomez, J.A. Research model for measuring the impact of customer relationship management (CRM) on performance indicators. *Econ. Res. Ekon. Istraživanja* **2021**, *34*, 2669–2691. [[CrossRef](#)]

8. Vukovic, D.B.; Romanyuk, K.; Ivashchenko, S.; Grigorieva, E. Are CDS spreads predictable during the COVID-19 pandemic? Forecasting based on SVM, GMDH, LSTM and Markov switching autoregression. *Expert Syst. Appl.* **2022**, *194*, 116553. [[CrossRef](#)] [[PubMed](#)]
9. Le, T.D.B.; Ngo, M.M.; Tran, L.K.; Duong, V.N. Applying LSTM to Predict Firm Performance Based on Annual Reports: An Empirical Study from the Vietnam Stock Market. In *Data Science for Financial Econometrics. Studies in Computational Intelligence*; Ngoc Thach, N., Kreinovich, V., Trung, N.D., Eds.; Springer: Cham, Switzerland, 2021; Volume 898. [[CrossRef](#)]
10. Miyakawa, D.; Miyauchi, Y.; Perez, C. Forecasting Firm Performance with Machine Learning: Evidence from Japanese Firm-Level Data. Discussion Papers of Research Institute of Economy, Trade and Industry (RIETI), 17068. 2017. Available online: <https://www.rieti.go.jp/jp/publications/dp/17e068.pdf> (accessed on 14 March 2023).
11. Lado-Sestayo, R.; Vivel-Búa, M. Hotel profitability: A multilayer neural network approach. *J. Hosp. Tour. Technol.* **2019**, *11*, 35–48. [[CrossRef](#)]
12. Park, K.; Jang, S. Firm growth patterns: Examining the associations with firm size and internationalization. *Int. J. Hosp. Manag.* **2010**, *29*, 368–377. [[CrossRef](#)]
13. Alawiyah, I.; Humairoh, P.N. The impact of customer relationship management on company performance in three segments. *J. Ekon. Bisnis* **2017**, *22*, 132–144.
14. Habrosh, A.A. Impact of cash flow, profitability, liquidity, and capital structure ratio on predict financial performance. *Adv. Sci. Lett.* **2017**, *23*, 7177–7179. [[CrossRef](#)]
15. Hung, C.; Vinh, T.; Thai Binh, D. The impact of firm size on the performance of Vietnamese private enterprises: A case study. *Probl. Perspect. Manag.* **2021**, *19*, 243–250. [[CrossRef](#)]
16. Jang, S.; Park, K. Inter-relationship between firm growth and profitability. *Int. J. Hosp. Manag.* **2011**, *30*, 1027–1035. [[CrossRef](#)]
17. Mueller, D. The persistence of profits above the norm. *Economica* **1977**, *44*, 369–380. [[CrossRef](#)]
18. Poonawala, S.H.; Nagar, N. Gross profit manipulation through classification shifting. *J. Bus. Res.* **2019**, *94*, 81–88. [[CrossRef](#)]
19. Bansal, M.; Kumar, A.; Kumar, V. Gross profit manipulation in emerging economies: Evidence from India. *Pac. Account. Rev.* **2022**, *34*, 174–196. [[CrossRef](#)]
20. Lee, S. The relationship between growth and profit: Evidence from firm-level panel data. *Struct. Change Econ. Dyn.* **2014**, *28*, 1–11. [[CrossRef](#)]
21. Yoo, S.; Kim, J. The dynamic relationship between growth and profitability under long-term recession: The case of Korean construction companies. *Sustainability* **2015**, *7*, 15982–15998. [[CrossRef](#)]
22. Federico, J.S.; Capelleras, J.L. The heterogeneous dynamics between growth and profits: The case of young firms. *Small Bus. Econ.* **2015**, *44*, 231–253. [[CrossRef](#)]
23. Coad, A. Testing the principle of ‘growth of the fitter’: The relationship between profits and firm growth. *Struct. Chang. Econ. Dyn.* **2007**, *18*, 370–386. [[CrossRef](#)]
24. Coad, A. Exploring the processes of firm growth: Evidence from a vector auto-regression. *Ind. Corp. Chang.* **2010**, *19*, 1677–1703. [[CrossRef](#)]
25. Coad, A.; Rao, R.; Tamagni, F. Growth processes of Italian manufacturing firms. *Struct. Chang. Econ. Dyn.* **2011**, *22*, 54–70. [[CrossRef](#)]
26. Dolega, L.; Rowe, F.; Branagan, E. Going digital? The impact of social media marketing on retail website traffic, orders and sales. *J. Retail. Consum. Serv. (Electron. J.)* **2021**, *60*, 102501. [[CrossRef](#)]
27. Shantharam, B.B.; Balaji, P.; Jagadeesan, P. Impact of Customer Commitment in Social Media Marketing on Purchase Decision—An Empirical Examination. *J. Manag.* **2019**, *6*, 320–326. [[CrossRef](#)]
28. Stoica, V. Developing Customer Relationship Management Operations during the COVID-19 Pandemic. A Digitalization Perspective. *Strateg. Shap. Future Bus. Econ. (Electron. J.)* **2022**, 273–284. Available online: <https://strategica-conference.ro/wp-content/uploads/2022/04/21-2.pdf> (accessed on 10 March 2023).
29. Zhang, H.; Yang, F.; Li, Y.; Li, H. Predicting profitability of listed construction companies based on principal component analysis and support vector machine—Evidence from China. *Autom. Constr.* **2015**, *53*, 22–28. [[CrossRef](#)]
30. Kim Oanh, T.T.; Thu Hien, D.T.; Phuong Anh, H.T.; Thu Ha, D.T. Ownership Structure and Firm Performance: Empirical Study in Vietnamese Stock Exchange. *Stud. Comput. Intell.* **2020**, *3*, 53–67. [[CrossRef](#)]
31. Lee, J.; Jang, D.; Park, S. Deep Learning-Based Corporate Performance Prediction Model Considering Technical Capability. *Sustainability* **2017**, *9*, 899. [[CrossRef](#)]
32. Anagnostis, A.; Papageorgiou, E.; Bochtis, D. Application of Artificial Neural Networks for Natural Gas Consumption Forecasting. *Sustainability* **2020**, *12*, 6409. [[CrossRef](#)]
33. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The Performance of LSTM and BiLSTM in Forecasting Time Series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019. [[CrossRef](#)]
34. Araujo, G.S.; Gaglianone, W.P. Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Lat. Am. J. Cent. Bank.* **2023**, *4*, 100087. [[CrossRef](#)]
35. Effrosynidis, D.; Spiliotis, E.; Sylaios, G.; Arampatzis, A. Time series and regression methods for univariate environmental forecasting: An empirical evaluation. *Sci. Total Environ.* **2023**, *875*, 162580. [[CrossRef](#)]

36. Kim, T.; Sharda, S.; Zhou, X.; Pendyala, R.M. A stepwise interpretable machine learning framework using linear regression (LR) and long short-term memory (LSTM): City-wide demand-side prediction of yellow taxi and for-hire vehicle (FHV) service. *Transp. Res. Part C Emerg. Technol.* **2020**, *120*, 102786. [[CrossRef](#)]
37. Jahn, M. Artificial neural network regression models in a panel setting: Predicting economic growth. *Econ. Model.* **2020**, *91*, 48–54. [[CrossRef](#)]
38. Maiti, M.; Vykylyuk, Y.; Vukovic, D. Cryptocurrencies Chaotic Co-movement Forecasting with Neural Networks. *Internet Technol. Lett.* **2020**, *3*, e157. [[CrossRef](#)]
39. Kock, A.B.; Teräsvirta, T. Forecasting performances of three automated modelling techniques during the economic crisis 2007–2009. *Int. J. Forecast.* **2014**, *30*, 16–31. [[CrossRef](#)]
40. Acharya, M.S.; Armaan, A.; Antony, A.S. A Comparison of Regression Models for Prediction of Graduate Admissions. In Proceedings of the 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 21–23 February 2019. [[CrossRef](#)]
41. Zaheer, S.; Anjum, N.; Hussain, S.; Algarni, A.D.; Iqbal, J.; Bourouis, S.; Ullah, S.S. A Multi Parameter Forecasting for Stock Time Series Data Using LSTM and Deep Learning Model. *Mathematics* **2023**, *11*, 590. [[CrossRef](#)]
42. Al-Ali, E.M.; Hajji, Y.; Said, Y.; Hleili, M.; Alanzi, A.M.; Laatar, A.H.; Atri, M. Solar Energy Production Forecasting Based on a Hybrid CNN-LSTM-Transformer Model. *Mathematics* **2023**, *11*, 676. [[CrossRef](#)]
43. Dietterich, T.G. Ensemble Methods in Machine Learning. In *Multiple Classifier Systems*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2000; Volume 1857. [[CrossRef](#)]
44. Nguyen, H.V.; Byeon, H. Prediction of Parkinson’s Disease Depression Using LIME-Based Stacking Ensemble Model. *Mathematics* **2023**, *11*, 708. [[CrossRef](#)]
45. Alsalem, K.O.; Mahmood, M.A.A.; Azim, N.; Abd El-Aziz, A.A. Groundwater Management Based on Time Series and Ensembles of Machine Learning. *Processes* **2023**, *11*, 761. [[CrossRef](#)]
46. Gomes, C.P.; Selman, B. Algorithm portfolios. *Artif. Intell.* **2001**, *126*, 43–62. [[CrossRef](#)]
47. Kotthoff, L. Algorithm Selection for Combinatorial Search Problems: A Survey. In *Data Mining and Constraint Programming*; Lecture Notes in Computer Science; Bessiere, C., De Raedt, L., Kotthoff, L., Nijssen, S., O’Sullivan, B., Pedreschi, D., Eds.; Springer: Cham, Switzerland, 2016; Volume 10101. [[CrossRef](#)]
48. Calderín, J.F.; Masegosa, A.D.; Pelta, D.A. An algorithm portfolio for the dynamic maximal covering location problem. *Memetic Comp.* **2017**, *9*, 141–151. [[CrossRef](#)]
49. Yawen, L.; Liu, Y.; Bohan, Y.; Ning, W.; Tian, W. Application of interpretable machine learning models for the intelligent decision. *Neurocomputing* **2019**, *333*, 273–283. [[CrossRef](#)]
50. Yuen, S.Y.; Zhang, X. On composing an algorithm portfolio. *Memetic Comp.* **2015**, *7*, 203–214. [[CrossRef](#)]
51. Kourentzes, N. On intermittent demand model optimisation and selection. *Int. J. Prod. Econ.* **2014**, *156*, 180–190. [[CrossRef](#)]
52. Spark Information System. 2022. Available online: <https://www.spark-interfax.ru/> (accessed on 14 March 2023).
53. Seranking. 2021. Available online: <https://seranking.com/> (accessed on 14 March 2023).
54. Lovallo, D.; Brown, A.L.; Teece, D.J.; Bardolet, D. Resource re-allocation capabilities in internal capital markets: The value of overcoming inertia. *Strateg. Manag. J.* **2020**, *41*, 1365–1380. [[CrossRef](#)]
55. Munjal, S.; Requejo, I.; Kundu, S.K. Offshore outsourcing and firm performance: Moderating effects of size, growth and slack resources. *J. Bus. Res.* **2019**, *103*, 484–494. [[CrossRef](#)]
56. Chatterjee, S. The impact of working capital on the profitability: Evidence from the Indian firms. *SSRN Electron. J.* **2012**. [[CrossRef](#)]
57. Vaicondam, Y.; Ramakrishnan, S. Capital structure and profitability across Malaysian listed firms. *Adv. Sci. Lett.* **2017**, *23*, 9275–9278. [[CrossRef](#)]
58. Holland, C.P.; Thornton, S.C.; Naudé, P. B2B analytics in the airline market: Harnessing the power of consumer big data. *Ind. Mark. Manag.* **2020**, *86*, 52–64. [[CrossRef](#)]
59. Plaza, B. Google Analytics for measuring website performance. *Tour. Manag.* **2011**, *32*, 477–481. [[CrossRef](#)]
60. Anokhin, S.; Spitsin, V.; Akerman, E.; Morgan, T. Technological leadership and firm performance in Russian industries during crisis. *J. Bus. Ventur. Insights* **2021**, *15*, e00223. [[CrossRef](#)]
61. Vithessonthi, C.; Tongurai, J. The effect of firm size on the leverage–performance relationship during the financial crisis of 2007–2009. *J. Multinat. Financ. Manag.* **2015**, *29*, 1–29. [[CrossRef](#)]
62. Liang, D.; Tsai, C.F.; Lu, H.Y.R.; Chang, L.S. Combining corporate governance indicators with stacking ensembles for financial distress prediction. *J. Bus. Res.* **2020**, *120*, 137–146. [[CrossRef](#)]
63. Spitsin, V.; Vukovic, D.; Anokhin, S.; Spitsina, L. Company performance and optimal capital structure: Evidence of transition economy (Russia). *J. Econ. Stud.* **2020**, *48*, 313–332. [[CrossRef](#)]
64. Marquardt, D.W. Comment. You should standardize the predictor variables in your regression models. *J. Am. Stat. Assoc.* **1980**, *75*, 87–91. [[CrossRef](#)]
65. Liao, S.; Chen, J.; Ni, H. Forex Trading Volatility Prediction using Neural Network Models. *arXiv* **2021**, arXiv:2112.01166.
66. Thakallapelli, A.; Ghosh, S.; Kamalasan, S. Real-time frequency based reduced order modeling of large power grid. In Proceedings of the 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, USA, 17–21 July 2016. [[CrossRef](#)]
67. Jais, I.K.M.; Ismail, A.R.; Nisa, S.Q. Adam Optimization Algorithm for Wide and Deep Neural Network. *Knowl. Eng. Data Sci.* **2019**, *2*, 41. [[CrossRef](#)]

68. Salem, H.; Kabeel, A.E.; El-Said, E.M.; Elzeki, O.M. Predictive modelling for solar power-driven hybrid desalination system using artificial neural network regression with Adam optimization. *Desalination* **2022**, *522*, 115411. [[CrossRef](#)]
69. Ding, S.; Wang, G. Research on intrusion detection technology based on deep learning. In Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; pp. 1474–1478. [[CrossRef](#)]
70. Kingma, D.P.; Lei Ba, J. ADAM: A Method for Stochastic Optimization. Published as a Conference Paper at ICLR 2015. Available online: <https://arxiv.org/pdf/1412.6980.pdf> (accessed on 10 March 2023).
71. Zhu, Z.; Zhang, P.; Liu, Z.; Wang, J. Static Voltage Stability Assessment Using a Random UnderSampling Bagging BP Method. *Processes* **2022**, *10*, 1938. [[CrossRef](#)]
72. Christoffersen, P.; Jacobs, K. The importance of the loss function in option valuation. *J. Financ. Econ.* **2004**, *72*, 291–318. [[CrossRef](#)]
73. Biau, G.; Scornet, E. A random forest guided tour. *TEST* **2016**, *25*, 197–227. [[CrossRef](#)]
74. Pekša, J. Extensible Portfolio of Forecasting Methods for ERP Systems: Integration Approach. *Inf. Technol. Manag. Sci.* **2018**, *21*, 64–68. [[CrossRef](#)]
75. Wawrzyniak, J.; Drozdowski, M.; Sanlaville, E. Selecting algorithms for large berth allocation problems. *Eur. J. Oper. Res.* **2020**, *283*, 844–862. [[CrossRef](#)]
76. Valentini, G.; Masulli, F. Ensembles of Learning Machines. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 3–20. [[CrossRef](#)]
77. Beluch, W.H.; Genewein, T.; Nurnberger, A.; Kohler, J.M. The Power of Ensembles for Active Learning in Image Classification. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9368–9377. [[CrossRef](#)]
78. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient Boosting with Categorical Features Support. CoRR abs/1810.11363. 2018. Available online: <http://dblp.uni-trier.de/db/journals/corr/corr1810.html#abs-1810-11363> (accessed on 10 March 2023).
79. Adler, A.I.; Painsky, A. Feature Importance in Gradient Boosting Trees with Cross-Validation Feature Selection. *Entropy* **2022**, *24*, 687. [[CrossRef](#)]
80. Refaeilzadeh, P.; Tang, L.; Liu, H. Cross-Validation. In *Encyclopedia of Database Systems*; Liu, L., Özsu, M.T., Eds.; Springer: Boston, MA, USA, 2009. [[CrossRef](#)]
81. Shcherbakov, M.V.; Brebels, A.; Shcherbakova, N.L.; Tyukov, A.P.; Janovsky, T.A.; Kamaev, V.A. A Survey of Forecast Error Measures. *World Appl. Sci. J.* **2013**, *24*, 171–176.
82. Guo, P.; Liu, T.; Zhang, Q.; Wang, L.; Xiao, J.; Zhang, Q.; Luo, G.; Li, Z.; He, J.; Zhang, Y.; et al. Developing a dengue forecast model using machine learning: A case study in China. *PLoS Negl. Trop. Dis.* **2017**, *11*, e0005973. [[CrossRef](#)] [[PubMed](#)]
83. Flores, B.E. A pragmatic view of accuracy measurement in forecasting. *Omega* **1986**, *14*, 93–98. [[CrossRef](#)]
84. Kim, S.; Kim, H. A new metric of absolute percentage error for intermittent demand forecasts. *Int. J. Forecast.* **2016**, *32*, 69–79. [[CrossRef](#)]
85. Kim, S.; Kang, S.; Ryu, K.R.; Song, G. Real-time occupancy prediction in a large exhibition hall using deep learning approach. *Energy Build.* **2019**, *199*, 16–22. [[CrossRef](#)]
86. Anyaeche, C.O.; Ighravwe, D.E. Predicting performance measures using linear regression and neural network: A comparison. *Afr. J. Eng. Res.* **2013**, *1*, 84–89.
87. Gaytan, J.C.T.; Ateeq, K.; Rafiuddin, A.; Alzoubi, H.M.; Ghazal, T.M.; Ahanger, T.A.; Chaudhary, S.; Vijju, G.K. AI-Based Prediction of Capital Structure: Performance Comparison of ANN SVM and LR Models. *Comput. Intell. Neurosci.* **2022**, *2022*, 8334927. [[CrossRef](#)]
88. Erdal, H.; Karahanoğlu, İ. Bagging ensemble models for bank profitability: An empirical research on Turkish development and investment banks. *Appl. Soft Comput.* **2016**, *49*, 861–867. [[CrossRef](#)]
89. Goyeneche, D. Predicting Profitability of Neighbourhood Stores in Colombia. *Rev. Integr. Bus. Econ. Res.* **2022**, *11*, 1–24.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.