УДК 004.8
DOI: 10.26102/2310-6018/2022.38.3.004

On the approach to forecasting indicators of socio-economic development of the region based on indirect indicators

V.R. Abbazov[1,2], V.A. Baluev[1], V.V. Burlutsky[1,3], A.V. Melnikov[1,2], M.A. Rusanov[1,2]✉

[1]Yugorsky Research Institute of Information Technologies, Khanty-Mansiysk, Russian Federation,
[2]Yugra State University, Khanty-Mansiysk, Russian Federation
[3]Yugoria Group of Insurance Companies, Khanty-Mansiysk, Russian Federation
RusanovMA@uriit.ru✉

Abstract. Economic and social development requires constant modernization of the regional management system based on the system of key socio-economic indicators of the region's development and methods of their analysis and forecasting. The article proposes a comprehensive approach to forecasting based on the application of classical forecasting methods for existing time series of statistical indicators and by identifying and analyzing indirect semantically close indicators to a new indicator in the absence of the necessary time series for forecasting. The article provides a general methodology for obtaining a forecast and describes in detail the method for constructing a forecast estimate of the change dynamics in the estimated indicator as well as a description of the AutoML library with open source FEDOT, which was used to build forecasts. The issue of constructing and optimizing a combined forecast with the aid of automatic machine learning tools is considered. At the end of the article, the result of an experiment on predicting the indicators “Population of the subject of the Russian Federation” and “Life expectancy at birth” according to the proposed approaches and a comparison of the findings is presented. It can be concluded that the suggested approach to making a predictive assessment of the change dynamics in the estimated indicator by identifying indirect indicators can be applied to socio-economic indicators of the development of the region.

Keywords: socio-economic indicators, forecasting, incompleteness, AutoML, indicator of senior official activity effectiveness.

For citation: Abbazov V.R., Baluev V.A., Burlutsky V.V., Melnikov A.V., Rusanov M.A. On the approach to forecasting indicators of socio-economic development of the region based on indirect indicators. *Modeling, Optimization and Information Technology*. 2022;10(3). Available from: https://moitvivt.ru/ru/journal/pdf?id=1202 DOI: 10.26102/2310-6018/2022.38.3.004

О подходе к прогнозированию показателей социально-экономического развития региона на основе косвенных показателей

В.Р. Аббазов[1,2], В.А. Балуев[1], В.В. Бурлуцкий[1,3], А.В. Мельников[1,2], М.А. Русанов[1,2]✉

[1]Югорский научно-исследовательский институт информационных технологий, Ханты-Мансийск, Российская Федерация,
[2]Югорский государственный университет, Ханты-Мансийск, Российская Федерация,
[3]Группа страховых компаний «Югория», Ханты-Мансийск, Российская Федерация
RusanovMA@uriit.ru✉

Резюме. Экономическое и общественное развитие требует постоянной модернизации системы управления, основанной на системе ключевых социально-экономических показателей развития региона и методах их анализа и прогнозирования. В статье предлагается комплексный подход к построению прогноза как на основе применения классических методов для существующих временных рядов статистических показателей, так и посредством выявления и анализа косвенных, семантически близких к новому показателей, в случае отсутствия у него необходимого для прогноза временного ряда. Приведена общая методика получения прогноза и подробно описана методика построения прогнозной оценки динамики изменения расчетного показателя, а также приведено описание библиотеки AutoML с открытым исходным кодом FEDOT, которая использовалась для построения прогноза. Рассмотрен вопрос построения и оптимизации комбинированного прогноза на основе автоматических средств машинного обучения. В заключении представлен результат эксперимента по прогнозированию показателей «Население субъекта Российской Федерации» и «Ожидаемая продолжительность жизни при рождении» по предложенным подходам и сравнение полученных результатов. Сделан вывод, что предложенный подход к формированию прогнозной оценки динамики изменения расчетного показателя на основе выделения косвенных показателей может быть применен к социально-экономическим показателям развития региона.

Ключевые слова: социально-экономические показатели, прогнозирование, неполнота, AutoML, показатель деятельности высших должностных лиц.

Для цитирования: Аббазов В.Р., Балуев В.А., Бурлуцкий В.В., Мельников А.В., Русанов М.А. О подходе к прогнозированию показателей социально-экономического развития региона на основе косвенных показателей. *Моделирование, оптимизация и информационные технологии*. 2022;10(3). Доступно по: https://moitvivt.ru/ru/journal/pdf?id=1202 DOI: 10.26102/2310-6018/2022.38.3.004 (на англ.).

Introduction

The ability to assess and forecast socio-economic indicators is important for the task of development planning and monitoring the performance of state power.

In 2021, the methodology for calculating socio-economic effectiveness indicators of the senior official activities in the constituent entities of the Russian Federation was approved on the basis of the Decree of the President of the Russian Federation dated 04.02.2021 No.68 "On estimating the effectiveness of the activities of senior officials of the constituent entities of the Russian Federation and the activities of executive authorities of the constituent entities of the Russian Federation" [1,2].

This methodology describes the calculation of 20 different indicators such as population size, life expectancy at birth, poverty level, housing construction volume, and others.

There are explanations in the methodology [2] regarding the calculations of indicators and data sources for these calculations, but problems exist with extracting and insufficient volume of primary data from various regional and state information systems for the application of forecasting methods.

The availability of primary data can be described in two scenarios: there is a sufficiently large time series for the direct calculation of the desired indicator and its mathematical forecast; statistical data is insufficient and, therefore, it is necessary to use expert and other subjective methods of predicting the indicator.

Despite the difficulties described above in obtaining primary data and the problems of their quality, it is necessary to calculate the value and make a forecast of statistical indicators in both cases.

Currently, methods for forecasting statistical indicators with a sufficient size of time series have been studied quite fully, but there is no universal forecasting approach for indicators with insufficient data for previous time periods.

The purpose of the research is to test the hypothesis about the possibility of forecasting key effectiveness indicators of senior official activities on the basis of indirect indicators.

This article describes the forecasting methodology for scenarios of data availability and completeness, including an experimental approach to obtain indirect indicators when standard forecasting methods are not applicable. The method of constructing a time series of an estimated indicator, which uses time series of indirect indicators, is described.

Literature review

The analysis of scientific articles has shown that for the tasks of selecting indirect indicators, an expert approach is used or a case is considered when the desired indicator is known and it is necessary to test the hypothesis of the desired indicator relationship with an indirect or a set of indirect indicators. Otherwise, they construct knowledge graphs that allow them to identify factors related to the desired indicator.

The article [3] provides an overview of forecasting methods which examines the quality of forecasting economic indicators by statistical methods and machine learning (ML) methods and also considers various estimates for the validation of the forecast. The methods under review are suitable for predicting the indicator in the presence of a statistically significant time series. In particular, the article examines a retrospective of prediction methods based on neural networks using the example of the M3-Competition [4]. The authors of the article in their research employed time series from the M3-Competition to compare methods according to sMAPE and concluded that statistical forecasting methods show better results compared to machine learning methods.

The article [5] compares the methods of combining forecasts with each other and with statistical forecasting methods. Variations of the Granger-Ramanathan method, the ridge regression combining forecasts, the method of harmonic weights, adaptive exponential smoothing with tracking signal, the method of simple exponential smoothing, and Box-Jenkins model (ARIMA) are considered. The comparison takes place on data for the period from 1957 to 2017:

– steel production,
– metallurgical coke production,
– cellulose production,
– plywood production,
– cement production.

In [6], the authors continue their studies of combining forecasts, adding for comparison the method of the matrix of paired preferences and the method of linear combination of partial indicators with different weights. The findings of the studies are similar, the Granger-Ramanathan method gives the greatest forecasting accuracy.

The article [8] proposes an approach to the development of composite mathematical models controlled by data. To verify the correctness and effectiveness of the proposed approach and substantiate the selected solutions, an experiment was conducted, the results of which show that the approach to model construction facilitates greater diversity and quality of the models obtained. The open source FEDOT framework for automatic modeling and machine learning (AutoML) was used to conduct the experiment.

In addition, articles devoted to methods of forecasting socio-economic indicators [9-14] and articles devoted to the construction of knowledge graphs for text analysis [15-17] were considered.

It is worth mentioning that there is a number of articles that use machine learning for text mining to identify the relationship of a calculated indicator with a set of indirect indicators. For example, article [9] describes an approach that employs text mining based on Chinese financial news on the Internet to predict a stock price trend based on a support vector machine (SVM). 2,302,692 news items were processed in the period 2008-2014. With the aid of the news corpus, a stop word dictionary and an accurate sentiment dictionary are formed. In reliance on the described corpus, an original forecasting model using SVM is proposed.

It is also interesting to note that there is a fairly large number of articles devoted to knowledge graphs. Thus, in the article [15], an approach is described for constructing thematic graphs of knowledge about world events on the basis of newspaper articles and it is shown that the entities extracted from such graphs improve the forecasts of industrial production in the USA, Germany and Japan. A corpus of over a billion news articles from 2015 to 2021 was used to validate the model.

Articles on the use of AutoML for forecasting and approaches to the modification of such systems for better prediction quality as well as optimization of calculations were also considered [18-21].

Materials and Methods

Problem statement

To test the hypothesis about the applicability of indirect indicators obtained using machine learning methods and statistical analysis of normative legal acts to forecast an arbitrary statistical indicator, we will clarify a number of key terms.

A statistical indicator is a generalized quantitative characteristic of qualitatively defined properties of a set of socio-economic objects or phenomena.

Senior official activity effectiveness indicator is a statistical indicator from the list of indicators for effectiveness analysis of senior official activities in the constituent entities of the Russian Federation according to the decree of the president of the Russian Federation from 04.02.2021 No.68 “On assessing the effectiveness of the activities of senior officials” [1].

By a **direct indicator** we will imply the senior official activity effectiveness indicator, the value of which is determined by a direct calculation in accordance with the methodology.

It is important to note that each of the 20 indicators defined in the above-mentioned government decree has a strictly defined name and a brief text description.

By **estimated indicator** we will imply the senior official activity effectiveness indicator the value of which cannot be calculated in accordance with the calculation methodology [2] due to the incompleteness of the time series of the initial data.

We will define the **indirect indicator** as an indicator that allows us to assess the dynamics of change for the estimated indicator in any other way other than described in the calculation methodology [2].

In the process of forecasting the senior official activity effectiveness indicator, machine learning, statistical or mathematical modeling methods are used, for which the completeness of the time series is an important characteristic. By **completeness of the time series**, we will imply the data quality indicator which determines the sufficiency of filling in indicators and their attributes to build a forecast for a given time series.

It is important to note that the use of indirect indicators is justified only if the time series for the corresponding senior official activity effectiveness indicator does not have the completeness necessary to obtain a forecast with a given accuracy.

To test our target hypothesis, the following tasks will need to be solved sequentially.

A. To develop a general methodology for obtaining a forecast for an arbitrary direct indicator and predictive assessment of the change dynamics using indirect indicators for an estimated indicator.

B. To develop a method for constructing a forecast estimate of the change dynamics in the estimated indicator based on aggregation from the forecast estimates of a set of indirect ones.

C. To carry out a comparative analysis of the forecast construction and the forecast assessment of the change dynamics for the senior official activity effectiveness indicator obtained using the developed methods.

General forecasting methodology

To solve the problem of constructing a general methodology for obtaining a forecast for an arbitrary direct indicator and predictive assessment of the change dynamics using indirect indicators for the calculated indicator, the following methodology is proposed.

1. For the performance indicator of senior officials, the availability and completeness of statistically significant time series for its calculation and forecasting is determined.

2. In case of incompleteness and/or unavailability of the time series described in step 1 (estimated indicator):

2.1. the expert selects a set of indirect indicators that have the time series necessary for forecasting;

2.2. for each indirect indicator, statistical data is collected in the form of a time series;

2.3. each time series of the indirect indicator is normalized;

2.4. for each time series, a forecast estimate of the dynamics of changes in the indirect indicator is formed;

2.5. the formed forecasts are aggregated into the final forecast estimate of the change dynamics in the estimated indicator.

3. In case of completeness and availability of the time series described in step 1, the forecasting process (direct indicator) is performed:

3.1. for a direct indicator, a time series is collected;

3.2. the resulting time series is forecasted.

The method of constructing a predictive assessment of the change dynamics in the estimated indicator

At the first stage, for each estimated indicator, the expert determines a set of indirect indicators related to the estimated indicator and having sufficient time series for forecasting. Having received a set of indirect indicators and their time series, it is necessary to forecast these series, and then proceed to the estimated indicator.

First, it is necessary to assess the quality of indirect indicator time series. In order to make a decision on the inclusion of this time series in the integral assessment, it must meet the following criteria:

A. Duration. The duration of the time series should be at least two years. This is necessary so that forecasting models can track seasonal trends.

B. Relevance. The time interval between the last point of the time series and the date before which the forecast is made should not be more than three years since the proposed solution is not intended for a long-term forecast.

C. The maximum measurement step. To take into account the seasonality factor, it is necessary that the time interval does not exceed a quarter, otherwise we can only talk about a rough assessment of the trend. This requirement is due to a strong loss of accuracy when interpolating values.

The selected time series should be normalized, and then forecasted by the most adequate method of forecasting time series.

The obtained forecasting results must be aggregated in such a way as to obtain a relative change in the estimated indicator. To do this, it is proposed to present the findings of forecasting indirect indicators in the form of a relative change in the last value. The averaged vector of forecast series will be a relative forecast of the estimated indicator.

Time series normalization method

For correct forecast aggregation of several indirect indicators related to a given estimated indicator, it is necessary to normalize the time series, namely, it is necessary to perform the following preprocessing stages:

1. The step of time series for indirect indicators is determined by the smallest timeframe between two consecutive points for any time series. The missing values in all time series are interpolated over the smallest timeframe.
2. Normalize the values of time series in the range from 0 to 1. Normalization is performed for each series independently, relative to the maximum value of the indicator on this time series.
3. If an inverse correlation is found for an indirect indicator, then it is necessary to invert the values of the time series.

For the time series prepared in this way, it is necessary to make a forecast until the given date. The forecasting is performed iteratively until the target date is reached. The choice of the best forecasting method should be carried out for each time series separately. In order to do that it is proposed to use AutoML models that automatically select the best forecasting models.

Time series forecasting and aggregation method

For forecasting it is proposed to use the AutoML open-source framework FEDOT [8]. The framework builds a composite model consisting of one or more models combined into a pipeline.

It is built from a variety of models/algorithms such as: ARIMA (autoregressive integrated moving average), AR (two-lag autoregression model), Linear (linear model), kNN regression (regression based on k-nearest neighbors), RANSAC (random sample consensus), etc. In total, about 30 models are used. An evolutionary algorithm is employed to select the most suitable in a particular situation, which performs optimization by means of genetic operators of selection, crossover and mutation.

After building a chain of applied models, the hyperparameters are tuned at all nodes of the resulting model using Bayesian optimization methods.

The resulting forecast of each indirect indicator must be presented as a relative change from the previous point. Since all time series were initially normalized relative to each other, the length of the forecast vector for all indirect indicators will be the same. This makes it possible to summarize the predictive vectors of indirect indicators and obtain the predictive vector of the estimated indicator.

Results and discussion

On the basis of the developed general methodology, an experiment was conducted to forecast the senior official activity effectiveness indicator "Population of the subject of the Russian Federation".

The expert identified the following indirect indicators:

1. life expectancy at birth, number of years;
2. number of births;
3. number of deaths;
4. migration increase or decrease;
5. number of doctors.

For the indicator "Life expectancy at birth" the expert identified the following indirect indicators:

1. number of doctors;
2. the number of people who have been assisted on an outpatient basis and on trips;

3. poverty rate;
4. number of deaths.

For each of the identified indicators, we search for statistical indicators and their time series. Each time series undergoes normalization.

After that, we form models to predict the time series of each indirect indicator and then use them to form forecasts. At the next step of the general methodology, we present each forecast as a relative change from the previous point and form a forecast vector. Afterwards, we calculate the arithmetic means of the indirect indicator forecast vectors, which will be the forecast vector for the desired estimated indicator.

Figure 1 and 2 shows the forecasts of the calculated indicators "Population of the subject of the Russian Federation" and "Life expectancy at birth" for the Khanty-Mansiysk Autonomous Okrug – Yugra from 2017 for 3 years ahead and forecasts were obtained in the form of relative average increases. Forecasting of these indicators was also carried out without the use of indirect indicators. The value of the forecast error according to the SMAPE metric (symmetric average absolute percentage error) for the indicator "Population of the subject of the Russian Federation" by calculating the direct indicator and the method of integral predictive evaluation of the indirect indicator aggregate was 3.019% and 2.3%, respectively, and for the indicator "Life expectancy at birth" 12.94% and 6.64%, respectively.

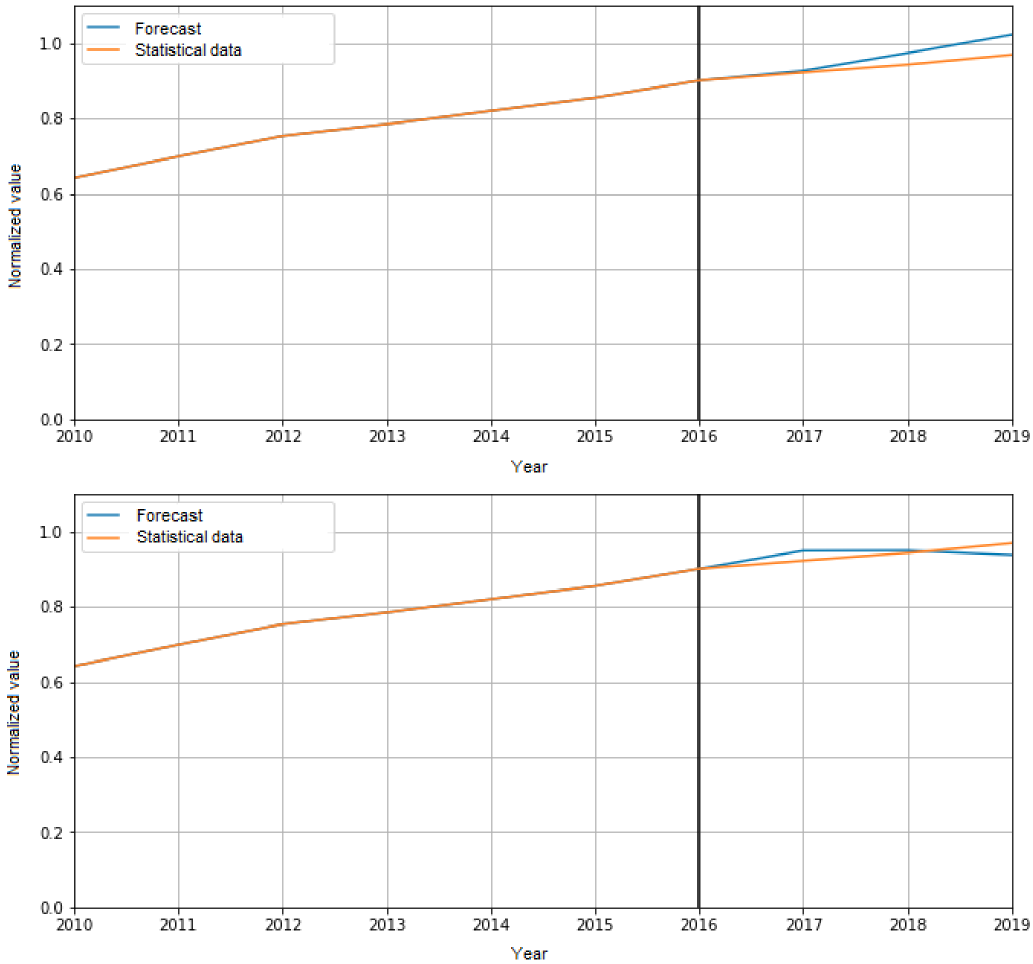

Figure 1 – Forecast of the indicator “Population of the subject of the Russian Federation»” by the ARIMA method and using indirect indicators
Рисунок 1 – Прогноз показателя «Численность населения субъекта Российской Федерации» методом ARIMA и с использованием косвенных показателей

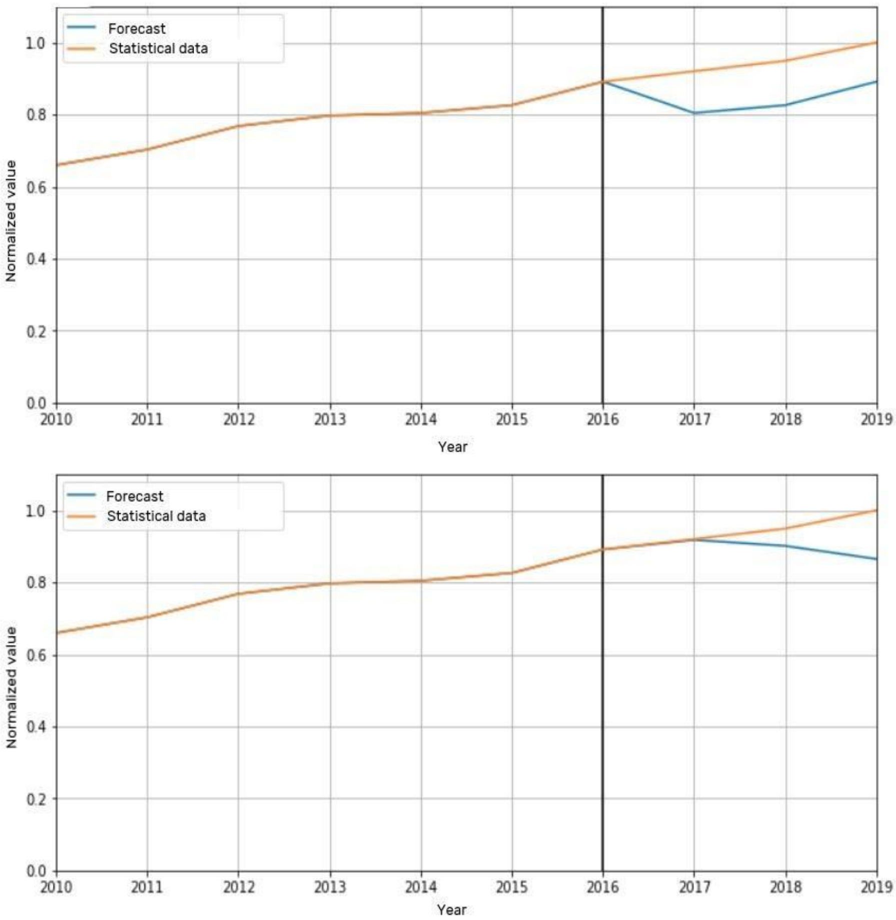

Figure 2 – Forecast of the indicator “Life expectancy at birth” by the ARIMA method and using indirect indicators.
Рисунок 2 – Прогноз показателя «Ожидаемая продолжительность жизни при рождении» методом ARIMA и с использованием косвенных показателей

Conclusion

Following on from the experiment described above, a comparative analysis of the forecast obtained for the direct indicator by classical forecasting methods and the forecast obtained by the proposed methodology for two indicators – “Population size” and “Life expectancy at birth” was carried out. The results of the experiment suggest that the initial hypothesis about the possibility of predicting key performance indicators of public authorities based on indirect indicators obtained using machine learning methods and statistical analysis of regulatory legal acts has been confirmed experimentally. The developed methodology for obtaining a forecast for an arbitrary indicator effectiveness of senior official activities, including the method of constructing an indirect indicator in the conditions of incompleteness of the time series and the method of constructing a forecast estimate of the change dynamics in the calculated indicator based on aggregation from the forecast estimates of a set of indirect ones, can be used for practical construction of specialized software for experts and analysts based on it.

At the same time, the great role of the expert in the interpretation of the obtained N-grams and the selection of candidates for indirect indicators should be noted. It is possible to reduce the subjectivity of an expert's assessment by creating a recommendation system based on the method of automatically determining the parameters of the n-gram ranking or using knowledge graphs.

REFERENCES

1. Decree of the President of the Russian Federation dated February 4, 2021 No. 68 «On estimating the effectiveness of the activities of senior officials (heads of the highest executive bodies of state government) of the constituent entities of the Russian Federation and the activities of executive authorities of the constituent entities of the Russian Federation». (In Russ.)
2. Resolution of the Government of the Russian Federation dated April 3, 2021 No. 542 «On approval of methods for calculating indicators for evaluating the effectiveness of the activities of senior officials (heads of the highest executive bodies of state government) of the constituent entities of the Russian Federation and the activities of executive authorities of the constituent entities of the Russian Federation». (In Russ.)
3. Makridakis S., Spiliotis E. and Assimakopoulos V., Statistical and machine learning forecasting methods: Concerns and ways forward». *PLoS ONE*. 2018;13(3):e0194889.
4. Makridakis S., Hibon M. The M3-Competition: results, conclusions and implications. *International jornal of forecasting*. 2000;16(4):451–476.
5. Frenkel A., Volkova N., Surkov A. and Romanyuk E. The application of ridge regression methods when combining forecasts. *Finance: Theory and Practice*. 2018:124–133. (In Russ.)
6. Frenkel A., Volkova N., Surkov A. and Romanyuk E. Comparative analysis of methods for constructing a combined forecast. *Voprosy statistiki*. 2017:17–27. (In Russ.)
7. Frenkel A. and Surkov A. Determination of weighting factors in combining forecasts. *Voprosy statistiki*. 2017:17–27. (In Russ.)
8. Polonskaia I., Nikitin N., Revin I., Vychuzhanin P. and Kalyuzhnaya A. Multi-objective evolutionary design of composite data-driven models. *2021 IEEE Congress on Evolutionary Computation (CEC)*. 2021:926–933.
9. Xie Y. and Jiang H. Stock market forecasting based on text mining technology: A support vector machine method. *Journal of Computers*. 2017;12(6):500–510.
10. Kuh S., Chiu G. and Westveld A. Modeling national latent socioeconomic health and examination of policy effects via causal inference. 2019. Available by: https://arxiv.org/abs/1911.00512.
11. Yagi I., Masuda Y. and Mizuta T. Analysis of the impact of high-frequency trading on artificial market liquidity. *IEEE Transactions on Computational Social Systems*. 2020;7:1324–1334.
12. He Q., Panp P. and Si Y. Multi-source transfer learning with ensemble for financial time series forecasting. *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. 2020;1:227–233.
13. Weeraddana D., Khoa N., O Neil L., Wang W. and Cai C. Energy Consumption Forecasting Using a Stacked Nonparametric Bayesian Approach. Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track. *ECML PKDD 2020. Lecture Notes in Computer Science*. 2021;12461.
14. Rajapaksha D., Bergmeir C. and Hyndman R.J. LoMEF: A Framework to Produce Local Explanations for Global Model Time Series Forecasts. 2021. Available by: https://arxiv.org/abs/2111.07001.
15. Tilly S. and Livan G. Macroeconomic forecasting with statistically validated knowledge graphs. 2021. Available by: https://arxiv.org/abs/2104.10457.

16. Huang J., Chang K., Xiong J. and Hwu W. Open relation modeling: Learning to define relations between entities. 2021. Available by: https://arxiv.org/abs/2108.09241v1.
17. Nimishakavi M., Saini U. and Talukdar P. Relation schema induction using tensor factorization with side information. *2016 Conference on Empirical Methods in Natural Language Processing*. 2016:414–423.
18. Feurer M., Klein A., Eggensperger K., Springenberg J.T., Blum M., and Hutter F. Auto-sklearn: efficient and robust automated machine learning. *Automated Machine Learning. Springer, Cham*. 2019:113–134.
19. Drori I., Krishnamurthy Y., Rampin R., Lourenco R., One J., Cho K., Silva C. and Freire J. AlphaD3M: Machine learning pipeline synthesis. *ICML AutoML workshop*. 2018.
20. Coors S., Schalk D., Bischl B. and Rügamer D. Automatic Componentwise Boosting: An Interpretable AutoML System. 2021. Available by: https://arxiv.org/abs/2109.05583.
21. Luo Z., He Z., Wang J., Dong M., Huang J., Chen M. and Zheng B. AutoSmart: An Efficient and Automatic Machine Learning framework for Temporal Relational Data. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2021:3976–3984.

СПИСОК ИСТОЧНИКОВ

1. Указ Президента РФ от 04.02.2021 N 68 «Об оценке эффективности деятельности высших должностных лиц (руководителей высших исполнительных органов государственной власти) субъектов Российской Федерации и деятельности органов исполнительной власти субъектов Российской Федерации».
2. Постановление Правительства РФ от 03.04.2021 N 542 «Об утверждении методик расчета показателей для оценки эффективности деятельности высших должностных лиц (руководителей высших исполнительных органов государственной власти) субъектов Российской Федерации и деятельности органов исполнительной власти субъектов Российской Федерации.
3. Makridakis S., Spiliotis E. and Assimakopoulos V., Statistical and machine learning forecasting methods: Concerns and ways forward». *PLoS ONE*. 2018;13(3):e0194889.
4. Makridakis S., Hibon M. The M3-Competition: results, conclusions and implications. *International jornal of forecasting*. 2000;16(4):451–476.
5. Френкель А., Волкова Н., Сурков А., Романюк Э. Использование методов гребневой регрессии при объединении прогнозов. *Финансы: теория и практика*. 2018;4:124–133.
6. Френкель А., Волкова Н., Сурков А., Романюк Э. Сравнительный анализ методов построения объединенного прогноза. *Вопросы статистики*. 2017;7:17–27.
7. Френкель А., Сурков А. Определение весовых коэффициентов при объединении прогнозов. *Вопросы статистики*. 2017;12:3–15.
8. Polonskaia I., Nikitin N., Revin I., Vychuzhanin P. and Kalyuzhnaya A. Multi-objective evolutionary design of composite data-driven models. *2021 IEEE Congress on Evolutionary Computation (CEC)*. 2021:926–933.
9. Xie Y. and Jiang H. Stock market forecasting based on text mining technology: A support vector machine method. *Journal of Computers*. 2017;12(6):500–510.
10. Kuh S., Chiu G. and Westveld A. Modeling national latent socioeconomic health and examination of policy effects via causal inference. 2019. Available by: https://arxiv.org/abs/1911.00512.
11. Yagi I., Masuda Y. and Mizuta T. Analysis of the impact of high-frequency trading on artificial market liquidity. *IEEE Transactions on Computational Social Systems*. 2020;7:1324–1334.

12. He Q., Panp P. and Si Y. Multi-source transfer learning with ensemble for financial time series forecasting. *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. 2020;1:227–233.
13. Weeraddana D., Khoa N., O Neil L., Wang W. and Cai C. Energy Consumption Forecasting Using a Stacked Nonparametric Bayesian Approach. Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track. *ECML PKDD 2020. Lecture Notes in Computer Science*. 2021;12461.
14. Rajapaksha D., Bergmeir C. and Hyndman R.J. LoMEF: A Framework to Produce Local Explanations for Global Model Time Series Forecasts. 2021. Available by: https://arxiv.org/abs/2111.07001.
15. Tilly S. and Livan G. Macroeconomic forecasting with statistically validated knowledge graphs. 2021. Available by: https://arxiv.org/abs/2104.10457.
16. Huang J., Chang K., Xiong J. and Hwu W. Open relation modeling: Learning to define relations between entities. 2021. Available by: https://arxiv.org/abs/2108.09241v1.
17. Nimishakavi M., Saini U. and Talukdar P. Relation schema induction using tensor factorization with side information. *2016 Conference on Empirical Methods in Natural Language Processing*. 2016:414–423.
18. Feurer M., Klein A., Eggensperger K., Springenberg J.T., Blum M., and Hutter F. Auto-sklearn: efficient and robust automated machine learning. *Automated Machine Learning. Springer, Cham*. 2019:113–134.
19. Drori I., Krishnamurthy Y., Rampin R., Lourenco R., One J., Cho K., Silva C. and Freire J. AlphaD3M: Machine learning pipeline synthesis. *ICML AutoML workshop*. 2018.
20. Coors S., Schalk D., Bischl B. and Rügamer D. Automatic Componentwise Boosting: An Interpretable AutoML System. 2021. Available by: https://arxiv.org/abs/2109.05583.
21. Luo Z., He Z., Wang J., Dong M., Huang J., Chen M. and Zheng B. AutoSmart: An Efficient and Automatic Machine Learning framework for Temporal Relational Data. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2021:3976–3984.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Аббазов Валерьян Ринатович, программист информационно-аналитического отдела Югорского научно-исследовательского института информационных технологий, Ханты-Мансийск, Российская Федерация.
e-mail: AbbazovVR@uriit.ru

Valeryan Rinatovich Abbazov, Programmer of Information Analysis Department of Yugorsky Research Institute of Information Technologies, Khanty-Mansiysk, Russian Federation.

Балуев Владимир Александрович, руководитель центра информационно-аналитических систем Югорского научно-исследовательского института информационных технологий, Ханты-Мансийск, Российская Федерация.
e-mail: BaluevVA@uriit.ru

Vladimir Aleksandrovich Baluev, Head of the Center for Information and Analytical Systems of Yugorsky Research Institute of Information Technologies, Khanty-Mansiysk, Russian Federation.

Бурлуцкий Владимир Владимирович, начальник отдела развития скоринга Группа страховых компаний «Югория», Ханты-Мансийск, Российская Федерация.
e-mail: Vladimir.Burlutskiy@gmail.com

Vladimir Vladimirovich Burlutsky, Head of Scoring Development Department of Yugoria Group of Insurance Companies, Khanty-Mansiysk, Russian Federation.

Мельников Андрей Витальевич, директор Югорского научно-исследовательского института информационных технологий, Ханты-Мансийск, Российская Федерация.
e-mail: MelnikovAV@uriit.ru
ORCID: 0000-0002-1073-7108

Andrej Vitalievich Melnikov, Director of Yugorsky Research Institute of Information Technologies, Khanty-Mansiysk, Russian Federation.

Русанов Михаил Александрович, старший преподаватель кафедры информационных технологий Югорского государственного университета, Ханты-Мансийск, Российская Федерация.
e-mail: RusanovMA@uriit.ru
ORCID: 0000-0002-9926-4609

Mikhail Aleksandrovich Rusanov, Lecturer of the Department of Digital Technologies of Yugra State University, Khanty-Mansiysk, Russian Federation.

Статья поступила в редакцию 08.06.2022; одобрена после рецензирования 06.07.2022; принята к публикации 18.07.2022.

The article was submitted 08.06.2022; approved after reviewing 06.07.2022; accepted for publication 18.07.2022.