mathematics

MDPI

Article

Predicting Firm's Performance Based on Panel Data: Using Hybrid Methods to Improve Forecast Accuracy

Nikita V. Martyushev [1,*], Vladislav Spitsin [2], Roman V. Klyuev [3], Lubov Spitsina [2], Vladimir Yu. Konyukhov [4], Tatiana A. Oparina [4] and Aleksandr E. Boltrushevich [1]

1 Department of Information Technologies, Tomsk Polytechnic University, 634050 Tomsk, Russia; aeb20@tpu.ru
2 Business School, National Research Tomsk Polytechnic University, Lenin Avenue, 30, 634050 Tomsk, Russia; spitsin_vv@mail.ru (V.S.); spicyna@tpu.ru (L.S.)
3 Technique and Technology of Mining and Oil and Gas Production Department, Moscow Polytechnic University, 107023 Moscow, Russia; kluev-roman@rambler.ru
4 Department of Automation and Control, Irkutsk National Research Technical University, 664074 Irkutsk, Russia; konyukhov_vyu@mail.ru (V.Y.K.); tatianaop@istu.edu (T.A.O.)
* Correspondence: martjushev@tpu.ru

Abstract: The problem of predicting profitability is exceptionally relevant for investors and company owners making decisions about investment and business development. The global literature contains a number of studies where researchers predict the profitability of firms using various methods, including modern machine learning. However, these works hardly take advantage of panel data. This paper takes advantage of additional capabilities offered by panel data and proposes hybrid forecasting methods based on panel data, which allow significantly improving the accuracy of predicting the profitability. Our calculations show that when predicting the profitability, investors and company owners should take into account the profitability of the previous years and the trend in its change. The work shows that this approach can be successfully applied to high-tech companies whose profitability is characterised by increased volatility. Prediction forecasting includes STL-decomposition of time series, regression with random effects and machine learning (LSTM and CatBoost), and clustering. The training sample includes 1811 companies and data for 2013–2018 (panel data, 10,866 observations). The test sample contains data for these companies for 2019. As a result, the authors propose an approach significantly improving the accuracy of predicting ROA and ROE based on the panel nature of the data. The panel data allowed using the profitability of the previous years in forecast models and applying the STL-decomposition of the profitability of the previous years into three variables (Trend, Seasonal, and Residual), considerably improving the quality of the constructed forecast models (STL-CatBoost, STL-LSTM, and STL-RE hybrid models).

Keywords: firm's performance; profitability prediction; ROA; ROE; panel data; machine learning; CatBoost; long short-term memory (LSTM); clustering; seasonal decomposition of time series by LOESS (STL); hybrid methods

MSC: 68T20

Academic Editors: Hany Guirguis, Hyeon Park and Raymond Lee

Received: 9 February 2025
Revised: 25 March 2025
Accepted: 8 April 2025
Published: 10 April 2025

Citation: Martyushev, N.V.; Spitsin, V.; Klyuev, R.V.; Spitsina, L.; Konyukhov, V.Y.; Oparina, T.A.; Boltrushevich, A.E. Predicting Firm's Performance Based on Panel Data: Using Hybrid Methods to Improve Forecast Accuracy. *Mathematics* **2025**, *13*, 1247. https://doi.org/10.3390/math13081247

1. Introduction

Scientists distinguish two main goals of an enterprise: sales growth and increasing the activity efficiency (or profitability of the firm) [1,2]. This paper explores the second direction. Two important measures of profitability are addressed. The first is ROA (return on assets), a fairly stable indicator studied by many scientists. The second is ROE (return on

capital), a key indicator for investors characterised by high volatility. Based on advanced mathematical methods, this paper searches for effective methodologies and techniques and takes advantage of panel data to improve the accuracy of forecasting these indicators [3–5].

The object of forecasting in this paper is the profitability of firms operating in high-tech industries and services of the Russian economy. Russia is a large country in transition, having many firms that operate in high-tech industries, representing appropriate samples for modelling. Currently, the Russian business and stock market are highly undervalued as foreign investors have left the country. However, Russian high-tech companies show good profitability and have high growth potential, making them attractive to investors. Moreover, high-tech business is a priority for investment in the global financial markets, as it demonstrates high growth rates of business and shareholder value. At the same time, the high volatility of their financial performance complicates forecasting the profitability of such companies. This problem can be dealt with by applying hybrid forecasting and taking advantage of panel data [6,7].

Predicting profitability using modern advanced forecasting remains little studied in economic science. Linear methods and, in particular, regression are often used to evaluate and predict various indicators. Despite the relative simplicity of application, linear methods can give errors and incorrectly reflect real interrelationships of economic indicators. To solve this problem, machine learning is used, including neural networks, Random Forest, and others, as well as combinations of these methods. Machine learning already uses non-linear hidden dependencies [8]. A significant number of authors predict profitability as a binary variable when using machine learning, for instance, a binary state (positive and negative) of profitability [9], a decrease or increase in profitability [10], etc. Further work on increasing the accuracy of forecasts using machine learning can be a consideration of profitability as an interval variable. Very few such works use this approach and base it on panel data [11].

The feature of this study is using complex (composite, hybrid) forecasting models based on panel data. Panel data are multivariate data used in social sciences and econometrics obtained by a series of measurements or observations over several time periods for the same companies or people. Panel data realise complex forecasting techniques where the target variable (profitability of a firm) is defined depending on other financial indicators of firms and on the time factor, analysing the time trend in the target variable.

This study predicts a firm's profitability as an interval variable for three models (sets of variables) and random effects regression (RE), selected machine learning (CatBoost Regressor (CatBoost)), and a recurrent neural network known as long short-term memory (LSTM).

Researchers note that one of the promising areas of research in forecasting is the creation of hybrid models combining several data-mining methods [12–15]. Hybrid approaches are developed taking into account the structure of the source data, sample size and a number of other indicators, obtaining qualitatively new forecasting results compared to separately using combined methods. Clustering and Seasonal Decomposition of Time Series by LOESS (STL-decomposition) were used in this study for data processing. Based on the results, variables known as factors in forecast models were obtained: STL-CatBoost, STL-LSTM, STL-RE, STL-Cluster-CatBoost, STL-Cluster-LSTM, and STL-Cluster-RE.

The peculiarity, contributions and novelty of this study are as follows:

(A) From the economic point of view:

- The article confirms the findings of the previous studies on the effect of past years' profitability on the current year's profitability. This result is obtained for companies in high-tech industries and services of an emerging economy and for ROA and ROE measures of profitability.

- For the first time, the work finds a strong influence of past years' profitability trends obtained by the STL-decomposition of the profitability of the previous years on the current year's profitability. This finding is obtained for enterprises in high-tech industries and services sectors of an emerging economy and for ROA and ROE profitability measures.
- The above results are of great practical interest to investors, corporate managers of firms, policymakers, and economists.

(B) From the mathematical point of view:

- Methods for predicting the profitability of firms based on the analysis of panel data are tested and developed.
- A hybrid approach is proposed, which consists of using the results of Seasonal Decomposition of Time Series according to LOESS (STL) decomposition of panel data as factors for forecast ROA and ROE models of profitability.
- A hybrid approach is proposed, which consists of using the results of clustering time series by revenue growth as a factor for forecasting ROA and ROE models of profitability.

In this paper, the second section is a review of the literature devoted to forecasting a firm's profitability which formulates the hypotheses of this study. The third section describes the data, variables, forecasting and their combinations. The forecasting results are presented in the fourth section, comparing the accuracy of the constructed models and analysing variance and mathematical tests. The fifth section tests the robustness of the proposed forecasting techniques and evaluates their accuracy in economic instability. The discussion of the obtained results, practical recommendations and conclusions are given in the last two sections.

2. Literature Review and Hypothesis Development

2.1. Firm's Profitability Prediction

Sales growth and profitability are the most important forecasting economic indicators [1,2] characterising the success of a business as a whole. At the same time, sales growth is difficult to model and predict, causing a low proportion of an explained variation when applying regression models [16]. Far more studies are devoted to issues of predicting profitability and to the study of factors affecting a company's profitability [7,17–19].

The first main scientific direction related to the study of profitability is modelling the impact of factors on profitability. The literature explores a wide range of factors: leverage and firm size [1–4], working capital management and global crises [5,6], customer relationship management, and innovation [7]. An identification of basic and main factors influencing profits is an extremely important scientific task. Most scientific papers build regression models of the effect of factors on profitability to identify patterns and to provide recommendations to the owners and managers of the company.

The second direction is forecasting the firm's profitability using the results of the previous direction and including the factors influencing profitability in the forecasting models. However, the objectives and forecasting techniques differ significantly from the first direction. Predicting profitability is of interest and involves special forecasting techniques and indicators of forecast accuracy. Moreover, predicting profitability using modern advanced forecasting methods remains little studied in economic science, and some empirical studies are summarised in Table 1.

Table 1. Empirical studies, approaches, and methods for predicting profitability.

Scholars	Panel Data	Type of Dependent Variable	Dependent Variable	Methods	Last Year's Profitability as One of Factors
David Goyeneche [20]	-	binary variable	profit	logistic regression; decision tree; Random Forest; discriminant analysis; and artificial neural network	-
Anyaeche C. O. and Ighravwe D. E. [21]	-	interval variable	profitability	linear regression and neural network	-
Hamit Erdal Ilhami Karahanoglub [22]	+	interval variable	ROE	bagging ensemble models (DecisionStum, RandomTree, and Reduced Error Pruning Tree)	-
Rubén Lado-Sestayo Milagros Vivel-Búa [11]	+	interval variable	profitability	multilayer neural network (deep neural networks)	+
Darko B. Vukovic, Lubov Spitsina, Ekaterina Gribanova, Vladislav Spitsin, and Ivan Lyzin [23]	+	interval variable	ROA	random effects regression, individual machine learning (deep neural networks—DNN, LSTM, and Random Forest), and advanced machine learning consisting of sets of algorithms (portfolios and ensembles)	+

It is important to analyse each of the papers presented in Table 1 to understand the existing approaches, their strengths and weaknesses, and the place of our study in the overall context. The study from ref. [20] focuses on predicting the profitability of small shops in Colombia using non-panel data. The author applies a wide range of classification techniques, including logistic regression, decision trees, Random Forest, discriminant analysis and artificial neural networks. The results of this work allow comparing the performance of different machine learning methods for profitability forecasting tasks but do not take into account the temporal dynamics and relationships that can be extracted from panel data. This analysis does not rely on the inclusion of past profitability as a predictor. The study [22] compares linear regression and neural networks for profitability prediction. The authors treat profitability as an interval variable but also do not use panel data. The paper highlights the potential of neural networks to model non-linear relationships but does not consider the temporal structure of the data. Past profitability is not included in the analysis. Bagging ensemble methods (bagging ensemble models) based on decision trees are used to forecast the ROE of Turkish investment banks in [24]. The author uses panel data, which allows the time dynamics to be taken into account. This study shows the effectiveness of ensemble methods for financial forecasting tasks but does not investigate other machine learning methods such as neural networks. It also does not consider past profitability. The authors of [11] use multi-layer deep neural networks to predict hotel profitability using panel data. This work demonstrates the benefits of deep learning for modelling complex relationships between factors affecting profitability. The inclusion of past profitability as one of the factors significantly improves forecast accuracy. The authors show that deep neural networks give very good results compared to those of regression

models. The research [23] compares regression with random effects, individual machine learning (DNN, LSTM, and Random Forest) and advanced machine learning consisting of sets of algorithms (portfolios and ensembles) for predicting ROA of retail companies. The authors use panel data and include past profitability as a predictor. The results show that machine learning, particularly ensembles, outperforms regression models in terms of prediction accuracy. This study also emphasises the importance of considering temporal dynamics and relationships in the data. However, the authors have found no significant differences in the accuracy of machine learning.

In general, the above-mentioned works confirm the advantages of advanced machine learning over traditional regression models. Works [11,23] take an important step in using panel data to improve the accuracy of profitability forecasting. They show that the forecast accuracy increases significantly if the variable 'past profitability' is included in the model. Below we will use these results in developing the hypotheses of our study. Our study further develops these directions by proposing hybrid forecasting based on panel data. In contrast to the previous work, we use STL-decomposition to extract the trend, seasonality and a residual component of past profitability, which allows explaining temporal dynamics in more detail. We also investigate the impact of a firm's clustering on forecast accuracy. Our study contributes to the development of profitability forecasting and may be useful for investors and managers making decisions under uncertainty.

The problem of predicting profitability is relevant for investors and other stakeholders, but it has not been sufficiently studied in the world literature due to a number of factors.

First, in the case of profitability forecasting, the methodology is significantly different:

- Two samples are used: a training sample and a test sample.
- As a rule, the profitability forecast is based on the data of the previous years, and only certain indicators can be referred to the current year.
- Various forecasting methods are used, including machine learning.
- The aim of forecasting is to minimise errors in the text sample, and special indicators are used to assess the accuracy of the forecast.

Therefore, the literature on profitability forecasting allows identifying significant factors intended for inclusion in forecasting models but does not solve the problem of profitability forecasting and the search for effective forecasting.

Second, very few papers in the world literature are devoted to profitability forecasting based on panel data. This problem is currently under-researched, and there are only emerging works that look for benefits gained from the use of panel data [24]. Some studies predict profitability as a binary variable (positive or negative profitability, growth or decline in profitability) rather than as an interval variable. Such work is unlikely to be of much help to investors. Other studies treat profitability as an interval variable and use various approaches and techniques to predict it. Analysis of these approaches and techniques has revealed the following patterns:

- It is advisable (necessary) to include last year's profitability as one of the independent variables in forecast models. Indeed, a number of studies allow finding a positive effect between the profitability of a firm in the current year and the profitability of a firm in the previous year [25]. Last year's profitability is included in forecast models in [11,23]. Moreover, the work [23] shows that adding last year's profitability to forecast models can significantly improve the accuracy of predicting a company's performance.
- Individual machine learning or advanced machine learning, consisting of sets of algorithms (portfolios and ensembles), can improve forecast accuracy compared to regression models, but there is little difference in the accuracy of these methods. New approaches and techniques are required to improve forecasting accuracy [23].

Guided by these patterns, the authors of this paper propose new approaches and techniques to achieve significant improvements in the accuracy of predicting a firm's profitability.

2.2. Hypotheses Development

In forecast models, researchers typically add an independent variable (last year's profitability) [26], proving that it:

- Significantly increases R^2 in regression models.
- Noticeably enhances the accuracy of predicting the profitability of the current year in forecast models.

Therefore, the profitability of the previous year is the main variable determining the profitability of the current year.

This study goes further and suggests that the profitability of previous years ($t - 1$, $t - 2$, $t - 3$, etc.) affects the profitability of the current year. Using the panel nature of the data and the Seasonal Decomposition of Time Series by LOESS (STL), this paper introduces additional variables (Trend, Seasonal, and Residual) to explain the impact of past years' profitability on the current year's profitability.

Seasonal Decomposition of Time Series by LOESS (STL) is decomposing time series into additive components: trend, seasonal and residual. This method differs from its analogues (for example, wavelet transform or singular spectral analysis) in its high resistance to outliers and the availability of ready-made libraries and modules in Python. They significantly simplify the technical implementation of the method. Series decomposition occurs by smoothing the series using locally fitted regression models (LOESS). The LOESS algorithm applies locally weighted polynomial regression at each point in the dataset. Another advantage of STL-decomposition is providing good results for data with different frequencies. This approach is widely used, for example, in forecasting fuel prices [27], climate indices (El Niño Index) [28], tourist flow [29], vegetable prices [30] and in the study of time series [31].

The widespread use of STL-decomposition is due to the ability of the algorithm to identify temporal patterns. The main limitation of STL-decomposition is the selection of parameters (the choice of periodicity in the data) requiring visual (graphical) evaluation.

However, in some studies, the use of STL-decomposition did not improve the quality of the forecast model; when forecasting fuel prices [27], several approaches were compared, including STL-decomposition. However, the best generalising ability was obtained by the Autoregressive Integrated Moving Model Average (ARIMA).

In other studies, results indicate the effectiveness of using methods for decomposing time series into additive components, involving STL. When using STL-decomposition in combination with LSTM based on the attention mechanism in forecasting vegetable prices, the forecast error was reduced by 4–5% [30]. Another example of using STL is forecasting temperature time series [31], where, using additive decomposition of the temperature time series, it was possible to achieve the best approximation of the forecast model to the actual data. It significantly improves prediction accuracy for such problems. However, we are the first to use it to predict a firm's profitability based on panel data and a wide set of explanatory variables.

We expect a significantly increased accuracy of predicting firm's profitability based on this approach and are going to test the following hypotheses:

Hypothesis 1. *Additional variables (Trend, Seasonal, and Residual) obtained for last year's profitability based on Seasonal Decomposition of Time Series by LOESS significantly increase the accuracy of ROA prediction.*

Hypothesis 2. *Additional variables (Trend, Seasonal, and Residual) obtained for last year's profitability based on Seasonal Decomposition of Time Series by LOESS significantly increase the accuracy of ROE prediction.*

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

- Most works confirm the advantage of machine learning (LSTM [36], neural networks [10,37,38], and so on) over regression. Classical approaches (random effects model for panel data) were considered, whose use in [35] allowed obtaining fairly good results. More modern approaches were also considered: artificial neural networks, including LSTM [39,40]; the CatBoost algorithm [36]; and other machine learning models [39], confirming the high quality of machine learning models compared to the regression model.
- Some works do not reveal significant differences between machine learning and regression [41] or find that regression is better than machine learning [42].

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

Hypothesis 3. *Machine learning assuming the presence of non-linear latent dependencies provides greater accuracy in predicting ROA compared with traditional regression.*

Hypothesis 4. *Machine learning assuming the presence of non-linear latent dependencies provides greater accuracy in predicting ROE compared with traditional regression.*

Clustering is a machine learning problem widely used in applications for grouping many objects into some subsets (clusters) based on the similarity of these objects. That is, objects within a cluster should be more similar to each other than to the objects from other clusters. Clustering is used in such applications as mobile networks [43], economics [44], energy [45,46], ecology [47], driver behaviour on the roads [48], medicine [49], construction [50], agriculture [51], seismology [52], etc. In most works, clustering allows better performance of the forecast model to be obtained. Hence, the clustering of pharmaceutical sales data, described in [53], made it possible to identify and evaluate the impact of seasonality on drug sales and improve the forecast model based on these data. Another example of clustering is grouping days by meteorological factors when predicting electricity generation by photovoltaic installations [54], supplementing the factor space of the forecast model and improving its quality.

However, clustering does not always provide the best quality of the model. The main limitation of using clustering is the large amount of data for participation in clustering to achieve better results [55]. Therefore, we have formulated hypotheses tested experimentally using our dataset.

We tested the following hypotheses:

Hypothesis 5. *Adding clustering for an independent variable (Growth) improves the accuracy of ROA prediction.*

Hypothesis 6. *Adding clustering for an independent variable (Growth) improves the accuracy of ROE prediction.*

To test hypotheses, 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

In this work, a hybrid approach is proposed that combines the advantages of panel regression and machine learning to improve the accuracy of a firm's profitability forecasting. Panel regression, in particular random effects models, allows considering firm-specific unobservable characteristics and controlling sample heterogeneity. However, regression methods are limited in modelling complex non-linear dependencies that may exist between the predictors and the target variable. To address this issue, machine learning methods (CatBoost, XGBoost, and LSTM) are integrated that are able to identify and exploit these non-linear dependencies. The integration is achieved by training ML-models on the same features as in panel regression models, as well as on additional variables obtained as a result of STL-decomposition. It is important to note that we do not train separate ML-models for each firm but use a pooled sample, which allows the model to generalise patterns to the entire dataset. To account for interfirm differences, ML-models include variables reflecting the individual peculiarities of each firm (e.g., industry affiliation, size, age, etc.).

The method used in this paper to forecast the profitability (ROA and ROE) of companies is shown in Figure 1. The proposed approach to forecasting the profitability (ROA and ROE) of companies includes several key stages presented in Figure 1. The process begins with data collection and preprocessing, followed by the formation of variables, including standard factors and variables based on the hypotheses under study, as well as the results of STL-decomposition. Next, companies are clustered by growth indicators, after which the data are divided into training and test samples. Various models (fixed effects regression, CatBoost, XGBoost, LSTM) are built and trained on the training sample. Forecasts obtained on the test sample are assessed using various accuracy metrics. The best model is selected based on the comparison of forecast accuracy using the Wilcoxon criterion. Finally, the stability of the results is checked in conditions of economic instability.

The step-by-step forecasting algorithm is as follows:

(1) Data collection and preprocessing:
- Collect the data on the financial performance of companies for the period of 2013–2019 from the SPARK IS.
- Perform data cleaning, missing value processing, and data formatting.

(2) Formation of variables:
- Calculate dependent variables (ROA, ROE) and independent variables (Last year's profitability, Size, Growth, FATA, Leverage, Turnover, Age, and Mean_ind).
- Perform STL-decomposition of profitability time series for previous periods (2013–2018) to obtain Trend, Seasonal, and Residual variables.
- Cluster companies by the Growth indicator using the K-means algorithm to determine the Cluster variable.

(3) Separation of the data into training and test samples:
- Separate the data into training (2013–2018) and test (2019) samples.
- Normalise (Z-normalisation) the variables separately for training and test samples.

(4) Training the models:
- Train fixed effects regression models, CatBoost, XGBoost and LSTM on the training sample.
- Set the model hyperparameters using cross-validation.

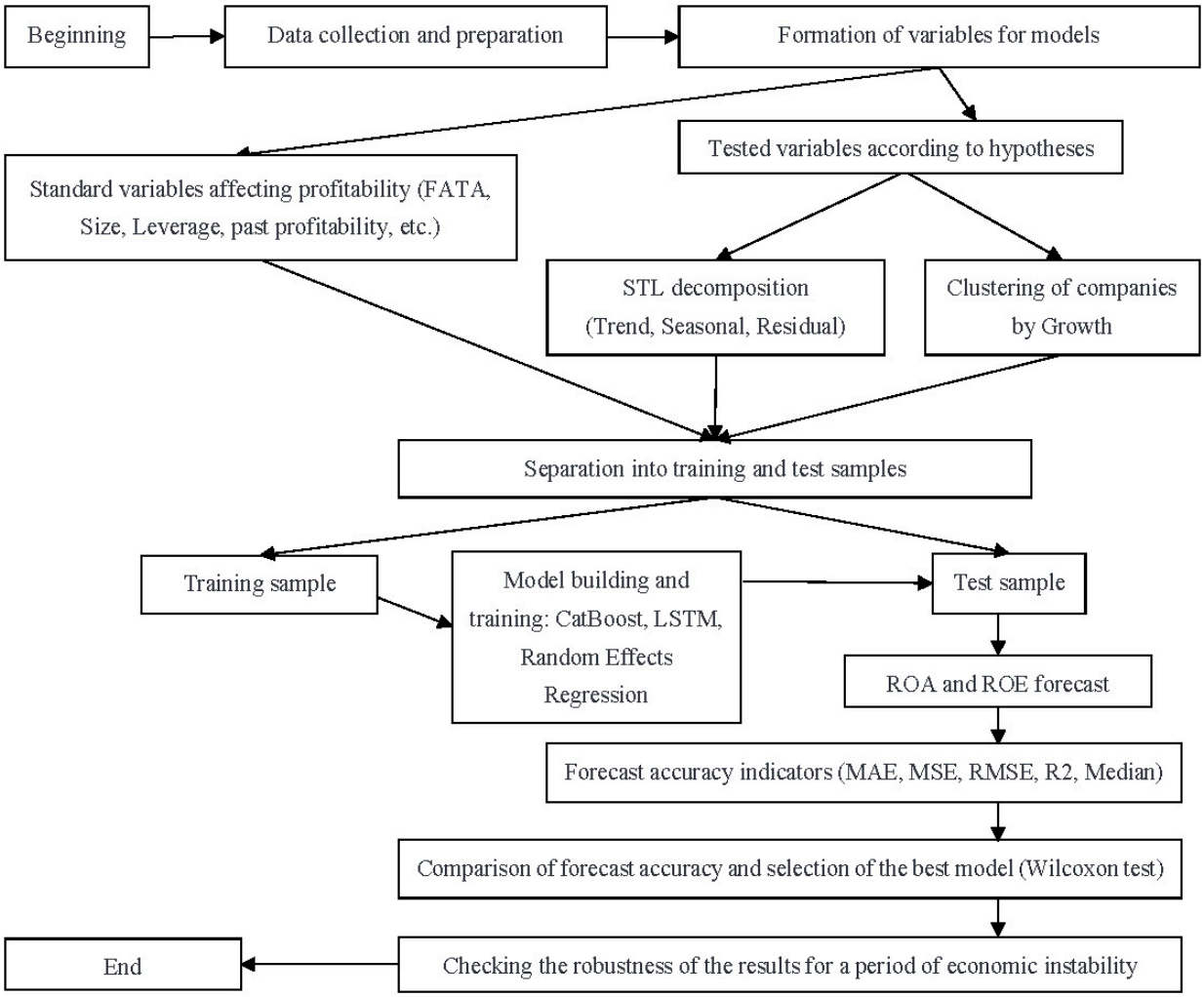

Figure 1. Simplified representation of the algorithm of forecasting the profitability (ROA and ROE) of companies.

(5) Model evaluation:
- Apply the trained models to the test sample for obtaining profitability forecasts.
- Calculate forecast accuracy metrics (MAE, MSE, RMSE, R^2, and median).
- Analyse the dispersion of absolute forecast errors to compare different models and methods.
- Check the hypotheses using the nonparametric Wilcoxon test.

(6) Choice of the best model:
- Select the model that provides the best forecast accuracy on the test sample.

(7) Testing the stability of the results:
- Test the robustness of the results for 2020 and 2021 to ensure that the hybrid method does not lose its advantages during the periods of economic instability.

The forecasting process starts with data collection and preparation followed by the formation of variables, including standard factors and variables based on the research hypotheses and STL-decomposition results. Next, the companies are clustered by growth indicators, after which the data are divided into training and test samples. Various models (regression, CatBoost, LSTM, regression with random effects) are built and trained using the training sample. The predictions obtained on the test sample are evaluated using various accuracy metrics. The best model is selected based on a comparison of the prediction

accuracy using the Wilcoxon criterion. Finally, the robustness of the results during economic instability is tested.

The choice of CatBoost and LSTM algorithms is due to their high performance for time series forecasting problems [29–31,39]. The advantages of CatBoost are the speed of the algorithm owing to training on several GPUs ensuring high accuracy by reducing overfitting and the ability to automatically select hyperparameters using the GridSearchCV cross-validation tool. The LSTM model is a type of recurrent neural network whose advantage is mechanisms for remembering and combating gradient attenuation. The LSTM architecture includes "gates" or special structures consisting of sigmoid neural networks and element-wise multiplication operations. They regulate the information flow in a memory cell, improving the LSTM performance compared to other recurrent networks.

In addition to economic indicators, the results of the decomposition of time series of each company (with a lag of 1 year) obtained using the STL algorithm were submitted as input factors to the forecast models, which improved their quality for all regression models (random effects, XGBoost, CatBoost, and LSTM).

- The influence of the factors obtained after STL-decomposition was assessed in comparison with those (lags of the target variable), concluding a greater influence of the former. Consequently, the inclusion of lags of the target variable, decomposed into additive components using STL, in forecast models increases its generalising ability of profitability (ROA and ROE).
- The accuracy of predictive methods is assessed by parametric and non-parametric (rank) prediction measures. This study uses ANOVA to identify differences between absolute forecast errors and to determine the best models and methods for predicting a firm's profitability.

3.1. Data

The sample of companies consists of 1811 Russian firms operating in high-tech industrial or service sectors. The company data were obtained from SPARK IS; the upload date was 14 March 2023. The criteria for inclusion were as follows:

- Sales of products of more than RUB 50 million annually from 2012 to 2019.
- A firm belongs to one of the following high-tech sectors (according to OKVED 2.0 or NACE Rev. 2): manufacture of basic pharmaceutical products and pharmaceutical preparations; manufacture of computer, electronic and optical products; computer programming, consultancy and related activities; information service activities; and scientific research and development.

The firms meeting these criteria were included in the sample. Companies' financial indicators were sourced from the Spark Information Systems [56].

For the case of ROA, the scope of this study was from 2012 to 2019. According to forecasting techniques, it was divided into two periods:

- The training period (2013–2018) allowed training models and identifying relationships between variables. The panel data include 10866 observations (1811 firms $\times$ 6 years). We lost one year (2012) of observations as we calculated the growth rates of sales and used the "last year's profitability" variable.
- The test period (2019) predicted the profitability of firms. It included 1811 observations (1811 firms $\times$ 1 year).

For the case of ROE, the training period and the test period are the same. However, the number of firms in the sample decreased to 1031 firms. We excluded highly leveraged firms with debt capital exceeding 80% of assets to eliminate division by zero when calculating ROE.

3.2. Variables

Dependent Variables. The company's net return on assets (ROA) and net return on capital (equity) (ROE) are dependent variables characterising the efficiency of the enterprise. This approach to measuring a firm's performance is widely used in modern economic research [57–60]. The choice of these indicators is due to several reasons. Firstly, ROA is a relatively stable indicator of profitability because net profit is divided by the value of assets, being less volatile. ROA characterises the return on all of the company's assets, simplifying modelling and forecasting. Secondly, ROE is a key metric for investors reflecting the return on capital invested in a company and allowing comparison with alternative investments. Despite its volatility, ROE assesses the attractiveness of a business to investors. The authors exclude highly leveraged firms from the analysis to avoid distorting ROE when dividing by small or negative equity values.

ROA is calculated as the ratio of net profit to the firm's assets multiplied by 100%. ROE is calculated as the ratio of net profit to the firm's capital (equity) multiplied by 100%.

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

- Last year's profitability (ROA t − 1 or ROE t − 1) [11,25].
- Firm's size (Size) operationalised using the natural logarithm of the firm's total assets. To ensure temporal consistency in value terms, adjustments are applied based on the inflation index [61,62].
- Sales growth (Growth) measured as the ratio of difference in the revenue between t and (t − 1) years to the revenue in the (t − 1) year [63–65].
- Share of fixed assets in total assets (FATA) [59,66].
- Leverage calculated as the share of borrowed funds in the assets [4,67].
- Asset turnover (Turnover) measured as the ratio of revenue to the company's assets serving as a control for the company's efficiency to generate sales [68].
- Firm's age (Age) measured as a number of years since the company's establishment according to the SPARK database [26,67].
- Mean_ind variable reflecting differences in the firm's performance across industries and years [23].
- Trend, Seasonal, and Residual variables obtained by means of STL-decomposition of the variable of last year's profitability. A technique of STL-decomposition is described below.
- Cluster variable obtained when clustering by a Growth variable. The clustering technique is described below.

Profitability was predicted based on current-year sales and values of most other variables in the previous years. That is, to predict the profitability for year t, we used the Growth variable for the t year and most of the other variables (FATA, Leverage, Turnover, Mean_ind, etc.) for the past year (t − 1).

3.3. Formation of Training and Test Samples

A set of initial data was divided into training (2013–2018) and test (2019) samples. The mean (Z-normalisation) normalised peculiarities according to (1) separately for a training sample and a test sample:

$$\hat{x}_i = \frac{x_i - \overline{X}}{\sigma_x} \quad (1)$$

The variables are standardised according to [69]. Standardisation reduces problems of multicollinearity, especially in cases of interaction between variables (e.g., the square of revenue growth). Then, the square of the sales growth rate ($Growth^2$) was calculated by squaring the normalised Growth value.

3.4. STL-Decomposition and Calculation of Trend, Seasonal, and Residual Variables

To obtain new factors for models and to study their influence, decomposition took place according to seasonal trends using locally selected regression models (LOESS) of time series of firms (hereinafter an STL-decomposition) into trend, seasonal and residual components according to (2):

$$Y_i = T_i + S_i + R_i \quad (2)$$

where T_i, S_i, and R_i are trend, seasonal, and residual components; Y is the target variable (ROA and ROE); and $i = 1, \ldots, n$, where n is the length of the time series.

STL-decomposition is decomposing the time series into additive components, according to hypothesis 3, allowing the obtainment of new results by applying them as predictors taken with a lag of 1 time interval (in this case, a year).

Stages of STL-decomposition were summarised as follows. An STL module was imported from the Python 3.12 statsmodels.tsa.seasonal library. Then, a function performing STL-decomposition in a cycle for each company was created. The decomposition parameters were period = 2 and seasonal = 3; other parameters were set by default. A Period parameter is a periodicity of a sequence, and a Seasonal parameter specifies a length of a season (always an odd integer). These parameters were selected experimentally when decomposing a series and removing a noise component. An error calculated in (3) with an initial series did not exceed 2.5%, which is acceptable.

The period and seasonal parameters of STL-decomposition play an important role in the capacity of time series decomposition. The period parameter determines the period of seasonality in the data. In our case, the data represent annual financial indicators of companies, and we aim to highlight intra-annual seasonality. This is absent from annual data, but there may be cycles associated with reporting and planning. Therefore, we decided to set period = 2, allowing the algorithm to highlight relatively short cycles in the data. The seasonal parameter determines the window length for smoothing the seasonal component. The seasonal = 3 value was chosen to ensure sufficient smoothing without causing excessive loss of information about time changes.

To justify the choice of the parameters and to assess their impact on the forecasting results, a sensitivity analysis was conducted. Within the framework of the analysis, the period parameter (from 2 to 3) and the seasonal parameter (from 3 to 5) varied. The results showed that changing the parameters in the specified ranges had a minor effect on the overall forecasting accuracy. But in some cases, it could lead to small changes in the values of the MAE and RMSE metrics. In particular, increasing the seasonal parameter to 5 led to a slight decrease in forecasting accuracy for some companies due to excessive data smoothing and the loss of important information about trends. Based on the analysis, the values of period = 2 and seasonal = 3 were decided to be kept as optimal for this task, providing a balance between smoothing and preserving information about time changes:

$$\text{MAPE}\ (\mathrm{y}, \hat{\mathrm{y}}) = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=0}^{\mathrm{n}-1} \frac{|\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i}|}{\mathrm{y_i}} \times 100\% \quad (3)$$

where y_i is the actual value for the ith observation; $\hat{y}_i$ is the forecasted (calculated) value for the ith observation; and n is the number of observations.

Such an approach allows obtaining Trend, Seasonal, and Residual variables as additive components of the initial series (ROA and ROE).

However, to further assess an impact on a target result, all these factors, including a residual component, were included in models 2 and 3 lagged for 1 time period (year). Information Gain (IG) features were filtered. The essence is to calculate the (H(X)) entropy of information according to (4) and the relative (H(Y | X)) entropy according to (5). Then,

a difference in the values (6) is calculated, known as an "Information Gain" value. IG characterises a relationship between dependent variables and the target result: the higher the value of IG, the stronger the influence of the change in the factor on the change in the target variable. That is, IG is a value by which the uncertainty with respect to Y changes with an informational addition of each factor relative to the target (Y) variable:

$$H(X) = -\sum_{x_i \in X} p(x_i) \cdot \log_2(p(x_i)) \quad (4)$$

where $p(x_i)$ is the probability that the X variable will take a value of x_i,

$$H(Y \mid X) = \sum_{x_i \in X} p(x_i) \cdot H(Y \mid X = x_i), \quad (5)$$

where $H(Y \mid X = x_i)$ is the entropy calculated for records when $X = x_i$,

$$IG(Y \mid X) = H(Y) - H(Y \mid X) \quad (6)$$

This approach allows obtaining a ranked list of factors ordered by the magnitude of their influence on the target variable (Figure 2 for ROA, Figure 3 for ROE).

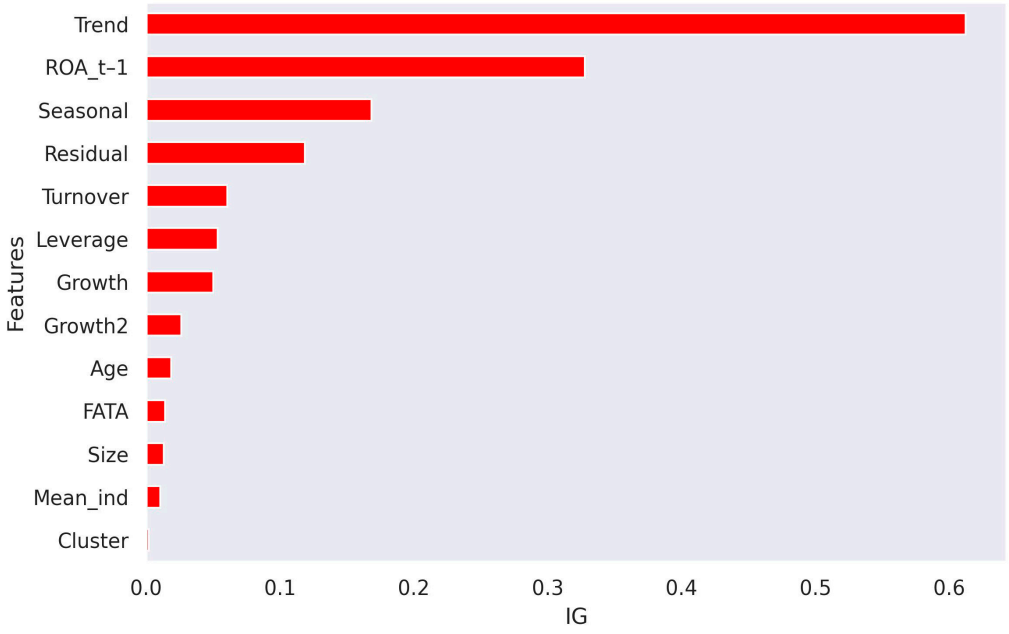

Figure 2. Influence of factors on return on assets (ROA).

3.5. Clustering Firms and Defining a Cluster Variable

Using a K-means algorithm, group firms were clustered depending on the values of target variables. Taking into account limitations for each of the samples, each sample was clustered separately. During clustering, K-means uses the Euclidean distance between points as a metric.

The optimal number of clusters was found using the Silhouette analysis algorithm involving the calculation of indicators of "cohesiveness" and "separation". "Cohesiveness" allows measuring a similarity of points in one cluster, being a kind of an intra-cluster metric. Let C be a cluster, and x_i, $x_j \in C$ be two points in this cluster. A distance (d) between x_i and x_j can be considered as a measure of their similarity. Based on it, we can determine the

connectivity of x_i in the C cluster according to (6) as an average distance between x_i and other points x_j in the C cluster:

$$a_i = \text{mean}_{xj \in C}(d(x_i, x_j)) \quad (7)$$

The "separation" indicator is an intercluster indicator characterising non-intersections of clusters. A separation of $x_i \in C_1$ is calculated as a minimum average distance between x_i and other clusters according to (7) when $C_2 \neq C_1$:

$$b_i = \min(\text{mean}_{xj \in C_2}(d(x_i, x_j))) \quad (8)$$

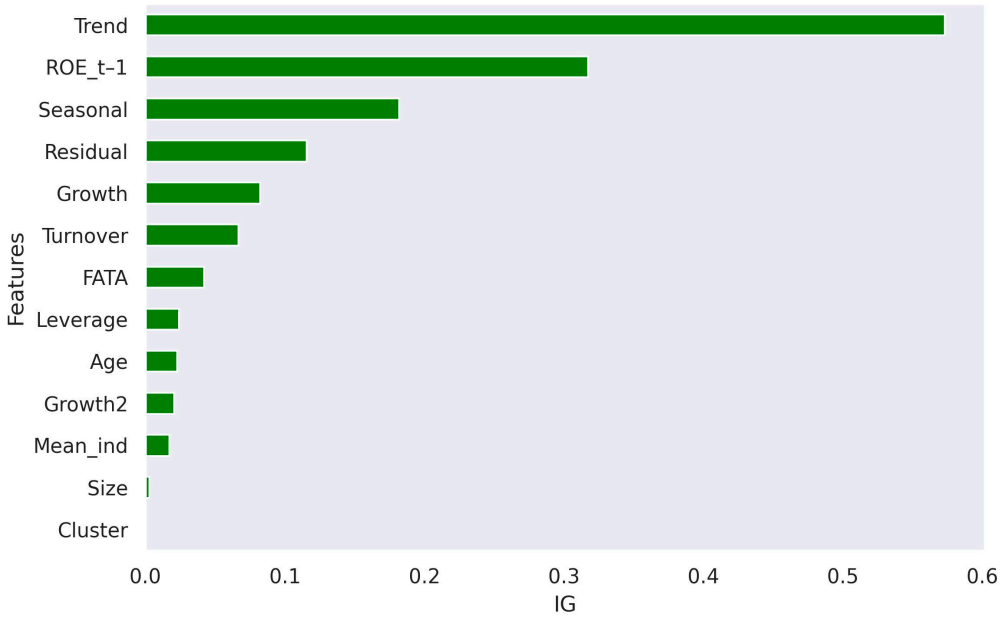

Figure 3. Influence of factors on return on equity (ROE).

A silhouette value of s_i represents a combination of connectivity and separation conditions according to (9). A range of acceptable values for the silhouette is [–1, 1]. The closer the silhouette value to 1, the better the clustering quality.

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (9)$$

In this study, for all the samples, a silhouette value was more than 0.5, indicating a sufficient quality of selection of the optimal number of clusters. After that, the optimal number of clusters was calculated for each sample; they were clustered using the K-means algorithm, including the following steps:

1. Setting the optimal number of k clusters as a parameter.
2. Randomly placing k centroids into the data space.
3. For each point in the dataset, calculating which centroid it is closer to.
4. Moving each centroid to the centre of the sample that we assigned to this centroid.
5. Repeating the last two steps until centroids "converge" (their displacement relative to the previous position does not exceed some predetermined small value) [70].

Such an approach allows obtaining a Cluster variable taking values from 0 to n, where n is an optimal number of clusters. The number of clusters was 5 for ROA and 4 for ROE.

Therefore, the Cluster variable is grouping firms by a target indicator, and its values at the input are the numbers of groups into which the firms were combined.

3.6. Models

3.6.1. Regression Analysis

This research applies regression analysis to the panel data. The regression model based on the Ordinary Least Squares (OLS) method is considered inadequate. For panel data analysis, either fixed-effects or random-effects models are commonly used. In this study, random-effects models are examined because the RE model assumes that individual effects (unobservable factors affecting each firm's profitability) are not correlated with the independent variables included in the model. In contrast to the fixed effects (FE) model, the RE model estimates these individual effects as random variables. They allow accounting for both intra-firm (changes in profitability within the same firm over time) and inter-firm (differences in profitability between different firms) variations. The choice of RE approach is driven by the desire for greater estimation efficiency because the RE model provides more accurate coefficient estimates than the FE model does. This assumes the fact that individual effects and regressors are uncorrelated. If the individual effects were correlated with the independent variables, the FE model would be more appropriate, as it eliminates the bias caused by such correlation. However, based on theoretical considerations and preliminary data analysis, we assumed that uncorrelatedness is a reasonable assumption for our sample, justifying the use of the RE model.

To confirm the adequacy of the choice of the random effects model, the Hausman test was conducted. The test results did not reveal statistically significant differences between the estimates obtained using fixed and random effects models ($p > 0.05$), which further justifies the choice of the random effects model.

In line with the common practice of analysing panel data using the random effects model (10), this paper assumes that the random individual effects (μ_i) have a normal distribution with mean 0 and constant variance. There is no supposed correlation between individual effects (μ_i) and first-level random disturbances (ε_{it}), and no autocorrelation between first-level random disturbances. The general formula for a regression model with random effects is [26]:

$$Y_{it} = \text{Intercept} + X_{it} \times \beta + \mu_i + \varepsilon_{it}, \quad (10)$$

where:

Intercept is the constant;
X_{it} is a variable;
β is a coefficient for each variable;
μ_i is random error invariant in time for each object;
ε_{it} is model regression residual.

Since the Ordinary Least Squares (OLS) method was inadequate for panel data analysis (Section 3.6.1), the Generalised Least Squares (GLS) method was used to estimate the parameters of the random effects model. GLS considers the correlation structure between observations within each firm, improving the efficiency of the estimates.

The models are presented in Table 2, where ROA and ROE are dependent variables.

The first model predicts the profitability of firms based on a standard set of variables. It is calculated using three methods (regression with random effects, CatBoost, and LSTM) and allows comparing their accuracy while working with a standard set of variables.

The second model adds Trend, Seasonal, and Residual variables derived from STL-decomposition. It tests hypotheses No. 1.1 and No. 1.2 about the influence of STL-decomposition on the accuracy of predicting the profitability (STL-CatBoost, STL-LSTM, and STL-RE). Model No. 2 compares the accuracy of forecasting methods (regression with random effects, CatBoost, and LSTM) and testing hypotheses Nos. 2.1 and 2.2.

Model No. 3 adds a Cluster variable derived from clustering (STL-Cluster-CatBoost, STL-Cluster-LSTM, and STL-Cluster-RE). It tests hypotheses No. 3.1 and No. 3.2 about the impact of clustering on the accuracy of predicting the profitability.

Table 2. Regression models and their variables.

N	Variables	Model 1	Model 2	Model 3
1	FATA	+	+	+
2	Size	+	+	+
3	Leverage	+	+	+
4	Turnover	+	+	+
5	Age	+	+	+
6	Mean_ind	+	+	+
8	ROA t − 1/ROE t − 1	+	+	+
9	Growth	+	+	+
10	$Growth^2$	+	+	+
11	Trend		+	+
12	Seasonal		+	+
13	Residual		+	+
14	Cluster			+

According to Table 2, formulas of the regression model with random effects for model 1 are as follows:

$$\begin{gathered} \text{ROA} = \text{Intercept} + \beta_1 \times \text{FATA} + \beta_2 \times \text{Size} + \\ + \beta_3 \times \text{Leverage} + \beta_4 \times \text{Turnover} + \beta_5 \times \text{Age} + \\ + \beta_6 \times \text{Mean\_ind} + \beta_7 \times \text{ROE}_{t-1} + \beta_8 \times \text{Growth} + \beta 8 \times \text{Growth}^2 + \mu_i + \varepsilon_{it}; \end{gathered} \tag{11}$$

$$\begin{gathered} \text{ROE} = = \text{Intercept} + \beta_1 \times \text{FATA} + \beta_2 \times \text{Size} + \\ + \beta_3 \times \text{Leverage} + \beta_4 \times \text{Turnover} + \beta_5 \times \text{Age} + \\ + \beta_6 \times \text{Mean\_ind} + \beta_7 \times \text{ROE}_{t-1} + \beta_8 \times \text{Growth} + \beta 8 \times \text{Growth}^2 + \mu_i + \varepsilon_{it}; \end{gathered} \tag{12}$$

To minimise the problems of multicollinearity, all independent and control variables of regression models were standardised according to [23].

The choice of the random effects (RE) instead of the fixed effects (FE) was driven by the assumption that there was no correlation between individual effects and the regressors included in the model (Section 3.6.1). While it is recognised that the estimates from the RE model may be biased when the number of time periods is small (T = 6), the use of GMM methods (e.g., Arellano–Bond estimator) was rejected for the following reasons. First, GMM required a larger number of time periods to obtain reliable estimates, which is not appropriate for our sample size. Second, preliminary experiments using GMM did not show a significant improvement in results compared to the RE model. Moreover, Hausman tests (if conducted) did not reveal a statistically significant difference between the estimates obtained using the RE and FE, which also confirms the adequacy of using the RE model.

To avoid inaccuracies, ε_{it} in formula 10 should be interpreted as random disturbances of the first level and not as residuals being their estimates.

It is important to note that when using ML-models (XGBoost, CatBoost, and LSTM) together with the random effects model, the random effects model is estimated firstly, and then the predicted random effects values were used for each firm as one of the input features for the ML-model. This allows the ML-model to directly take into account individual firms' effects and improve the prediction accuracy.

3.6.2. Gradient Boosting CatBoost Model

CatBoost Regressor implements a gradient boosting algorithm and is an ensemble method based on decision trees. The CatBoostRegressor uses categorical variables,

L2-regularisation and other opportunities. The choice of this algorithm is justified by results of analysing modern research on predicting economic indicators [71–73].

The CatBoost Regressor is an effective forecasting tool in various fields, including in economics. For example, ref. [71] shows that CatBoost Regressor allowed predicting the price of cryptocurrency with high accuracy (R^2 = 0.9858 calculated according to (13)), surpassing the results of XGBoost and LightGBM. Similar conclusions about the high efficiency of CatBoost are presented in [72,73].

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (13)$$

The main parameters of the model are iterations (a number of iterations), loss_function (a function of losses), learning_rate (rate of learning), and depth (a depth of the trees). A difference between this algorithm and the Random Forest algorithm is a consistent improvement of decision tree models when each subsequent decision tree model in the ensemble is built considering the results of the previous model.

At each iteration of a gradient descent, model parameters are updated according to (14) in the direction that is opposite to a gradient of the loss function:

$$\theta = \theta - \eta \cdot \nabla_\theta J(\theta), \quad (14)$$

where η is the step of a gradient descent known as learning rate; θ is the model parameter; and $\nabla_\theta J(\theta)$ is the gradient of the loss function (J) by the θ parameter.

In this study, the loss function is a root-mean square error (RMSE):

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (15)$$

CatBoost Regressor implements the gradient boosting algorithm and is an ensemble method based on decision trees. The choice of this algorithm is conditioned by its high performance, the ability to effectively process category features and resistance to overfitting, which is especially important when working with economic data. CatBoost effectively copes with heterogeneous data, providing the possibility to simultaneously process numerical and categorical features without the need for the preliminary encoding of the latter. However, CatBoost requires the careful setting of hyperparameters to achieve optimal accuracy and prevent overfitting.

For each of the models (Table 2), the parameters were selected using GridSearchCV. The hyperparameters in GridSearchCV were set as follows: cv = 3, estimator = CatBoostRegressor (verbose = 0), param_distributions = param_grid, where param_grid = {"iterations": [100, 500, 1000], "learning_rate": [0.01, 0.03, 0.05], "depth": [4, 6, 8, 10], "l2_leaf_reg": }. As a result, the optimal parameters of CatBoost are as follows: a depth = 10; L2 regularisation (l2_leaf_reg) = 1; a learning rate = 0.03; and a number of iterations = 1000. The results of the model quality assessment are presented in Tables 3–6.

The panel data were transformed into a format suitable for CatBoost as follows: each observation was a row with indicators for a specific company in a specific year. To account for time dependence, lags of dependent and independent variables were included in the model. In this work, a lag of one year ($t - 1$) was used. CatBoost works well with category features. The Mean_ind variable, reflecting differences in a firm's performance across industry and years, was treated as a category attribute.

Table 3. Accuracy of ROA forecast models (test dataset).

Model	Algorithm	MAE	MSE	RMSE	R^2	Median
1	CatBoost	8.80	206.36	14.36	0.47	5.01
2		5.99	90.13	9.49	0.77	3.58
3		5.94	89.39	9.46	0.77	3.47
1	LSTM	9.12	229.44	15.15	0.41	5.22
2		6.41	112.37	10.6	0.71	3.55
3		6.43	114.18	10.68	0.71	3.8
1	Random-effects model (RE)	11.85	320.85	17.91	0.18	8.04
2		6.39	101.99	10.09	0.74	4.01
3		6.38	101.68	10.08	0.74	3.99
1	XGBoost	9.06	212.35	14.86	0.43	5.13
2		6.19	103.84	9.97	0.73	3.56
3		6.22	105.58	9.92	0.73	3.59

Table 4. Accuracy of ROA forecast models (train dataset).

Model	Algorithm	MAE	MSE	RMSE	R^2	Median
1	CatBoost	7.45	151.22	12.25	0.65	4.08
2		6.11	95.22	9.75	0.75	3.64
3		6.05	90.85	9.10	0.76	3.51
1	LSTM	8.50	181.95	13.40	0.55	4.50
2		7.00	119.12	10.8	0.68	4.00
3		7.10	122.72	10.14	0.67	4.10
1	Random-effects model (RE)	12.42	353.56	18.81	0.19	8.29
2		6.63	107.35	10.36	0.76	3.98
3		6.63	107.27	107.27	0.76	4.02
1	XGBoost	8.31	198.27	14.15	0.49	4.80
2		6.15	95.25	9.48	0.74	3.51
3		6.10	93.31	9.41	0.75	3.49

To prevent overfitting, L2 regularisation (l2_leaf_reg = 1) and early stopping with error monitoring on the validation sample were used. Early_stopping_rounds was set to 50. This avoided overfitting and improved the generalisation ability of the model. CatBoost does not have built-in mechanisms for explicitly modelling individual effects, as is carried out in fixed- or random-effects models. However, including time lags and category features allows the model to indirectly consider these effects.

The calculations were performed using the CatBoost library in Python. The model was trained using a CPU. Using the graphical processor did not significantly increase the training speed due to the small sample size.

To prevent overfitting, CatBoost used L2 regularisation (l2_leaf_reg = 1) and early stopping (early stopping) with error monitoring on the validation sample. This avoided overfitting and improved the generalisability of the model. Subsequent analysis of the metrics on the training and test samples showed that CatBoost demonstrated little evidence of overfitting. The differences between the RMSE values on the training and test samples (factor set according to model 3) are 0.36 for ROA and 2.86 for ROE. L2 reg-

ularisation (l2_leaf_rag = 1) allowed reducing the overfitting effect and improving the model's generalisability.

Table 5. Accuracy of ROE forecast models (test dataset).

Model	Algorithm	MAE	MSE	RMSE	R^2	Median
1	CatBoost	14.81	627.66	25.05	0.52	8.53
2		10.66	317.06	17.81	0.76	6.19
3		10.59	324.43	18.01	0.75	5.72
1	LSTM	15.87	786.97	28.05	0.4	8.74
2		10.65	352.19	18.77	0.73	5.79
3		11.58	380.08	19.49	0.71	6.65
1	Random-effects model (RE)	24.21	1141.26	33.78	0.13	18.27
2		11.97	385.03	19.62	0.71	7.07
3		12.09	388.57	19.71	0.7	7.35
1	XGBoost	15.42	690.45	27.09	0.47	8.68
2		10.65	364.58	18.83	0.73	5.62
3		10.94	374.31	19.09	0.72	6.35

Table 6. Accuracy of ROE forecast models (train dataset).

Model	Algorithm	MAE	MSE	RMSE	R^2	Median
1	CatBoost	10.24	226.65	19.09	0.75	6.61
2		9.68	203.57	15.27	0.78	5.97
3		9.65	200.23	15.15	0.79	5.94
1	LSTM	15.37	590.24	24.29	0.49	8.93
2		12.85	405.91	19.15	0.68	7.88
3		13.40	425.88	18.64	0.66	8.09
1	Random-effects model (RE)	27.42	1816.43	42.62	0.21	19.39
2		13.41	520.61	22.82	0.78	7.92
3		13.41	519.54	22.79	0.78	7.93
1	XGBoost	14.20	489.53	22.13	0.67	8.13
2		10.30	292.09	17.09	0.79	6.27
3		10.35	297.72	17.25	0.78	6.27

CatBoost, in contrast to fixed- or random-effects models, does not have built-in mechanisms for explicitly modelling individual effects. However, our model includes the 'Mean_ind' variables (a category feature reflecting differences in a firm's efficiency across industries and years) and profitability lags ('ROA t − 1' or 'ROE t − 1'), indirectly allowing taking into account individual effects. Therefore, CatBoost is trained on a combined data sample of all firms, but at the same time, it takes into account their individual characteristics, which model sample heterogeneity. The training process occurs jointly for all companies.

3.6.3. Recurrent Neural Network LSTM Model

A model of a recurrent neural network of long short-term memory (LSTM) was created in TensorFlow Keras. Artificial neural LSTM networks are widely used in predicting economic indicators when forecasting gold prices [74], economic activity and individual

economic indicators analysing news sentiment [24]. LSTM networks are described in detail in [75].

A structure of the recurrent LSTM neural network was selected during a series of experiments. The Keras library tuner tool was used for its automated selection. The selected parameters were the number of hidden layers, the number of neurones in the layers, the type of the activation function, and the type of the optimiser. The hyperparameters in the tuner for the best of the selected architectures were objective = RMSE, max_trials = 5, seed = 42, and project_name = Regression. It consisted of an input, output and five hidden layers: two LSTM layers of 16 and 32 neurons, a linear layer of 32 neurones, an LSTM layer of 25 neurones, a linear layer of 25 neurones, and an output layer. A ReLU activation function was used for all hidden layers (16):

$$\mathrm{ReLU}(x) = \max(0,\ x), \tag{16}$$

where x is the value calculated by summing offset and multiplying input values of a neurone by weight coefficients. If input data are negative, ReLU takes the value of 0.

An optimisation Adaptive Moment Estimation (Adam) algorithm was used, according to which a parameter update rule was used to calculate network parameters (17):

$$\theta_{t+1} = \theta_t - \frac{\eta}{\hat{\vartheta}_t + \varepsilon}\hat{m}_t \tag{17}$$

where $\hat{m}_t$ is the vector of the first moment (accumulation of a gradient of the target function using an exponential moving average) calculated according to (17); $\hat{\vartheta}_t$ is the second moment vector (a moving average of squares of recent gradients) calculated according to (18); and ε is the smoothing parameter required to avoid division by 0.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{18}$$

where β is the variable controlling the moving average. Usually, β = 0.9 means that gradients are averaged according to the last 10 iterations. Moreover, g_t is the $\nabla_\theta J(\theta_t)$ gradient.

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{19}$$

The rest of the neural network parameters were the loss function as an average square error (15); the learning rate (learning_rate) was taken as 0.01. The training of the neural network was divided into batches, whose optimal size was 64. A convergence of the results was obtained at 200 epochs.

For each company, a time sequence of values of the dependent and independent variables was formed. The sequence length corresponded to the number of years in the training sample (6 years). LSTM took a three-dimensional tensor (batch_size, timesteps, features) as input, where batch_size was the number of companies in the package, timesteps were the length of the time sequence (6 years), and features were the number of features (independent variables). In this LSTM re-implementation, stateful = True was not used because the LSTM state was reset after processing each sequence. Using stateful = True could allow LSTM to retain information about the previous states and account for individual effects, but this would require more careful tuning and could lead to overfitting. Since statefulness was not used, individual effects were accounted for by including variables such as Mean_ind (a categorical attribute reflecting industry and year characteristics) and ROA $t - 1$ or ROE $t - 1$ (lagged profitability values) in the signs. LSTM, as well as CatBoost, did not have built-in mechanisms for the explicit modelling of individual effects, as it was performed in models with fixed or random effects. However, the formation of time

sequences and the inclusion of appropriate attributes allowed the model to consider the temporal structure and individual characteristics of companies.

One-time lag (lookback = 1) was used when generating input data for LSTM, i.e., the data for the previous year (t − 1) were used to forecast the profitability of the company in the current year (t). The use of more lags did not lead to a significant improvement in forecasting accuracy and increased the risk of overfitting. Subsequent analysis of the metrics on the training and test samples showed that the LSTM model demonstrated little evidence of overfitting. The difference between the RMSE values on the training and test samples was 0.54 for ROA and 0.85 for ROE.

The computations were performed using TensorFlow and Keras libraries in Python with a CPU. Using the graphic processor did not result in a significant increase in training speed due to the relatively small sample size.

Similarly to CatBoost, LSTM does not have built-in mechanisms for explicitly modelling individual effects. However, generating time series and including relevant attributes (e.g., 'Mean_ind', 'ROA t − 1' or 'ROE t − 1') allows the model to consider the time structure and individual characteristics of firms. LSTM is trained on the pooled data sample of all firms, rather than on individual data for each firm.

3.6.4. Extreme Gradient Boosting XGBoost Model

Extreme Gradient Based Boosting (XGBoost) is an ensemble algorithm; the ensemble of models is combined using boosting and trains a sequence of models using information about the errors of the previous model while also combining predictions and taking into account the weights of each model in the ensemble (weighted voting). Similarly to CatBoost, XGBoost has the option of stopping the learning cycle early (early_stopping_rounds parameter equal to 5 in this study), improving the performance.

According to the models described in Table 2, the parameters were selected using the GridSearchCV cross-validation tool for each set of factors. The best set of parameters included the number of trees (n_estimators = 314), the learning rate (learning_rate = 0.03), the regularisation parameter (l2_leaf_reg = 1), the maximum depth of the decision tree (depth = 4), and the total number of iterations (iterations = 100).

The panel data were transformed into a format suitable for XGBoost similarly to CatBoost. Lags of the variables and a category feature (Mean_ind) were included in the model to consider time dependencies and individual characteristics of firms. XGBoost does not have built-in mechanisms for explicitly modelling individual effects, as is carried out in fixed- or random-effects models. But the inclusion of time-lag variables and a category feature allows the model to indirectly take into account these effects.

The calculations were performed using the XGBoost library in Python using a CPU. Using the graphic processor did not lead to a significant increase in training speed due to the relatively small sample size.

The computational cost of training the models is presented in Tables 7 and 8. The training time is short due to the small sample size and the possibility of early stopping the training cycle (early_stopping_rounds) if each subsequent model in the ensemble does not improve.

The difference between the RMSE values on the training and test samples, set according to model 3, is 0.51 for ROA and 1.84 for ROE.

Table 7. Computational efficiency—ROA.

Algorithm	Model	Training Time (s)	Memory Consumption (MB)
CatBoost	1	29.7	0.4
	2	38.5	0.4
	3	38.9	0.4
LSTM	1	245.5	1.0
	2	260.9	0.9
	3	255.7	0.9
RE	1	0.2	0.3
	2	0.1	0.3
	3	0.3	0.3
XGBoost	1	5.5	0.3
	2	4.8	0.3
	3	3.3	0.3

Table 8. Computational efficiency—ROE.

Algorithm	Model	Training Time (s)	Memory Consumption (MB)
CatBoost	1	35.6	0.4
	2	43.1	0.4
	3	44.0	0.4
LSTM	1	127.1	1.0
	2	129.9	1.0
	3	138.4	0.9
RE	1	0.1	0.3
	2	0.4	0.3
	3	0.2	0.3
XGBoost	1	6.5	0.3
	2	5.5	0.3
	3	2.9	0.3

3.7. Assessing the Accuracy of Forecast Models

The following metrics were used to assess the quality of forecast models: RMSE, a mean absolute error (MAE), a mean quadratic error (MSE), a coefficient of determination (R^2), and a median absolute error (median) calculated by (13), (15), and (20)–(23), respectively:

$$\mathrm{MAE}(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{20}$$

$$\mathrm{MSE}(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{21}$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{22}$$

$$\mathrm{Median}(y, \hat{y}) = \mathrm{median}(|y_1 - \hat{y}_1|, \ldots, |y_n - \hat{y}_n|) \tag{23}$$

The authors of the work additionally use variance analysis to assess the significance of differences between absolute forecast errors according to different forecasting methods and models. Since error distributions do not correspond to the normal distribution law, we use nonparametric indicators and criteria. Absolute errors are visualised using a span diagram and its characteristics (median, mean, quartile range, etc.). A nonparametric Wilcoxon criterion is applied to assess the significance of differences in absolute forecast errors.

The calculations were performed in Python in the Google Colab programming environment in several stages.

3.8. Evaluation of Computational Efficiency

To evaluate the computational efficiency of the proposed hybrid methods and to compare them with traditional models, we measured the training time and estimated the memory consumption. All the measurements were performed in the Google Colab cloud service with the following configuration of computing hardware: Intel(R) Xeon(R) CPU @ 2.20 GHz and 13 GB of RAM.

The training time was the time required to train the model on the training sample. The memory consumption was estimated using the memory_profiler library in Python. The maximum memory consumption was estimated during the training time. To obtain reliable results, the measurements were repeated nine times and the mean value was calculated.

4. Results

4.1. Predicting the Results and the Accuracy of the Developed Models

The results and the accuracy of the developed models are predicted in Tables 3 and 4. The results are given using CatBoost, XGBoost, LSTM and the random-effects model. Model 1 includes a standard set of variables; model 2 has a standard set of variables and the results of STL-decomposition of the previous years' profitability; and model 3 contains a standard set of variables, the results of STL-decomposition of the previous years' profitability, and the clustering of companies. The XGBoost model, as well as other machine methods, demonstrated an improvement in the quality of forecasts as the model became more complex. The analysis of the indicators in Table 3 shows that model 2 using the results of STL-decomposition turned out to be the best for predicting ROA (MAE = 6.19, RMSE = 9.97, R^2 = 0.73). For predicting ROE, according to the data in Table 5, model 2 also showed the best results (MAE = 10.65, RMSE = 18.83, R^2 = 0.73). This indicates a significant impact of the profitability trend in past years on the current level of profitability of companies. A detailed list of the variables included in each model is presented in Table 2.

When evaluating forecasting models, it is important to consider the possibility of overfitting. Overfitting occurs when a model overfits the training data, causing it to lose its ability to generalise patterns to new, previously unseen data. In Tables 3–6, the accuracy of the models on the training and test samples is given to identify signs of overfitting and to assess the generalisation ability of the developed models.

The analysis of the results for the test sample (ROA) shows that CatBoost provides the best forecasting accuracy (MAE = 5.94, RMSE = 9.46, R^2 = 0.77). Comparison of the indicators with the training sample (Table 4) shows a slight increase in RMSE on the test sample (9.46 vs. 9.10), indicating the absence of significant overfitting and good generalisation ability of the model.

The analysis of the results for the test sample (ROE) shows that CatBoost provides better forecasting accuracy (MAE = 10.59, RMSE = 18.01, R^2 = 0.75). Comparison of the indicators with the training sample (Table 6) shows a slight increase in RMSE on the test sample (18.01 versus 15.15), which indicates the absence of significant overfitting and good generalisation ability of the model.

The results on the training sample (ROE) demonstrate good forecasting accuracy, especially for CatBoost (MAE = 9.65, RMSE = 15.15, R^2 = 0.79). The difference between RMSE values on the training (15.15) and test (18.01) samples is insignificant, indicating the absence of significant overfitting and good generalisation ability of the model.

Based on the results shown in Tables 4 and 6, the following conclusions can be drawn:

1. Hypotheses 1 and 2 are confirmed. The STL-decomposition of a time series of a target variable into additive components significantly increases the quality and accuracy of the forecast for both types of profitability (ROA and ROE). The introduction of the variables (Trend, Seasonal, and Residual) obtained using the STL-decomposition into forecast models significantly improves all indicators (MAE, MSE, R^2, and median).
2. Hypotheses 3 and 4 require further research to confirm the advantage of machine learning methods over the regression method. Calculations show that machine learning methods give better results and surpass the regression method for both types of profitability (ROA and ROE). However, the difference in the indicators (MAE, median, etc.) is small. The authors of this study performed additional calculations and applied the method of variance analysis to assess the significance of the differences between these methods. The calculation results are presented below.
3. Hypotheses 5 and 6 also require additional research to assess the impact of clustering on forecasting accuracy. Calculations provide contradictory data when, in some models, the clustering slightly improves individual indicators, and in other models, on the contrary, it worsens the forecast indicators. The work will present additional calculations and analysis of variance to assess the influence of clustering on the accuracy of forecast models. The calculation results are presented below.

Figures 2 and 3 show a diagram of factors sorted out in a descending order by the magnitude of IG.

It should be noted that the inter-industry variable (Mean_ind) takes into account differences in a firm's performance across industries and years. Regularisation used to prevent overfitting in CatBoost and LSTM helps reduce the impact of outliers that may be due to interactions (dependencies) between firms. Also, the clustering results group firms into groups with similar growth dynamics.

Figures 2 and 3 demonstrate that the variables formed on the basis of profitability of the previous years have the greatest impact on the profitability of the current year (ROA and ROE): Trend, ROA $t-1$ or ROE $t-1$, Seasonal, Residual. This result further confirms Hypotheses 1.1 and 1.2. Clustering (Cluster variable) has the least impact. Consequently, the use of clustering seems impractical.

A comparison of XGBoost with other methods (CatBoost and LSTM) shows that XGBoost provides comparable forecasting accuracy. However, it is worth noting that the XGBoost model requires careful hyperparameter setting to achieve optimal performance. Despite this, XGBoost remains a competitive tool for forecasting a firm's profitability, offering a balance between accuracy and computational efficiency.

4.2. Assessment of the Forecast Accuracy

The results of the variance analysis of the accuracy of forecast models are presented in Figures 4 and 5.

In the case of ROA, advantages of machine learning methods over regression with random effects have been revealed. An absolute prediction error is highly significantly ($p < 0.001$) lower for the CatBoost method in models 2 and 3 and strongly significantly ($p < 0.01$) lower for the LSTM method in model 2 compared with that of the regression. However, the differences are insignificant when compared with the regression of the LSTM

method in model 3. Comparing CatBoost and LSTM, we have found that CatBoost is highly significantly ($p < 0.001$) better in both models 2 and 3.

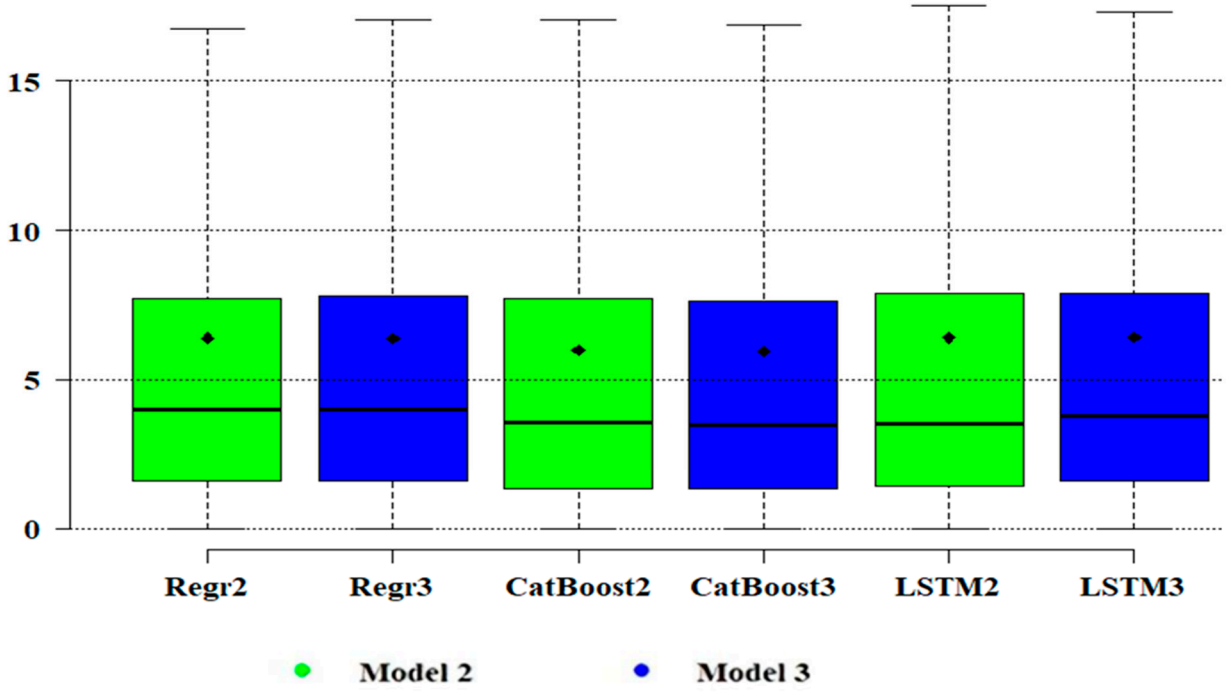

Figure 4. Boxplots of absolute errors for various models and methods for predicting ROA on the test sample. Hereinafter, a point denotes the mean value; a line means the median; a rectangle is a 25–75% quartile range; and whiskers are the minimum and maximum values or a 1.5 interquartile range.

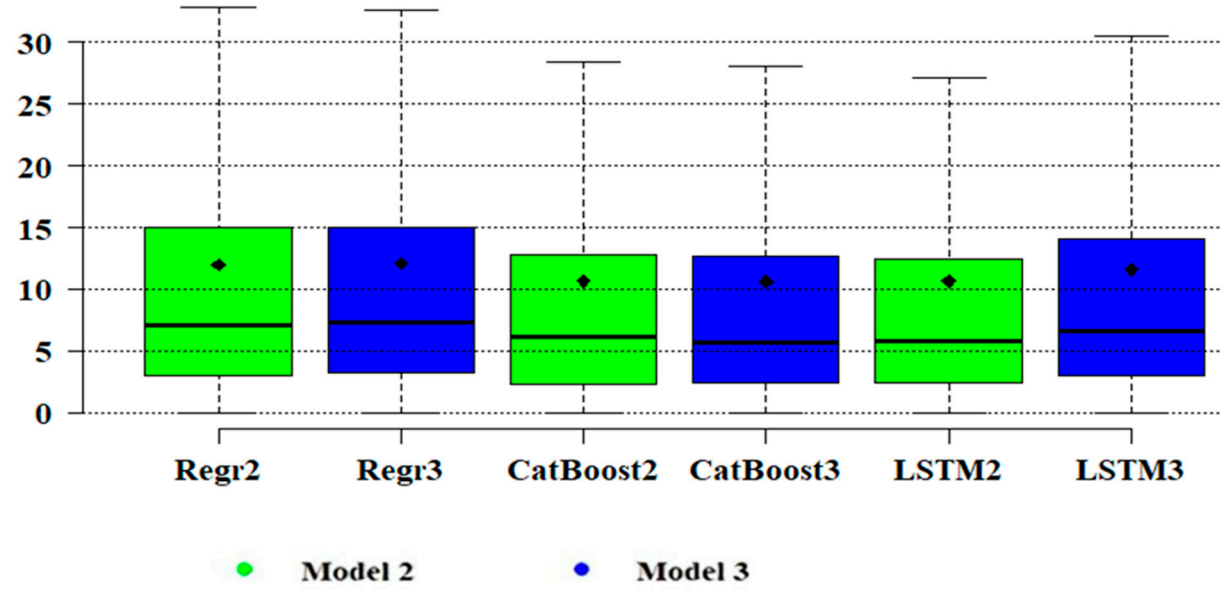

Figure 5. Boxplots of absolute errors for various models and methods for predicting ROA on the test sample. Hereinafter, a point is the mean value; a line is the median; a rectangle is a 25–75% quartile range; and whiskers are the minimum and maximum values or a 1.5 interquartile range.

In the case of ROA, including XGBoost in the analysis showed that the comparison of absolute forecast errors revealed a significant ($p < 0.01$) advantage of XGBoost models 2 and 3 over random effects regression. However, the differences in accuracy between XGBoost and CatBoost, as well as between XGBoost and LSTM, were statistically insignificant. This

suggests that all three machine learning methods (XGBoost, CatBoost, and LSTM) provide approximately the same accuracy in forecasting ROA.

In the case of ROE, advantages of machine learning methods over regression with random effects have also been revealed. An absolute prediction error is highly significantly ($p < 0.001$) lower for the CatBoost method in models 2 and 3 and the LSTM method in model 2 compared with the regression. However, the differences are insignificant when comparing the regression and the LSTM method in model 3. Comparing CatBoost and LSTM, we have found that CatBoost is better in model 3, and in model 2 the differences are not significant.

Hence, Hypotheses 3 and 4 are fully confirmed when comparing CatBoost and the regression and partially confirmed for model 2 when comparing LSTM and the regression. Machine learning methods show higher prediction accuracy compared to that of the regression. The CatBoost method shows the best results when compared with LSTM.

To evaluate the effect of clustering, we compared models 2 and 3 for different methods. The results were contradictory. In the case of ROA, CatBoost showed advantages (lower absolute error) of model 3; LSTM and the regression show negligible differences. In the case of ROE, CatBoost demonstrates insignificant differences; LSTM and the regression prefer model 2 (differences are highly significant). Consequently, clustering does not improve the quality of the constructed forecast models. Hypotheses 5 and 6 are not confirmed.

Similar results were obtained for ROE: XGBoost models 2 and 3 significantly outperform random effects regression but do not show significant differences in accuracy compared to CatBoost and LSTM. This once again confirms that machine learning provides higher accuracy in forecasting profitability than traditional regression models do. At the same time, the use of clustering does not improve the quality of forecasts regardless of the chosen machine learning method.

4.3. Computational Efficiency Results

The results of the measured training time, as well as memory consumption for all models, are presented in Tables 7 and 8.

The results of the computational efficiency measurements showed that the CatBoost and LSTM models required significantly longer training times and consumed more memory than the RE and XGBoost models did. In particular, the LSTM training time (254.03 for ROA and 131.8 for ROE) is more than 500 times greater than the RE model training time (0.3 s) and about 7 (for ROA) and 3 (for ROE) times greater than the CatBoost training time (35.7 ROA, 40.9 ROE). The training time of XGBoost (4.96 s ROE and 4.53 ROA) is much shorter than that of CatBoost (35.7 ROA, 40.9 ROE) due to the different number of iterations required to train the models. However, the computational and time costs of the ensemble models (CatBoost and XGBoost) are still higher than those of RE.

This is because LSTM and, to a lesser extent, CatBoost and XGBoost are more complex models that require more computational resources for training and prediction. LSTM is a recurrent neural network that processes data sequentially and has a complex structure, including layers with tens, hundreds or thousands of neurones. Training such a model that requires tuning a lot of parameters takes a lot of time and memory, especially when working with large amounts of data. CatBoost, although being a relatively efficient gradient-boosting algorithm, also requires significant computational cost to build an ensemble of decision trees, especially if the number of trees is large (in our study, the number of trees is parameter iterations = 1000, with the maximum tree depth = 10). In contrast, the RE (random effects regression) model is a simpler linear model, which explains its high training speed and low memory consumption. XGBoost is an efficient gradient boosting algorithm that supports stopping the learning cycle early and not adding models to the ensemble that

do not improve model performance. The learning rate depends, among other things, on the number of trees, the n_estimators parameter equal to 314 in this study. Due to the smaller number of iterations (selected using GridSearchCV) and the possibility of the early stopping of training, it shows lower time consumption than others do (LSTM and CatBoost).

The improvement in prediction accuracy achieved by hybrid methods may justify the additional computational cost in cases when high prediction accuracy is required. In cases when computational resources are limited, simpler models such as CatBoost or XGBoost can be considered. Optimisation of the proposed hybrid methods can also be considered to reduce the computational cost. One can reduce the number of layers in LSTM, simplifying the tree structure in CatBoost, using GPUs for training and inference, and quantising the models. The choice between accuracy and computational efficiency depends on the specific task and available resources.

5. Robustness Check

To confirm the effectiveness of the proposed method of forecasting the profitability of firms, the paper tests the robustness of the obtained results. The paper tests the robustness of the results by forecasting firms' profitability for 2020 and 2021. This period was chosen because it corresponds to the COVID-19 pandemic period, which was characterised by increased volatility in the financial performance of Russian firms. The purpose of this validation was to show that the proposed hybrid profitability forecasting method (model 2) retains its advantages and provides better results even under conditions of economic instability. The authors compared the forecasting results using model 2 with other models for 2020 and 2021 data to verify the stability and reliability of the proposed approach. The results of this verification are presented in Tables 9 and 10.

Table 9. Accuracy of ROA forecast models for 2020 and 2021 (robustness check).

Model	Algorithm	MAE	MSE	RMSE	R^2	Median
1	CatBoost	8.80	206.36	14.36	0.47	5.01
2		5.99	90.13	9.49	0.77	3.58

Table 10. Accuracy of ROE forecast models for 2020 and 2021 (robustness check).

Model	Algorithm	MAE	MSE	RMSE	R^2	Median
1	CatBoost	14.81	627.66	25.05	0.52	8.53
2		10.66	317.06	17.81	0.76	6.19

The predictions of results and accuracy of the developed models are shown in Tables 9 and 10.

6. Discussion and Practical Implementation

The results of our study are consistent with the previous scientific studies, and they significantly complement them. We are developing scientific approaches to predicting the profitability of a company and confirming that the accuracy of the forecast can be significantly improved by including the profitability of the previous years in the models and due to the panel nature of the analysed data.

The work [25] shows that the profitability of the previous years has a strong positive influence on the profitability of the current year. Using an example of a small sample of retail-orientated companies, scientists have proved that the inclusion of the profitability of the previous years into forecast models significantly increases the accuracy of predicting the

profitability of the current year for the case of ROA [23]. This work develops this direction and proves that the profitability of the previous years is the second most important factor influencing the forecast of the profitability of the current year. This result has been obtained for the first time using a sample of firms in the tech-intensive sector of industry and services for two types of profitability (ROA and ROE).

This study goes further and uses the panel nature of the data to significantly improve the accuracy of the prediction. Previously, panel data allowed scientists to propose portfolios of methods for predicting profitability, which gave good results, but a reduction in the absolute forecast error was insignificant [23]. This work offers STL-decomposition of the profitability of the previous years into three variable components (Trend, Seasonal, and Residual), which are the first, third and fourth most important factors influencing the forecasts of both types of profitability (ROA and ROE). An inclusion of these three variables into forecast models significantly increases the accuracy of prediction. This result is consistent with [30,31], where scientists confirm a positive effect of STL-decomposition on the accuracy of forecast models. Moreover, in order to successfully apply STL-decomposition, panel data from a training sample over a short time period (in our case, this is a six-year time period) is sufficient.

We have also identified advantages of machine learning methods over regression with random effects. Machine learning methods (CatBoost and LSTM) provide more stable and accurate results, which is consistent with [23,36,39,40]. However, their advantage over regression is small; MAE and median absolute errors are only slightly lower when using CatBoost and LSTM. On the contrary, this study has not revealed benefits of data clustering. An application of data clustering has not led to an increase in the accuracy of forecast models. The clustering algorithm was unable to identify the clusters that reflected meaningful differences in patterns important for prediction. Therefore, the preliminary clustering of the data did not have a significant impact on prediction accuracy. The main reason for this insignificant effect of the clustering on prediction accuracy is the lack of clear separable data into clusters. The reason for this is the vagueness and wide variation in the data. The main indicator, i.e., firm's sales, is not a constant variable and often varies widely. This is especially true for a country like Russia. A firm's sales can vary greatly from year to year. Therefore, it is not possible to identify clearly distinguishable clusters.

6.1. Limitations of This Study

In this paper, as in most econometric studies, there is potential for endogeneity and bias due to omitted variables. Although the models take into account a number of factors that affect a firm's profitability (size, asset structure, efficiency, age, industry affiliation, past profitability values, and their dynamics), there may still be unobserved or unincluded variables that are systematically related to both profitability and included predictors. For example, management quality and innovativeness, which are not directly measurable, may influence profitability and correlate with a firm's size or age. This can lead to biased coefficient estimates and misinterpretation of results. The use of panel data and fixed effects models, as well as instrumental variables (if available and valid), could partially mitigate this problem, but it is virtually impossible to eliminate bias completely.

In addition to the above, it should be noted that the STL-decomposition method used in this paper also has limitations that could potentially influence the results. First, as mentioned in Section 1, the choice of STL parameters, in particular period and seasonality ones, requires expert judgement and can be subjective, which may affect the quality of the decomposition and hence the accuracy of the forecasts. Second, STL assumes an additive time series model, which may not always adequately describe the behaviour of financial indicators. Third, there is a risk of overfitting when tuning the STL parameters, which

would require the use of validation techniques to assess the generalisability of the model. Finally, the lack of explicit treatment of irregular intervals and data omissions typical of financial time series may also introduce errors in the decomposition results and, as a consequence, in the final profitability forecasts.

It is important to note that in this paper, as in most econometric studies, there is potential for endogeneity and bias because of omitted variables. Despite the use of the variables reflecting interfirm differences, the ML-models (CatBoost, XGBoost, and LSTM) may not fully consider sample heterogeneity since they are trained on pooled data from all firms. Although these models take into account a number of factors influencing a firm's profitability (size, asset structure, efficiency, age, industry, historical profitability, and their dynamics), there may still be unobserved or omitted variables that are systematically related to both profitability and included predictors. The use of panel data and fixed effects models, as well as instrumental variables, mitigates this problem, but it is virtually impossible to eliminate bias completely. In future studies, it is planned to consider the possibility of training separate ML-models for each group of firms identified based on cluster analysis or other criteria, which will allow for a more accurate consideration of their specific peculiarities.

6.2. Practical Implementation of the Obtained Results

First, the obtained results can be useful for investors who predict the profitability of a company to make decisions about investing the funds. Indeed, publicly traded companies publish their financial statements, which have a strong influence on the movement of their stock prices. Investors pay particular attention to two aspects of firms' financial statements: sales growth and business profitability, in particular ROE, which characterises the return on funds invested in the company. Predicting a firm's profitability (which is investigated in this paper) will allow investors to make the right decisions to buy or sell its shares before the release of its financial statements. Our calculations show that when predicting the current profitability of a firm, investors should take into account not only the profitability of the previous years but also the trend in its change. The proposed approach will allow investors to significantly increase the accuracy of predicting two types of profitability (ROA and ROE) based on a wide range of variable factors of the previous years and sales of the current year. The work shows that this approach can be successfully applied to high-tech companies whose profitability is characterised by increased volatility.

Second, the obtained results and identified patterns can be used in corporate financial planning. It is advisable for firms to plan profit and profitability for the next year on the basis of the profitability of the previous years and trends in its change.

Third, it is advisable to take into account the obtained results by academic economists who model the influence of various factors on the profitability of firms. Our research proves that the profitability of a company strongly depends both on the profitability of the previous years and on STL-decomposition of the profitability of the previous years into three variables (Trend, Seasonal, and Residual). We believe that all these variables should be included in econometric models as control variables since they strongly influence profitability. Only after that can independent (testable) variables be included in the models. This approach will allow scientists to build more qualitative econometric models and to identify reliable patterns of the influence of factors on profitability.

7. Conclusions

In this work, the authors have proposed an approach that can significantly improve the accuracy of predicting ROA and ROE based on the panel nature of the data. The panel data have allowed using the profitability of the previous years in forecast models and

applying STL-decomposition of the profitability of the previous years into three variables (Trend, Seasonal, and Residual) improving the quality of the constructed forecast models.

The authors have compared various forecasting methods and proved the advantages of machine learning over regression; however, machine learning provides a slight reduction in absolute errors in forecast models. At the same time, CatBoost shows better results and a higher prediction accuracy. Clustering does not increase the accuracy of the profitability forecast.

In general, the constructed models allow achieving a good prediction accuracy when a median of an absolute error is reduced to 3.47 for ROA and to 5.72 for ROE. The error of 50% of firms does not exceed 3.47 and 5.72, respectively. This is a good predicted result for firms working in high-tech industries, whose profitability is characterised by high volatility. At the same time, the profitability of the current year has been forecasted on the basis of financial indicators of the previous years and sales of the current year.

However, there were some limitations of this study. The profitability of the previous years and its trend strongly influence the profitability of the current year and allow improving the accuracy of the forecast. Such a conclusion was obtained for Russian firms operating in high-tech industries, indicating a certain managerial policy followed by high-tech companies. In the case of profitability, the Russian taxation peculiarities must be considered. A number of studies confirm these conclusions for other countries [57]. However, such studies are few, and it would be interesting to see how this conclusion is realised in developed countries and industries. Perhaps such a conclusion is weakly applicable to offshore countries with low rates of taxation of profits.

Suggestions for further research. Calculations have shown a strong influence of the trend component of STL-decomposition of the profitability decomposition of the previous years on the forecast of the profitability of the current year. However, the impact of the trend component on profitability may strongly depend on the dynamics of the current year's sales:

- A trend in profitability growth may be disrupted when this year's sales drop.
- If sales increase in the current year, profitability will grow.

The trend component can be replaced by Seasonal or Residual components. In our further studies, we are going to investigate a combined effect of variables of STL-decomposition and sales dynamics on predicting profitability to additionally increase the forecast accuracy.

The obtained results characterise the management policy followed by high-tech companies in Russia. We plan to test the proposed hybrid models using samples of enterprises from other sectors of the Russian economy (manufacturing, mining, etc.) to assess their applicability in other industries and draw conclusions about similarities and differences in the management policies.

Author Contributions: Conceptualization, R.V.K.; Methodology, L.S.; Software, A.E.B.; Validation, A.E.B.; Formal analysis, V.Y.K. and T.A.O.; Investigation, R.V.K. and L.S.; Resources, V.Y.K. and T.A.O.; Data curation, V.Y.K. and T.A.O.; Writing—original draft, N.V.M., V.S. and R.V.K.; Writing—review & editing, L.S.; Visualization, A.E.B.; Supervision, N.V.M. and V.S.; Project administration, N.V.M. and V.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

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

References

1. Spitsin, V.; Vukovic, D.; Mikhalchuk, A.; Spitsina, L.; Novoseltseva, D. High-tech gazelle firms at various stages of evolution: Performance and distinctive features. *J. Econ. Stud.* **2023**, *50*, 674–695. [CrossRef]
2. Spitsin, V.V.; Mikhalchuk, A.; Vukovic, D.B.; Spitsina, L.Y. Technical Efficiency of High-Technology Industries in the Crisis: Evidence from Russia. *J. Knowl. Econ.* **2022**, *14*, 200–225. [CrossRef]
3. 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]
4. 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]
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]
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). 2017; p. 17068. 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. 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]
13. Mengash, H.A.; Alruwais, N.; Kouki, F.; Singla, C.; Abd Elhameed, E.S.; Mahmud, A. Archimedes Optimization Algorithm-Based Feature Selection with Hybrid Deep-Learning-Based Churn Prediction in Telecom Industries. *Biomimetics* **2024**, *9*, 1. [CrossRef] [PubMed]
14. Borrero, J.D.; Borrero-Domínguez, J.-D. Enhancing Short-Term Berry Yield Prediction for Small Growers Using a Novel Hybrid Machine Learning Model. *Horticulture* **2023**, *9*, 549. [CrossRef]
15. Vyalkova, S.; Morgoeva, A.; Gavrina, O. Development of a hybrid model for predicting the consumption of electrical energy for a mining and metallurgical enterprise. *Sustain. Dev. Mt. Territ.* **2022**, *14*, 486–493. [CrossRef]
16. Park, K.; Jang, S. Firm growth patterns: Examining the associations with firm size and internationalization. *Int. J. Hosp. Manag.* **2010**, *29*, 368–377. [CrossRef]
17. Alawiyah, I.; Humairoh, P.N. The impact of customer relationship management on company performance in three segments. *J. Ekon. Bisnis* **2017**, *22*, 132–144.
18. 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]
19. 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]
20. Goyeneche, D. Predicting Profitability of Neighbourhood Stores in Colombia. *Rev. Integr. Bus. Econ. Res.* **2022**, *11*, 1–24. [CrossRef]
21. 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.
22. 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]
23. Vukovic, D.B.; Spitsina, L.; Gribanova, E.; Spitsin, V.; Lyzin, I. Predicting the Performance of Retail Market Firms: Regression and Machine Learning Methods. *Mathematics* **2023**, *11*, 1916. [CrossRef]
24. Lukauskas, M.; Pilinkienė, V.; Bruneckienė, J.; Stundžienė, A.; Grybauskas, A.; Ruzgas, T. Evaluation of News Sentiment in Economic Activity Forecasting. *Eng. Proc.* **2023**, *31*, 7. [CrossRef]
25. Jang, S.; Park, K. Inter-relationship between firm growth and profitability. *Int. J. Hosp. Manag.* **2011**, *30*, 1027–1035. [CrossRef]
26. 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]

27. Sakib, A.N.; Razzaghi, T.; Bhuiyan, M.M.H. Forecasting the Fuel Consumption and Price for a Future Pandemic Outbreak: A Case Study in the USA under COVID-19. *Sustainability* **2023**, *15*, 12692. [CrossRef]
28. Chen, N.; Su, C.; Wu, S.; Wang, Y. El Niño Index Prediction Based on Deep Learning with STL Decomposition. *J. Mar. Sci. Eng.* **2023**, *11*, 1529. [CrossRef]
29. Adil, M.; Wu, J.-Z.; Chakrabortty, R.K.; Alahmadi, A.; Ansari, M.F.; Ryan, M.J. Attention-Based STL-BiLSTM Network to Forecast Tourist Arrival. *Processes* **2021**, *9*, 1759. [CrossRef]
30. Yin, H.; Jin, D.; Gu, Y.H.; Park, C.J.; Han, S.K.; Yoo, S.J. STL-ATTLSTM: Vegetable Price Forecasting Using STL and Attention Mechanism-Based LSTM. *Agriculture* **2020**, *10*, 612. [CrossRef]
31. Zhang, K.; Huo, X.; Shao, K. Temperature Time Series Prediction Model Based on Time Series Decomposition and Bi-LSTM Network. *Mathematics* **2023**, *11*, 2060. [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. Ponkumar, G.; Jayaprakash, S.; Kanagarathinam, K. Advanced Machine Learning Techniques for Accurate Very-Short-Term Wind Power Forecasting in Wind Energy Systems Using Historical Data Analysis. *Energies* **2023**, *16*, 5459. [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.; Vyklyuk, Y.; Vukovic, D. Cryptocurrencies Chaotic Co-movement Forecasting with Neural Networks. *Internet Technol. Lett.* **2020**, *3*, e157. [CrossRef]
39. Mahjoub, S.; Chrifi-Alaoui, L.; Marhic, B.; Delahoche, L. Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks. *Sensors* **2022**, *22*, 4062. [CrossRef]
40. Wang, Y.; Zhang, N.; Chen, X. A Short-Term Residential Load Forecasting Model Based on LSTM Recurrent Neural Network Considering Weather Features. *Energies* **2021**, *14*, 2737. [CrossRef]
41. 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]
42. 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]
43. Aziz, Z.; Bestak, R. Insight into Anomaly Detection and Prediction and Mobile Network Security Enhancement Leveraging K-Means Clustering on Call Detail Records. *Sensors* **2024**, *24*, 1716. [CrossRef]
44. Wei, T.; Wang, B. A Novel Curve Clustering Method for Functional Data: Applications to COVID-19 and Financial Data. *Analytics* **2023**, *2*, 781–808. [CrossRef]
45. Hu, S.; Yang, J.; Wang, Y.; Chen, C.; Nan, J.; Zhao, Y.; Bi, Y. Feature Extraction Approach for Distributed Wind Power Generation Based on Power System Flexibility Planning Analysis. *Electronics* **2024**, *13*, 966. [CrossRef]
46. Shi, J.; Wang, Z. A Hybrid Forecast Model for Household Electric Power by Fusing Landmark-Based Spectral Clustering and Deep Learning. *Sustainability* **2022**, *14*, 9255. [CrossRef]
47. Bărbulescu, A. Statistical Analysis and Modeling of the CO_2 Series Emitted by Thirty European Countries. *Climate* **2024**, *12*, 34. [CrossRef]
48. Pradhan, A.K.; Lin, B.T.W.; Wege, C.; Babel, F. Effects of Behavior-Based Driver Feedback Systems on the Speeding Violations of Commercial Long-Haul Truck Drivers. *Safety* **2024**, *10*, 24. [CrossRef]
49. Capoferri, D.; Mignani, L.; Manfredi, M.; Presta, M. Proteomic Analysis Highlights the Impact of the Sphingolipid Metabolizing Enzyme β-Galactosylceramidase on Mitochondrial Plasticity in Human Melanoma. *Int. J. Mol. Sci.* **2024**, *25*, 3062. [CrossRef]
50. Choi, Y.; Park, H.W.; Mi, Y.; Song, S. Crack Detection and Analysis of Concrete Structures Based on Neural Network and Clustering. *Sensors* **2024**, *24*, 1725. [CrossRef] [PubMed]
51. Mihaylova, D.; Popova, A.; Dincheva, I.; Pandova, S. HS-SPME-GC–MS Profiling of Volatile Organic Compounds and Polar and Lipid Metabolites of the "Stendesto" Plum–Apricot Kernel with Reference to Its Parents. *Horticulturae* **2024**, *10*, 257. [CrossRef]
52. Zhao, X.; Chen, H.; Li, B.; Yang, Z.; Li, H. Using Fuzzy C-Means Clustering to Determine First Arrival of Microseismic Recordings. *Sensors* **2024**, *24*, 1682. [CrossRef]
53. Fourkiotis, K.P.; Tsadiras, A. Applying Machine Learning and Statistical Forecasting Methods for Enhancing Pharmaceutical Sales Predictions. *Forecasting* **2024**, *6*, 170–186. [CrossRef]

54. Guo, W.; Xu, L.; Wang, T.; Zhao, D.; Tang, X. Photovoltaic Power Prediction Based on Hybrid Deep Learning Networks and Meteorological Data. *Sensors* **2024**, *24*, 1593. [CrossRef]
55. Liu, W.; Lei, P.; Xu, D.; Zhu, X. Anomaly Recognition, Diagnosis and Prediction of Massive Data Flow Based on Time-GAN and DBSCAN for Power Dispatching Automation System. *Processes* **2023**, *11*, 2782. [CrossRef]
56. Spark Information System. 2022. Available online: https://www.spark-interfax.ru/ (accessed on 14 March 2023).
57. 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]
58. 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]
59. Chatterjee, S. The impact of working capital on the profitability: Evidence from the Indian firms. *SSRN Electron. J.* **2012**. [CrossRef]
60. Vaicondam, Y.; Ramakrishnan, S. Capital structure and profitability across Malaysian listed firms. *Adv. Sci. Lett.* **2017**, *23*, 9275–9278. [CrossRef]
61. Bon, S.F.; Hartoko, S. The effect of dividend policy, investment decision, leverage, profitability, and firm size on firm value. *Eur. J. Bus. Manag. Res.* **2022**, *7*, 7–13. [CrossRef]
62. Dang, C.; Li, Z.F.; Yang, C. Measuring firm size in empirical corporate finance. *J. Bank. Financ.* **2018**, *86*, 159–176. [CrossRef]
63. Lee, S. The relationship between growth and profit: Evidence from firm-level panel data. *Struct. Chang. Econ. Dyn.* **2014**, *28*, 1–11. [CrossRef]
64. 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]
65. 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]
66. 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]
67. Vithessonthi, C.; Tongurai, J. The effect of firm size on the leverage–performance relationship during the financial crisis of 2007–2009. *J. Multinatl. Financ. Manag.* **2015**, *29*, 1–29. [CrossRef]
68. 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]
69. Marquardt, D.W. Comment. You should standardize the predictor variables in your regression models. *J. Am. Stat. Assoc.* **1980**, *75*, 87–91. [CrossRef]
70. Open Data Science. Open Machine Learning Course. Theme 7. Unsupervised Learning: PCA and Clustering. Available online: https://habr.com/ru/companies/ods/articles/325654/ (accessed on 14 January 2025).
71. Shukla, A.; Das, T.K.; Roy, S.S. TRX Cryptocurrency Profit and Transaction Success Rate Prediction Using Whale Optimization-Based Ensemble Learning Framework. *Mathematics* **2023**, *11*, 2415. [CrossRef]
72. Hwang, S.; Yoon, G.; Baek, E.; Jeon, B.-K. A Sales Forecasting Model for New-Released and Short-Term Product: A Case Study of Mobile Phones. *Electronics* **2023**, *12*, 3256. [CrossRef]
73. Rožanec, J.M.; Fortuna, B.; Mladenić, D. Reframing Demand Forecasting: A Two-Fold Approach for Lumpy and Intermittent Demand. *Sustainability* **2022**, *14*, 9295. [CrossRef]
74. Yurtsever, M. Gold Price Forecasting Using LSTM, Bi-LSTM and GRU. *Eur. J. Sci. Technol.* **2021**, *31*, 341–347. [CrossRef]
75. Hoehreiter, S.; Sehmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]