Research article
DOI: https://doi.org/10.48554/SDEE.2023.3.1

Modelling Profits Forecasts for the Russian Banking Sector Using Random Forest and Regression Algorithms

Nikolay Lomakin[1*], Anastasia Kulachinskaya[2], Svetlana Naumova[1], Maya Ibrahim[1], Evelina Fedorovskaya[1], Ivan Lomakin[1]

[1]Volgograd State Technical University, Volgograd, Russia
[2]Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
[*]Corresponding author: tel9033176642@yahoo.com

Abstract

This study is relevant because market uncertainty induces progressively more attempts at making accurate profits forecasts in the banking sector. The scientific novelty of this study lies in the profits forecasts for the Russian banking sector performed using a random forest machine learning (ML) model and a neural network regression model. Regarding technology, the two models are combined into a cognitive model, as they are executed in the same cloud service (Collab) and have a common dataset comprising a training set, scripts and result output. The aim of the study is to build two models: a random forest ML model and a neural network regression model. The dataset used in the random forest ML model and the regression model included data on the performance of the Russian banking sector and some macroeconomic data on the national economy and the stock market for the period 2017–2021. Specifically, the dataset for the models included the following: key rate (%), growth assets (%), overdue loans (%), gross domestic product (GDP, in billions of rubles), RTS index (points), USD rate (vs. RUB), investments in assets to GDP (%), exchange robots (%), capital outflow (in billions of rubles), bank assets (in trillions of rubles), stock accounts (pcs.), and bank profits (in billions of rubles). The practical relevance of this study is evidenced by the fact that the results of the digital profits forecasting for the Russian banking sector can be recommended for real-world use. In building the cognitive model, we used the Python language in the Collab cloud environment. The mean absolute error of the test set for the random forest ML model (DecisionTreeRegressor) was 414.67, which is 61% lower than for the linear regression model (LinearRegression), which had a mean absolute error of 667.65.

Keywords: digital model, cognitive model, ML model, random forest, profits forecast for banking sector

Citation: Lomakin, N., Kulachinskaya, A., Naumova, S., Ibrahim, M., Fedorovskaya, E., Lomakin, I., 2023. Modelling Profits Forecasts for the Russian Banking Sector Using Random Forest and Regression Algorithms. Sustainable Development and Engineering Economics 3, 1. https://doi.org/10.48554/SDEE.2023.3.1

Научная статья

УДК 368.519.86

DOI: https://doi.org/10.48554/SDEE.2023.3.1

Моделирование Прогноза Прибыли Банковского Сектора РФ с Использованием Модели Случайный Лес и Регрессии

Николай Ломакин[1*], Анастасия Кулачинская[2], Светлана Наумова[1], Майя Ибрахим[1], Эвелина Федоровская[1], Иван Ломакин[1]

[1]Волгоградский государственный технический университет, Волгоград, Россия
[2]Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия
[*]Автор, ответственный за переписку: tel9033176642@yahoo.com

Аннотация

В условиях рыночной неопределенности предпринимается все больше попыток используя системы искусственного интеллекта сформировать точный прогноз величины прибыли банковского сектора. Научная новизна данного исследования заключается в получении прогнозов величины прибыли российского банковского сектора с использованием модели машинного обучения (ML-модель) «Случайный лес» и нейросетевой модели регрессии. Технологически обе модели объединены в «Когнитивную модель», поскольку выполнены в одном «облачном сервисе» Collab, имеют общий датасет – обучающее множество, скрипты и вывод результата. Целью исследования является формирование моделей (ML-модель «Случайный лес» и модель регрессии) для получения прогнозных значений прибыли отечественного банковского сектора и сравнения результатов работы этих моделей. В целях формирования датасета, используемого для обучения модели машинного обучения «Случайный лес» и модели регрессии, использовались данные, отражающие результаты деятельности российского банковского сектора, некоторые макроэкономические показатели отечественной экономики и биржевого рынка за период 2017–2021 гг. В частности, в датасет моделей были включены: Ключевая ставка (%), Прирост банковских активов (%), Доля просроченных кредитов (%), ВВП (млрд руб.), Индекс RTS (пунктов), Курс USD (руб.), Инвестиции в активы к ВВП (%), Доля роботов на бирже (%), Отток капитала (млрд. руб.), Банковские активы (трлн. руб.), Количество счетов на бирже (шт.), Прибыль банков (млрд. руб.). Практическая значимость исследования заключается в том, что результаты цифрового прогнозирования прибыли банковского сектора РФ могут быть рекомендованы для дальнейшего практического применения. При формировании когнитивной модели, использовался язык Python в облачной среде Collab. Средняя ошибка прогноза на тестовом множестве у ML-модели «Случайный лес» (DecisionTreeRegressor) составила 414,67 и на 61% оказалась ниже в сравнении с моделью линейной регрессии (LinearRegression), средняя ошибка которой составила 667,65.

Ключевые слова: цифровая модель, когнитивная модель, ML-модель, случайный лес, прогнозирование прибыли банковского сектора

Цитирование: Ломакин, Н., Кулачинская, А., Наумова, С., Ибрахим, М., Федоровская, Э., Ломакин, И., 2023. Моделирование Прогноза Прибыли Банковского Сектора РФ с Использованием Модели Случайный Лес и Регрессии. Sustainable Development and Engineering Economics 3, 1. https://doi.org/10.48554/SDEE.2023.3.1

1. Introduction

The subject of this study is the performance of the banking sector, i.e., the profits it makes, which is determined by many factors. The focus of the study is the relationship between the profits made in the banking sector and how they are impacted by the factors we investigate. A critical problem in the banking sector and the finance sector is how to ensure financial stability and the stability of the economy as a whole, which cannot be done without having an accurate forecast of the sector's profits for the next year.

This study is relevant because market uncertainty induces progressively more attempts to use artificial intelligence systems to perform accurate profit forecasts for the banking sector. The scientific novelty of this research lies in the profits forecasts for the Russian banking sector made using a random forest machine learning (ML) model and a neural network regression model. Regarding technology, both models are combined into a cognitive model, as they are executed in the same cloud service (Collab) and have a common dataset comprising a training set, scripts and result output.

The practical significance of the study is that the results of the digital profits forecasting for the Russian banking sector can be recommended for real-world application. We used the Python language in the Collab cloud environment to build the cognitive model. The results of the study include the projected value for the sector's gross domestic product (GDP). This was obtained via the digital cognitive model, an integral component of which is the random forest ML model.

In choosing the determining factors to investigate, we relied on the findings of some previous studies; in particular, studies involving an analysis of the state of the banking sector in the Russian Federation (Polyanskaya, 2022) and an investigation of the quality of the loan portfolios and investment activities of banks as profitability factors impacting the financial sector (Vimalaratkhne, 2022).

The aim of this study is to build a random forest ML model and a regression model to forecast the profits of the nation's banking sector and then compare the results of these models.

To achieve this objective, the following hurdles had to be addressed:

1. Study the theoretical basis of a profitable operation of the banking system.

2. Understand the trends in the development of artificial intelligence (AI) systems in the banking sector and the finance sector.

3. Create a dataset for the model.

4. Calculate the projected value of the profits of the banking sector using the random forest ML model.

5. Analyse the results.

The results can be utilised in the credit and finance sector, as well as by investors, the business community, and the academic community. Anyone who needs an accurate profit forecast of the banking sector on an annual forecast horizon would be interested in employing the findings of this study. Economic and financial systems can become consumers of the information generated by the digital cognitive model, which has the random forest ML model as its critical component. It is essential to project the profits of the banking sector using input parameters that vary together with the changing global economic landscape and growing market uncertainty.

Badvan, Gasanov and Kuzminova, who researched various ways of ensuring the stability of financial markets, use cognitive modelling extensively in their study (Badvan et al., 2018). Cognitive modelling of the stability factors impacting financial markets and the creation of cognitive maps are considered in studies by Emelianenko and Kolesnik (Emelianenko et al., 2019).

Notably, given the digitalisation of the economy, all factors (both economic and technological) are essential. Their effect can be observed in the present and, even more importantly, will be felt in the future. Therefore, in the context of transitioning to a new technological paradigm (i.e., Industry 4.0), it is imperative to become familiar with the findings of a study conducted by Rodionov et al., researching the development of an innovation-industrial cluster strategy using a method that employs parallel and sequential real options (Rodionov et. al., 2022). Undoubtedly, attention should be paid to the proposals made by Balog et al. regarding human capital in the digital economy as a factor in sustainable development (Balog, 2022). According to Dianov, sustainable development can be achieved if effective organisational management systems are created (Dianov, 2022). Scientific interest has been sparked by the development of an innovative strategy for an industrial cluster using the concept of composite real options by Koshelev et al. (2023). It is quite possible that the factors studied by these aforementioned scientists can be parsed (collected, digitised and pre-processed) and used in the subsequent versions of the cognitive model.

2. Literature Review

This study is relevant because of the need to ensure the sustainability of the banking sector and the Russian economy as a whole in the face of growing market uncertainty and risk.

To frame the broad ideas and findings of previous studies clearly in this literature review, it is crucial to note—as the main thread—that many classical approaches to forecasting bank profits do not work well or are ineffective in many cases. Modern approaches in the literature are fragmented or inconsistent. However, the general vector of research studies shows that today's trends are characterised by the introduction of increasingly sophisticated AI forecasting systems and the extensive use of big data and business processes common to Industry 4.0.

AI and big data systems are fundamental tools for profits forecasting in the Russian banking sector. With these technologies, banks can analyse enormous amounts of data and identify trends that may affect business profitability. According to a report prepared by Accenture, using AI systems can increase a bank's profit by 34%. In addition, using big data can help banks reduce risks and improve their operational efficiency.[1]

An example of AI and big data systems being used successfully in the Russian banking sector is Sberbank. According to the Banki.ru portal, Sberbank uses an AI system for automatic decision-making regarding credit.[2] It should also be noted that an AI system and big data can help banks optimise costs. According to Forbes, banks can bring down their customer service costs by 20% using these technologies.[3]

Research indicates that the relationship between the categories of profitability and economic stability needs to be closely re-examined because the latter is a complex and multifaceted concept. Many studies by Russian and foreign scientists investigate the problem of the stability of economic systems. These problems have been explored by economists such as Gurvich, Prilepsky, Bobylev and Konishchev (Abdrakhmanova et al., 2019). The challenge of building a cognitive model of the national financial market—given its peculiarities—and the potential use of the model for assessing the operational safety of the market has been studied by Loktionova (2022).

Thus, AI and big data systems are essential tools for forecasting the profits of the Russian banking sector, as they help banks analyse enormous amounts of data, identify trends and make decisions that may affect the profitability of their businesses.

Today, it is important to study issues related to AI used to ensure sustainable economic development and reduce financial risks because of growing market uncertainty. Researchers such as Abdalmuttaleb and Al-Sartavi have reviewed the latest studies on AI applied to stable financing and sustainable technologies (Abdalmuttaleb, 2021). As presented by Lomakin et al. (2019) in the Global Economic Revolutions: The Era of Digital Economy international conference, the neural network model can be

[1]Accenture. Artificial intelligence in banking. URL: https://www.accenture.com/us-en/insights/banking/artificial-intelligence-in-banking Accessed on April 22, 2023.
[2]Banki.ru. Sberbank is using Artificial Intelligence when granting loans. URL: https://www.banki.ru/news/lenta/?id=10124323 Accessed on April 22, 2023.
[3]Forbes. How AI and big data can cut banks' costs by 20%.
URL: https://www.forbes.com/sites/tomgroenfeldt/2019/05/23/how-ai-and-big-data-can-cut-banks-costs-by-20/?sh=3b5f5a5d5c98 Accessed on April 22, 2023.

used to project the profits of enterprises operating in the real sector of the economy. Certain aspects of using neural networks in the financial sector intersect with economic analysis in financial management systems, as noted by Morozova, Polyanskaya, Zasenko, Zarubina and Verchenko. Notably, for an enterprise to operate effectively in today's economy, with ever-increasing competition, it must respond promptly to any change in any of the different factors that affect its operations (Morozova, et al., 2017).

A key aspect of the financial stability of the economy is the reliable operation of the banking sector. One of the most pressing issues regarding achieving this stability is preventing the growth of overdue debts. To achieve this goal, the creditworthiness and financial stability of enterprises must be assessed. Rybyantseva, Ivanova, Demin, Jamai and Bakharev studied various approaches to such an assessment and identified the most effective among them (Rybyantseva, et al., 2017). Hengxu Lin, Dong Zhou, Weiqing Liu and Jiang Bian proposed a deep risk model as a solution for deep learning and analysis of hidden risk factors. They experimented with stock market data and demonstrated the high efficiency of their solution. Their method allows users to achieve 1.9% more of the detected variance and reduces the risk of a global minimum variance portfolio (Hengxu et al., 2021). An important aspect of financial stability is the formulation of an investment portfolio. Of practical interest are the studies by Ni Zhang, Yijia Song, Aman Jakhar and He Liu on the development of graphical models of financial time series and the selection of a portfolio. They propose various graphical models for building the best portfolios (Zhan et al., 2021).

3. Materials and Methods

This study employs research methods such as monographic, analytical, statistical and cognitive models, including a random forest AI system and a program called Graphviz (a utility package developed by AT&T laboratories for automatic visualisation of graphs). The methodology employed in this study is based on a cognitive model.

A cognitive model is a software shell: a bot that collects information, creates a dataset, obtains and compares results, assesses the weight of parameters (based on the magnitude of correlation coefficients) if necessary, and removes weak factorial features from the training set. A cognitive model is expected to work cyclically.

With respect to technology, the two models (the random forest AI system and the multiple regression algorithm) are combined into a cognitive model, as they are executed in the same cloud service, Collab and have a common dataset (a training set, scripts and result output).

Financial and economic stability is modelled based on the cognitive model, which allows us to develop an original approach to supporting management decision-making at times of uncertainty through the ability to accurately forecast the profitability of the Russian banking sector.

This research proposes and attempts to substantiate the hypothesis that, at a time of uncertainty, when all types of risk are growing, the random forest ML model can be used to forecast the profits of the banking sector more accurately than a multivariate regression model.

The profitable operation of banks is closely related to their stability and the stability of the country's economy as a whole. Both Russian and foreign scientists are increasingly interested in the concept of the stability of financial and economic systems. Problems related to financial stability have been studied by many Western scientists, including John Chant, Andrew Crockett, Wim Duisenberg, Roger Ferguson, Michael Foot, Sir Andrew Large, Frederick Mishkin, and Garry Schinasi.

The deeper the tree, the more complex the decision-making rules and the more accurate the model. There are two types of decision trees used for both classification and regression problems. An understanding of the importance of variables in random tree forests is expressed in many studies, including one by Louppe et al. (2020).

A cognitive model acts as a trigger that launches methods as independent modular programs; in particular, it launches a decision tree that can be used to obtain forecasts of the profits of the banking system. The dataset of the decision tree model used in this study is presented in Table 1.

Table 1. Data used to create the dataset for the random forest ML model (fragment)

Year	Key Rate	Growth Assets (%)	Overdue Loans (%)	GDP (billions of rubles)	RTS Index	USD Rate
2021	8.50	16.0	23.5	131015	1608	73.7
2020	4.25	16.8	17.8	1073015	1376	73.8
2019	7.25	10.4	5.9	109241	1549	61.9
2018	7.75	6.4	7.5	103861	1157	69.8
2017	8.25	−3.5	9.3	91843	1154	57.6

Investments in Assets to GDP (%)	Exchange Robots (%)	Capital Outflow (billion rubles)	Bank Assets (trillion roubles)	Stock Accounts (pcs.)	Bank Profits (billion rubles)
21.2	58	72.0	120.0	38300	2400.0
16.5	55	53.0	103.7	32300	1608.0
20.6	55	25.2	92.6	3069	1715.0
20.6	51	60.0	92.1	1955	1705.0
21.4	51	33.3	85.2	1310	1300.0

The data presented in Table 1 were collected manually, but the process can be automated using a data parsing program. The ML model was generated in the cloud by Google Collab using Python programming language.

Describing the sample seems worthwhile. To create a dataset for training the random forest ML model and regression model, we used performance data on the Russian banking sector and macroeconomic data on the national economy and the stock market for the period 2017–2021. In particular, the dataset for the models included the following: key rate (%), growth assets (%), overdue loans (%), GDP (in billions of rubles), RTS index (points), USD Rate (vs. RUB), investments in assets to GDP (%), exchange robots (%), capital outflow (in billions of rubles), bank assets (trillion rubles), stock accounts (pcs.) and bank profits (in billions of rubles).

4. Results

4.1. Digital Cognitive Model

The Graphviz program was used to visualise the digital cognitive model. Graphviz is a utility package offered by AT&T laboratories for the automatic visualisation of graphs based on their textual descriptions. The package is distributed as an open-source code file and runs on Windows and other operating systems.

The cognitive model acts as a kind of trigger that launches methods as independent modular programs; in particular, it launches a decision tree that can be used to obtain a profits forecast of the banking system. Figure 1 is a schematic diagram of the digital cognitive model.

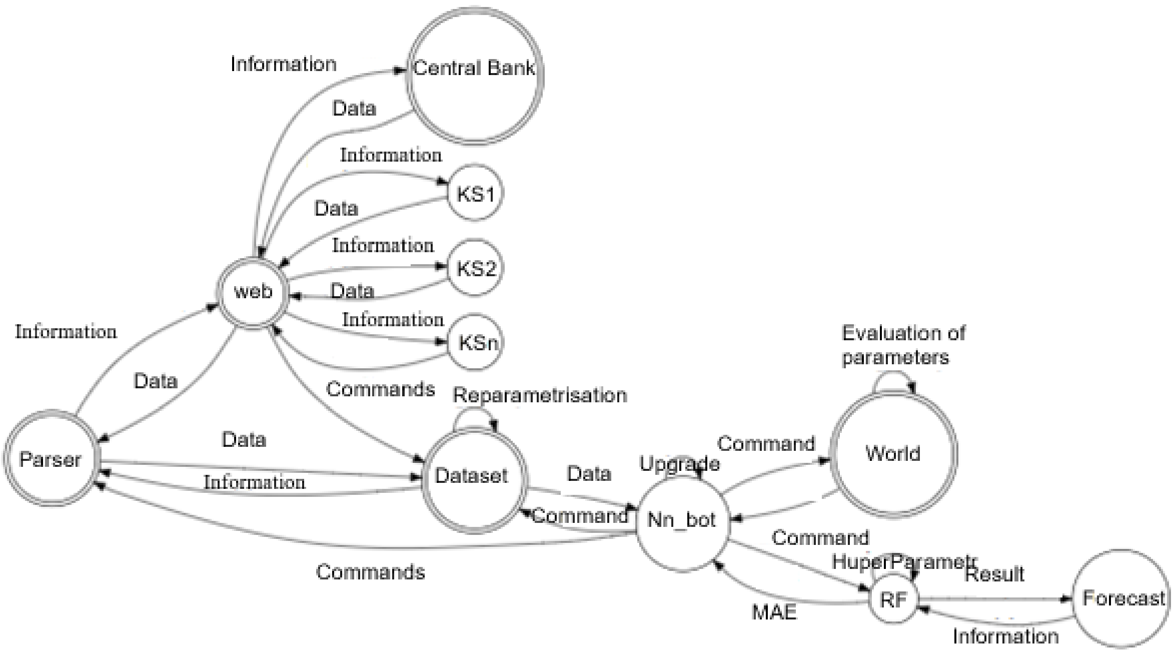

Figure 1. Digital cognitive model

The concept behind the cognitive model is based on the interaction of its main modules, the ultimate goal of which is to collect all the necessary information, process it and then create a dataset for the random forest ML model and a heat map of the pairwise coefficients of the multifactor linear regression model, which return the predicted profits of the banking sector. An integral element of the digital cognitive model is the random forest ML model, which performs the neural forecasting of banking sector profits.

4.2. Random Forest ML Model

A random forest is an ML learning algorithm that uses an assembly of decision trees to solve classification and regression problems. It is applied to different sectors, including finance, medicine and business, and is suitable for improving the accuracy of forecasts and reducing the probability of retraining the model.

Decision trees (DT) are based on a nonparametric learning method with a teacher and are used for classification and regression. The purpose of this method is to create a model that predicts the value of the target variable based on the study of simple decision-making rules obtained from the characteristics of the data. The tree can be considered a piecewise constant approximation. Table 2 presents the dataset of the random forest ML model.

Table 2. Random forest ML model dataset (fragment)

	Key rate	Growth assets	Overdue loans	GDP	RTS	USD	Investments	Exchange robots	Capital outflow	Bank assets	Stock accounts	target
0	8.50	16.0	23.5	131015.0	1609.7	73.70	21.2	58	72.0	120.0	38300	2400.0
1	4.25	16.8	17.8	107315.3	1376.4	73.80	16.5	55	53.0	103.7	32300	1608.0
2	7.25	10.4	5.9	109241.5	1549.4	61.98	20.6	55	25.2	92.6	3069	1715.0

A binary classification tree (i.e., regression) (Breiman et al., 1984) is an input-output model represented by a tree structure T from a random input vector ($X_1 \ldots X_p$), taking its values in ($X_1 * \ldots * X_p$)=X into a random output variable Y ϵ Υ. The tree is built from a training set of size N, taken from P($X_1 \ldots X_p$,Y) and using a recursive procedure that in each node t identifies partition $s_t=s^*$, for which the partition of the samples of node N_t into t_L and t_R maximises the reduction of a certain impurity measure i(t) (e.g., the Gini index, the Shannon entropy, and Y variance) (Equation 1).

$$\Delta_i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R), \quad (1)$$

where $p_L = N_t L / N_t$ and $p_R = N_t R / N_t$

The building of the tree stops, for example, when the nodes become pure along Y or when all variables X_i are locally constant. The tree is finally exported and mapped in the tree structure presented in Figure 2, which is visualised using a special service[4] by copying the data from the tree '.file' with a dot. Figure 2 shows the first level of the decision tree.

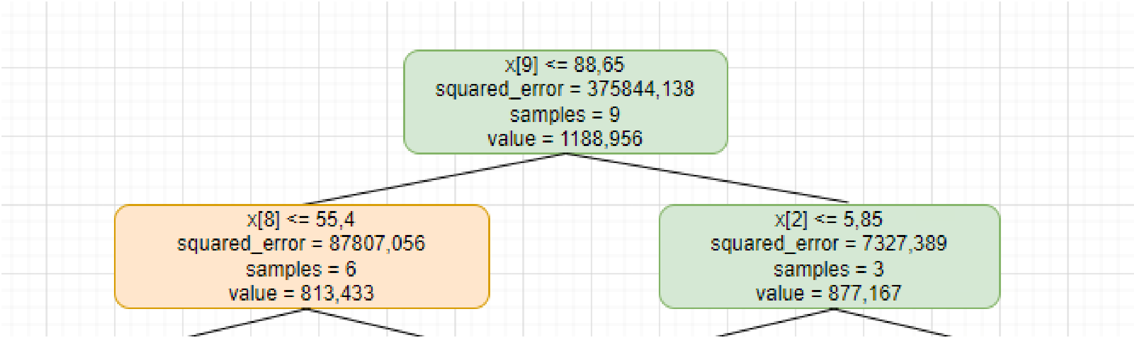

Figure 2. First two levels of the decision tree

To forecast the profits of the banking sector for the next year, you need to use a specific script in which the latest values of the input parameters are introduced.

4.3. Multivariate Linear Regression Model

An AI multivariate linear regression model was used to forecast the profits of the banking sector. The multivariate linear regression model is also used to project the value of a target indicator based on the values of several features; however, it relies on a linear combination of these features. In each *i-th* observation, we obtain a set of values of independent variables and the corresponding value of the dependent variable Y_i. If we assume that there is a linear relationship between the independent variables *x1, x2,... xi* and the dependent variable Y_i, then Equation 2

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m + \varepsilon \quad (2)$$

expressing the linear relationship between variables is called a theoretical multiple regression equation.

In the course of the study, a matrix of pairwise correlation coefficients was obtained (Table 3).

Table 3. Matrix of pairwise correlation coefficients

	Key rate	Growth assets	Overdue loans	GDP	RTS	USD	Investments	Exchange robots	Capital outflow	Bank assets	Stock accounts	target
Key rate	1.000	0.042	-0.072	-0.058	-0.550	0.280	0.445	0.207	0.123	-0.039	-0.300	-0.323
Growth assets	0.042	1.000	-0.034	-0.166	-0.152	-0.023	-0.255	-0.130	0.283	-0.163	0.078	-0.368
Overdue loans	-0.072	-0.034	1.000	0.720	0.175	0.620	-0.351	0.613	-0.051	0.755	0.948	0.675
GDP	-0.058	-0.166	0.720	1.000	-0.048	0.835	-0.199	0.901	-0.046	0.973	0.686	0.741
RTS	-0.550	-0.152	0.175	-0.048	1.000	-0.461	0.048	-0.373	-0.290	-0.088	0.320	0.518
USD	0.280	-0.023	0.620	0.835	-0.461	1.000	-0.367	0.957	-0.017	0.877	0.519	0.419
Investments	0.445	-0.255	-0.351	-0.199	0.048	-0.367	1.000	-0.249	-0.043	-0.280	-0.497	-0.119
Exchange robots	0.207	-0.130	0.613	0.901	-0.373	0.957	-0.249	1.000	-0.128	0.921	0.506	0.479
Capital outflow	0.123	0.283	-0.051	-0.046	-0.290	-0.017	-0.043	-0.128	1.000	-0.180	0.053	-0.306
Bank assets	-0.039	-0.163	0.755	0.973	-0.088	0.877	-0.280	0.921	-0.180	1.000	0.693	0.744
Stock accounts	-0.300	0.078	0.948	0.686	0.320	0.519	-0.497	0.506	0.053	0.693	1.000	0.686
target	-0.323	-0.368	0.675	0.741	0.518	0.419	-0.119	0.479	-0.306	0.744	0.686	1.000

[4]Graphviz in the Browser. URL: http://www.webgraphviz.com

The multifactorial linear regression model considers that the relations between mass economic phenomena are dependent on the fact that—in reality—a certain phenomenon is determined by a multitude of simultaneously and collectively acting causes. Therefore, in a general case, a dependent variable can be a function of several variables.

To visualise the matrix of pairwise correlation coefficients, it is advisable to use a heat map (Figure 3).

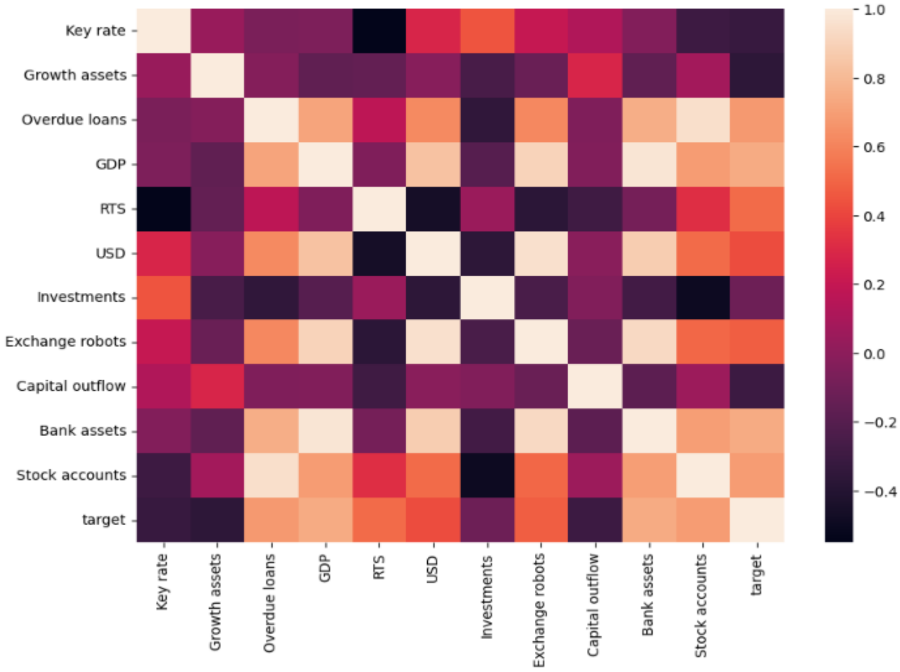

Figure 3. Heat map of the multivariate linear regression model

The correlation coefficients between factorial and resultant features are as follows: the key rate is −0.323, growth assets (%) are −0.368, overdue loans (%) are 0.675, GDP (in billions of rubles) is 0.741, the RTS index is 0.518, the USD rate is 0.419, investments in assets to GDP (%) are −0.119, exchange robots (%) are 0.479, capital outflow (in billions of rubles) is −0.306, bank assets (in trillions of rubles) are 0.744, and stock accounts are 0.686.

Using the Pandas and lin_reg.coef libraries, we calculated the regression equation coefficients, which are presented in Table 4.

Table 4. Regression equation coefficients

Key rate	Growth assets	Overdue loans	GDP	RTS	USD	Investments	Exchange robots	Capital outflow	Bank assets	Stock accounts
−4.0337	−39.572	103.239	−0.0953	−6.8288	−107.00	−11.027	−91.4689	−2.9446	89.20268	3.60045

It is important to analyse the results obtained.

4.4. Analysing the Results

The quality of the forecast was assessed based on a comparison of the following parameters:

1. Mean absolute error.

2. Mean squared error, which is applied in case we need to highlight large errors and then choose the model that results in fewer large errors for the forecast.

3. Root mean squared deviation (RMSD) or root mean squared error (RMSE), which is a commonly used measure of disparity between the values (sample or population) predicted by a model or an assessor and the actual observed values. The RMSD is the square root of the second sampling moment

of differences between the predicted values and the observed values, or the root mean squared value of these differences. These deviations are either called excesses, when the calculations are made with the data sample used for the assessment, or errors (also prediction errors), if the calculations are made beyond the sample.

An analysis of the findings shows that the ML model ensures a more precise result than the multifactor linear regression model (Table 5).

Table 5. Comparison of the results of using the ML model and a linear regression model

Name	DecisionTreeRegressor	LinearRegression	Deviation (%)
Mean Absolute Error	414.6666667	667.6533333	0.610096463
Mean Squared Error	232246	1325.48	−0.994292776
Root Mean Squared Error	481.9190803	1361.887	1.825966133

The mean absolute error of the forecast for the test set of the random forest ML model (DecisionTreeRegressor) was 414.67, which proved to be 61% lower than that for the linear regression model (LinearRegression), which had a mean absolute error of 667.65.

5. Discussion

It seems reasonable that the views and results obtained in this study should be thought over critically. Undoubtedly, the results are consistent with those of other published studies in the international academic domain.

In the course of this study, we solved the problems that had been identified as hurdles and obtained the following outcomes: the theoretical basis of profitable operation of the banking sector was investigated, the development trends of AI systems in the banking and finance spheres were studied, a dataset for the ML model was created, profits forecasts for the banking sector were calculated using a random forest ML model, and the results obtained were analysed.

The mean absolute error of the forecast for the test data was 414.67 for the random forest ML model (DecisionTreeRegressor), which is 61% lower than that for the linear regression model (LinearRegression), which has a mean absolute error of 667.65. Comparing the results obtained with the issues discussed in the introduction, we can say that other advanced neural network models should be used in future research.

A convolutional neural network (CNN) is a deep learning algorithm that can accept input parameters and assign weight (digestible weights and biases) to various areas/objects depending on the purpose of study. Due to the growing computing power of modern cloud clusters, modern neural CNN-based algorithms can be used with parallel calculations in open Hadoop and Spark frameworks to make complex economic and financial forecasts.

More sophisticated AI models should be applied in future research. AI is increasingly used in robotic advising, and the financial sector is no exception. Catherine D'Hondt, Rudy De Wynn, Eric Giesels and Steve Raymond studied the use of an AI alter ego system in the field of robotic investments, introducing the concept of AI AlterEgo, which is a type of shadow robot investor (D'Hondt, 2019). One of the promising areas where deep neural networks can be used is the banking sector. For example, Krzysztof et al. propose performing a neural risk assessment of networks with unreliable resources (Krzysztof, 2022).

Our cognitive model opens wide opportunities for AI systems that are suitable for providing management decision support, forecasting banking sector profits and increasing the stability of the economic and financial sector.

6. Conclusion

In this study, we came to the following conclusions:

Using a digital cognitive model, with a random forest ML system as its integral component, is essential for achieving stable economic growth based on forecasting banking sector profits because it stimulates the competitiveness of the national economy.

Using the results of the digital cognitive model, which has a random forest ML system as its integral component, opens ample opportunities for applying AI systems in management decision support, thus increasing the profitability of the banking sector and improving economic stability.

The results obtained in this study have practical significance, and the proposed algorithm can be used to forecast banking sector profits. The mean absolute error of the forecast for the test set of the random forest ML model (DecisionTreeRegressor) was 414.67, which is 61% lower than that of the linear regression model, which had a mean absolute error of 667.65.

References

Abdalmuttaleb, M.A., 2021. Artificial intelligence for sustainable finance and sustainable technology. M. Al-Sartawi. ICGER: The International Conference on Global Economic Revolutions, LNNS, Vol. 423, pp. 15–16. https://doi.org/10.1007/978-3-030-93464-4

Abdrakmanova, G.I., Vishnevsky, K.O., Gokhberg, L.M. et al., 2019. Digital Economics: A Brief Statistical Collection, National Research University, Higher School of Economics, Moscow: HSE, p. 96. https://doi.org/10.17323/978-5-7598-2599-9

Badvan, N.L., Hasanov, O.S., Kuzminov, A.N., 2018. Cognitive Modeling of the Stability Factors of the Russian Financial Market. Finance and Credit 24(5), 1131–1148. https://doi.org/10.24891/fc.24.5.1131

Balog, M., Demidova, S., Lesnevskaya, N., 2022. Human capital in the digital economy as a factor of sustainable development. Sustainable Development and Engineering Economics 1, 3. https://doi.org/10.48554/SDEE.2022.1.3

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., 2022. Classification and regression

D'Hondt, C., De Winne, R., Ghysels, E. and Steve Raymond 2019. Artificial Intelligence Alter Egos: Who Benefits from Robo-investing? Portfolio Management (q-fin.PM); Econometrics (econ.EM); Statistical Finance (q-fin.ST) https://doi.org/10.48550/arXiv.1907.03370

Dianov, S., Isroilov, B., 2022. Formation of effective organisational management systems. Sustainable Development and Engineering Economics 1, 2. https://doi.org/10.48554/SDEE.2022.1.2

Emelianenko, A.S., Kolesnik, D.V., 2019. Building Cognitive Maps. The Matters of Student Science 12(40), 309–316.

Hengxu, L., Dong, Z., Weiqing, L. and Jiang, B., 2021. Deep Risk Model: A Deep Learning Solution for Mining Latent Risk Factors to Improve Covariance Matrix Estimation. ICAIF'21. November 3–5. Virtual Event. USA https://arxiv.org/format/2107.05201 (accessed 10.12.2022).

Koshelev, E., Dimopoulos, T., Mazzucchelli, E.S., 2021. Development of innovative industrial cluster strategy using compound real options. Sustainable Development and Engineering Economics 2, 5. https://doi.org/10.48554/SDEE.2021.2.5

Krzysztof R., Piotr B., Piotr Jaglarz, Fabien G., Albert C., Piotr C., 2022. RiskNet: Neural Risk Assessment in Networks of Unreliable Resources. https://doi.org/10.48550/arXiv.2201.12263

Loktionova, E.A., 2022. Cognitive model of the national financial market: features of construction and the possibility of using it to assess the safety of its functioning. Finance: theory and practice 26(1), 126–132.

Lomakin, N., Lukyanov, G., Vodopyanova, N., Gontar, A., Goncharova, E. and Voblenko, E., 2019. Neural network model of interaction between real economy sector entrepreneurship and financial field under risk. Advances in Economics. Business and Management Research. volume 83. 2nd International Scientific Conference on 'Competitive. Sustainable and Safe Development of the Regional Economy' (CSSDRE 2019) (accessed 10.20.2022). https://doi.org/10.2991/cssdre-19.2019.51

Louppe, G., Wehenkel, L., Sutera, A. and Geurts, P., 2020. Understanding variable importances in forests of randomized trees. P. 9–10 https://proceedings.neurips.cc/paper/2013/file/e3796ae838835da0b6f6ea37bcf8bcb7-Paper.pdf (accessed 24.02.2023).

Morozova, T.V., Polyanskaya, T., Zasenko, V.E., Zarubin, V.I. and Verchenko Y.K., 2022. Economic Analysis in the Financial. Management System 15, 117–124.

Polyanskaya, A.A., Nersesov, V.S., Lomakin, N.I., 2022. Analysing the State of the Banking Sector in the Russian Federation. Russian Science in Today's World: Collection of papers of the XLVI International Scientific-Practical Conference. Moscow, May 31, 2022, 214–215.

Rodionov, D., Koshelev, E., Escobar-Torres, L., 2022. Formation of innovative-industrial cluster strategy by parallel and sequential real options. Sustainable Development and Engineering Economics 2, 1. https://doi.org/10.48554/SDEE.2022.2.1

Rybyantseva, M., Ivanova, E., Demin, S., Dzhamay, E. and Bakharev, V., 2017. Financial sustainability of the enterprise and the main methods of its assessment. International Journal of Applied Business and Economic Research 15, 139–146.

Vimalaratkhne, K., Fedorovskaya, E.O., Lomakin, N.I., 2022. Studying the Quality of a Loan Portfolio and Investment Activity of Banks as Profitability Factors of the Financial Sector. Problems of Competitiveness of Consumer Goods and Food Products: Collection of Papers of the 4th International Scientific-Practical Conference, April 13, 2022, Kursk, 59–64.

Zhan, N., Sun, Y., Jakhar, A. and Liu, H., 2021. Graphical Models for Financial Time Series and Portfolio Selection, In: ACM International Conference on A.I. in Finance (ICAIF '20). https://arxiv.org/format/2101.09214 (accessed 24.02.2023).

Список источников

Abdalmuttaleb, M.A., 2021. Artificial intelligence for sustainable finance and sustainable technology. M. Al-Sartawi. ICGER: The International Conference on Global Economic Revolutions, LNNS 423, 15–16. https://doi.org/10.1007/978-3-030-93464-4

Balog, M., Demidova, S., Lesnevskaya, N., 2022. Human capital in the digital economy as a factor of sustainable development. Sustainable Development and Engineering Economics 1, 3. https://doi.org/10.48554/SDEE.2022.1.3

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., 2022. Classification and regression

D'Hondt, C., De Winne, R., Ghysels, E. and Steve Raymond 2019. Artificial Intelligence Alter Egos: Who Benefits from Robo-investing? Portfolio Management (q-fin.PM); Econometrics (econ.EM); Statistical Finance (q-fin.ST) https://doi.org/10.48550/arXiv.1907.03370

Dianov, S., Isroilov, B., 2022. Formation of effective organisational management systems. Sustainable Development and Engineering Economics 1, 2. https://doi.org/10.48554/SDEE.2022.1.2

Hengxu, L., Dong, Z., Weiqing, L. and Jiang, B., 2021. Deep Risk Model: A Deep Learning Solution for Mining Latent Risk Factors to Improve Covariance Matrix Estimation. ICAIF'21. November 3–5. Virtual Event. USA https://arxiv.org/format/2107.05201 (accessed 10.120.2022).

Koshelev, E., Dimopoulos, T., Mazzucchelli, E.S., 2021. Development of innovative industrial cluster strategy using compound real options. Sustainable Development and Engineering Economics 2, 5. https://doi.org/10.48554/SDEE.2021.2.5

Krzysztof R., Piotr B., Piotr Jaglarz, Fabien G., Albert C., Piotr C., 2022. RiskNet: Neural Risk Assessment in Networks of Unreliable Resources. https://doi.org/10.48550/arXiv.2201.12263

Lomakin, N., Lukyanov, G., Vodopyanova, N., Gontar, A., Goncharova, E. and Voblenko, E., 2019. Neural network model of interaction between real economy sector entrepreneurship and financial field under risk. Advances in Economics. Business and Management Research. volume 83. 2nd International Scientific Conference on 'Competitive. Sustainable and Safe Development of the Regional Economy' (CSSDRE 2019). https://doi.org/10.2991/cssdre-19.2019.51

Louppe, G., Wehenkel, L., Sutera, A. and Geurts, P., 2020. Understanding variable importances in forests of randomized trees. P. 9-10 https://proceedings.neurips.cc/paper/2013/file/e3796ae838835da0b6f6ea37bcf8bcb7-Paper.pdf (accessed 24.02.2023).

Morozova, T.V., Polyanskaya, T., Zasenko, V.E., Zarubin, V.I. and Verchenko Y.K., 2022. Economic Analysis in the Financial. Management System 15, 117–124.

Rodionov, D., Koshelev, E., Escobar-Torres, L., 2022. Formation of innovative-industrial cluster strategy by parallel and sequential real options. Sustainable Development and Engineering Economics 2, 1. https://doi.org/10.48554/SDEE.2022.2.1

Rybyantseva, M., Ivanova, E., Demin, S., Dzhamay, E. and Bakharev, V., 2017. Financial sustainability of the enterprise and the main methods of its assessment. International Journal of Applied Business and Economic Research 15, 139–146.

Zhan, N., Sun, Y., Jakhar, A. and Liu, H., 2021. Graphical Models for Financial Time Series and Portfolio Selection, In: ACM International Conference on A.I. in Finance (ICAIF '20). https://arxiv.org/format/2101.09214 (accessed 24.02.2023).

Абдрахманова, Г.И., Вишневский, К.О., Гохберг, Л.М. и др., 2019. Цифровая экономика: краткий статистический сборник. Национальный исследовательский университет. Высшая школа экономики. Москва, НИУ ВШЭ, с.96. https://doi.org/10.17323/978-5-7598-2599-9

Бадван, Н.Л., Гасанов, О.С., Кузьминов, А.Н., 2018. Когнитивное моделирование факторов устойчивости финансового рынка России. Финансы и кредит 24(5), 1131–1148. https://doi.org/10.24891/fc.24.5.1131

Вималаратхне, К., Федоровская, Э.О., Ломакин, Н.И. Исследование качества кредитного портфеля и инвестиционной деятельности банков как факторов прибыльной работы финансового сектора. Проблемы конкурентоспособности потребительских товаров и продуктов питания, 4 Междунар. науч.-практ. конф., 13 апреля 2022, Курск, 59–64.

Емельяненко, А.С., Колесник, Д.В., 2019. Процесс построения когнитивных карт. Вопросы студенческой науки 12(40), 309–316.

Локтионова, Е.А., 2022. Когнитивная модель национального финансового рынка: особенности построения и возможности использования для оценки безопасности его функционирования. Финансы: теория и практика 26(1), 126–132. https://doi.org/10.26794/2587-5671-2022-26-1-126-143

Полянская, А.А., Нерсесов, В.С., Ломакин, Н.И., 2022. Анализ состояния банковского сектора Российской Федерации. Российская наука в современном мире, XLVI междунар. науч.-практ. конф., Москва, 31 мая 2022 г. 214–215.

The article was submitted 02.06.2023, approved after reviewing 28.07.2023, accepted for publication 03.08.2023.

Статья поступила в редакцию 02.06.2023, одобрена после рецензирования 28.07.2023, принята к публикации 03.08.2023.

About the authors:

1. Nikolay Lomakin, Candidate of Economics, Associate Professor, Volgograd State Technical University, Volgograd, Russia. https://orcid.org/0000-0001-6597-7195, tel9033176642@yahoo.com

2. Anastasia Kulachinskaya, Candidate of Economics, Associate Professor at the Graduate School of Industrial Economics, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia. https://orcid.org/0000-0002-6849-4313, a.kulachinskaya@yandex.ru

3. Svetlana Naumova, Postgraduate Student, Volgograd State Technical University, Volgograd, Russia. https://orcid.org/0000-0001-9932-9866, svetanaumova-sn2015@yandex.ru

4. Maya Ibrahim, Postgraduate Student, Volgograd State Technical University, Volgograd, Russia. https://orcid.org/0009-0003-4374-8625, maya.ibrahim1992@gmail.com

5. Evelina Fedorovskaya, Postgraduate Student, Volgograd State Technical University, Volgograd, Russia. https://orcid.org/0000-0002-3895-8930, evelinaelvis@gmail.com

6. Ivan Lomakin, Postgraduate Student, Volgograd State Technical University, Volgograd, Russia. https://orcid.org/0000-0001-7392-1554, ivan.grom0boy@yandex.ru

Информация об авторах:

1. Николай Ломакин, к.э.н., доцент, Волгоградский государственный технический университет, Волгоград, Россия. https://orcid.org/0000-0001-6597-7195, tel9033176642@yahoo.com

2. Анастасия Кулачинская, к.э.н., доцент Высшей инженерно-экономической школы, Санкт-Петербургский университет Петра Великого, Санкт-Петербург, Россия. https://orcid.org/0000-0002-6849-4313, a.kulachinskaya@yandex.ru

3. Светлана Наумова, магистрант, Волгоградский государственный технический университет, Волгоград, Россия. https://orcid.org/0000-0001-9932-9866, svetanaumova-sn2015@yandex.ru

4. Майя Ибрахим, магистрант, Волгоградский Государственный Технический Университет, Волгоград, Россия. https://orcid.org/0009-0003-4374-8625, maya.ibrahim1992@gmail.com

5. Эвелина Федоровская, магистрант, Волгоградский государственный технический университет, Волгоград, Россия. https://orcid.org/0000-0002-3895-8930, evelinaelvis@gmail.com

6. Иван Ломакин, магистрант, Волгоградский государственный технический университет, Волгоград, Россия. https://orcid.org/0000-0001-7392-1554, ivan.grom0boy@yandex.ru