

ГИБРИДНАЯ ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ПОТРЕБНОСТИ ЭКОНОМИКИ РОССИИ В КВАЛИФИЦИРОВАННЫХ КАДРАХ

А.В. Пролетарский

Д.В. Березкин

С.Т. Цаплин

pav@bmstu.ru

berezkind@bmstu.ru

tsaplin@bmstu.ru

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Рассмотрена гибридная интеллектуальная система прогнозирования потребности экономики России в квалифицированных кадрах с акцентом на анализ временных рядов вакансий. В условиях динамично изменяющегося рынка труда точные прогнозы позволяют работодателям и соискателям более эффективно планировать свои стратегии. Проведенные исследования охватывают период с июня 2022 года по ноябрь 2023 года и используют данные с ведущих сайтов вакансий. Применены классические статистические методы и современные технологии машинного обучения, включая Singular Spectrum Analysis, XGBoostRegressor и инновационный метод TimeLLM, основанный на методах Large Language Mode (LLM). Гибридный метод, сочетающий Singular Spectrum Analysis и XGBoostRegressor, а также метод TimeLLM, показал высокую точность прогнозирования, особенно для ИТ-специальностей и руководящих должностей. Оценка качества прогнозов выполнена с использованием метрик Mean Absolute Error и Mean Absolute Percentage Error. Перечислены преимущества гибридных подходов и LLM-методов по сравнению с классическими подходами и возможные ограничения, связанные с выходом за пределы допустимых значений. Отмечено, что гибридные методы и LLM-методы позволяют достигать высокой прогностической точности, но необходимо учитывать специфику данных и возможную коррекцию экстремальных значений. Даны рекомендации по использованию TimeLLM как гибкого инструмента, способного интегрировать экзогенные переменные и настраиваться под конкретные задачи прогнозирования

Ключевые слова

Временной ряд, SSA, LLM-метод, искусственный интеллект, открытые данные, гибридные алгоритмы, машинное обучение

Поступила 30.10.2025

Принята 21.05.2026

© Автор(ы), 2026

Введение. В современном мире в условиях динамично меняющихся экономики и рынка труда прогнозирование временных рядов становится ключевым инструментом при анализе и оценке спроса и предложения на рабочую силу. Точные прогнозы позволяют работодателям и соискателям эффективно планировать свои действия и стратегии в условиях нестабильности. В России, как и в других странах, факторы, влияющие на динамику вакансий, включают в себя экономическое положение, политику государства, демографические и социальные изменения, а также сезонные колебания, что делает прогнозирование временных рядов на рынке труда многокритериальным и важным аспектом анализа.

Существующие исследования в области прогнозирования временных рядов вакансий подтверждают необходимость учета многих факторов, в том числе контекста и характеристик рынка труда, при создании точных моделей [1].

Современные подходы к прогнозированию временных рядов основаны на применении глубокого и машинного обучения, а также других передовых статистических методов [2, 3], которые позволяют более точно учитывать многообразие сложных взаимосвязей и данных, что особенно важно для рынка труда, где данные могут быть нестационарными.

Проведена оценка применимости гибридных методов и моделей на основе LLM-методов (Large Language Mode) для прогнозирования временных рядов вакансий в России; проанализирована их эффективность на современном рынке труда.

Материалы и методы решения задач, принятые допущения. В качестве исходных данных использован датасет с портала «Университетский консорциум исследователей больших данных»¹. Датасет включает в себя информацию о вакансиях с основных сайтов вакансий (Headhunter, SuperJob, «Работа.ру», «Работа в России»), полученную в период с июня 2022 года по ноябрь 2023 года. В нем представлен широкий набор признаков, включая источники, уровень заработной платы, даты публикации, описания и различные детали о рабочем месте. Всего в датасете содержится 35 столбцов с информацией.

Для создания датасета, который будет использоваться для обучения моделей и последующего прогнозирования, выбраны два основных атрибута: дата публикации и название вакансии.

¹ Консорциум Big Data. URL: <https://opendata.university/access#rec814188930> (дата обращения: 25.09.2025).

Для решения поставленной задачи применены как традиционные статистические методы, так и методы машинного обучения и глубокие нейронные сети.

Цель настоящей работы — оценить применимость гибридных методов и LLM-методов в задаче прогнозирования потребности спроса на специалистов в России.

Метод SSA (Singular Spectrum Analysis) представляет собой мощный непараметрический инструмент для анализа и прогнозирования временных рядов [4]. Он объединяет элементы классического анализа временных рядов, мультивариативной статистики и обработки сигналов. Процесс применения SSA-метода включает в себя преобразование временного ряда в траекторную матрицу, ее разложение с использованием сингулярного разложения SVD (Singular Value Decomposition), а также группировку и реконструкцию компонентов [5]. Временной ряд встраивается в многомерное пространство и разлагается на составляющие с помощью разложения SVD [6]

$$X = U\Sigma V^T, \quad (1)$$

где X — траекторная матрица; U и V — ортогональные матрицы; Σ — диагональная матрица сингулярных чисел.

Эти компоненты могут быть сгруппированы по таким характеристикам, как тренды или периодические составляющие, и затем реконструированы для дальнейшего анализа или прогнозирования.

Алгоритм XGBoostRegressor, или Extreme Gradient Boosting, представляет собой эффективный инструмент машинного обучения, предназначенный для решения задач регрессии и прогнозирования временных последовательностей. Этот метод реализует принцип градиентного бустинга, объединяя прогнозы множества простых моделей (деревьев решений) для формирования более точной и надежной модели. Алгоритм XGBoostRegressor демонстрирует высокую результативность благодаря способности выявлять сложные нелинейные взаимосвязи в данных, а также устойчивости к переобучению за счет применения методов регуляризации [7].

Принципы функционирования XGBoostRegressor при прогнозировании временных рядов основаны на формировании признаков исторических данных, таких как лаговые переменные (значения предыдущих временных интервалов) и скользящие статистики (например, скользящее среднее) [8]. Эти признаки применяются для обучения модели, которая

в дальнейшем строит прогнозы будущих значений временного ряда. Приведем следующие ключевые формулы метода:

$$L(\Theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{i=1}^n \Omega(f_k) \quad (2)$$

— целевая функция градиентного спуска, $l(y_i, \hat{y}_i)$ — функция потерь; $\Omega(f_k)$ — регуляризация;

$$X_{lag_i}(t) = y(t-1) \quad (3)$$

— функция лаговых признаков.

Гибридный метод основан на комбинировании нескольких моделей в целях достижения более высокого качества. Для создания гибридных решений такого рода существует три основных подхода [9].

1. Бэггинг, используемый для уменьшения дисперсии модели и улучшения ее стабильности. При таком подходе создаются несколько подвыборок данных, затем каждая подвыборка используется для обучения отдельной модели. Результаты прогнозирования всех моделей усредняются [10].

2. Стэкинг, основанный на использовании нескольких различных моделей и объединении их предсказаний для улучшения общей точности. При таком подходе на исходных данных обучаются несколько базовых моделей. Затем результаты предсказаний этих моделей передаются в следующую модель в качестве метапараметров. Эта метамодель обучается на полученных предсказаниях и используется в качестве результирующей модели [11].

3. Бустинг, направленный на уменьшение смещения модели и улучшение ее точности. При данном подходе модели обучаются последовательно в целях уменьшения ошибки на каждом шаге. Каждая новая модель обучается только на тех данных, которые неверно предсказаны предыдущей моделью. Результат получается путем взвешенного суммирования предсказаний всех моделей [12].

Метод TimeLLM представляет собой инновационный подход к прогнозированию временных рядов, использующий возможности LLM-метода для обработки и анализа временных данных. Этот метод позволяет перепрограммировать LLM для выполнения задач прогнозирования, сохраняя при этом их основную архитектуру [13]. Метод TimeLLM включает в себя два ключевых компонента: перепрограммирование входных временных рядов в текстовые представления, которые более естественны

для LLM, и дополнение контекста ввода декларативными подсказками, такими как знания экспертов и инструкции по задачам для направления рассуждений LLM.

Основные принципы работы метода TimeLLM:

- преобразование временных данных в текстовый формат, который может быть понят LLM;
- использование этих данных для обучения модели прогнозирования.

Результаты. Для экспериментов выбрано две подгруппы вакансий. Это сделано для того, чтобы выполняемое прогнозирование имело ценность с экономической точки зрения, а также было наглядным для лиц, принимающих решение. Для первой подгруппы выделены вакансии ИТ-специальностей (включая программистов, разработчиков программного обеспечения, системных администраторов, аналитиков данных и специалистов по информационной безопасности). В данную подгруппу вошло 6812 записей. Для второй подгруппы выделены вакансии, связанные с руководящими должностями (руководители направлений, директора департаментов, начальники отделов, менеджеры проектов и исполнительные директора). В эту подгруппу вошло 19 763 записи. Для получения прогноза обучалось три модели: 1) SSA; 2) гибридная; 3) TimeLLM.

Для SSA-модели определены следующие параметры: размер окна 60; используются только первые три компонента; условие сходимости 0,001 [14].

Гибридная модель построена на основе стэкинга. В качестве базовой использована SSA-модель с параметрами, описанными ранее, а в качестве результирующей модели использовался алгоритм XGBoostRegressor с числом деревьев 1000. В модель также добавлены признаки, характеризующие конкретный день наблюдений: номера дней недели, месяца, года, номер квартала.

В качестве основы TimeLLM выбрана модель Qwen2.5-1.5B-Instruct. Данная модель адаптирована для работы с русским языком, имеет сравнительно малое число параметров, вследствие чего для нее требуется меньший объем вычислительных ресурсов. Модель обучалась на видеокарте NVIDIA RTX 4070 TI SUPER 16GB с размером батча 16.

В качестве метрик для оценки качества прогнозирования использованы MAE (Mean Absolute Error) и MAPE (Mean Absolute Percentage Error). Формулы для расчета ошибок [15]:

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|; \quad (4)$$

$$\text{MAPE} = \frac{1}{n} \sum_{j=1}^n \left| \frac{y_j - \hat{y}_j}{y_j} \right| \cdot 100 \%, \quad (5)$$

где n — число наблюдений; y_j — реальные значения; $\hat{y}_j$ — предсказанные значения.

Графики значений, предсказанных с помощью методов SSA, гибридного, TimeLLM, и реальных значений для вакансий ИТ-специальностей, приведены на рис. 1 (выбран временной период — июнь–сентябрь 2023 года). Получены следующие значения ошибок предсказаний: MAE = 62 и MAPE = 63 — SSA; MAE = 7,14 и MAPE = 67 — гибридный метод; MAE = 9 и MAPE = 0,9 — TimeLLM.

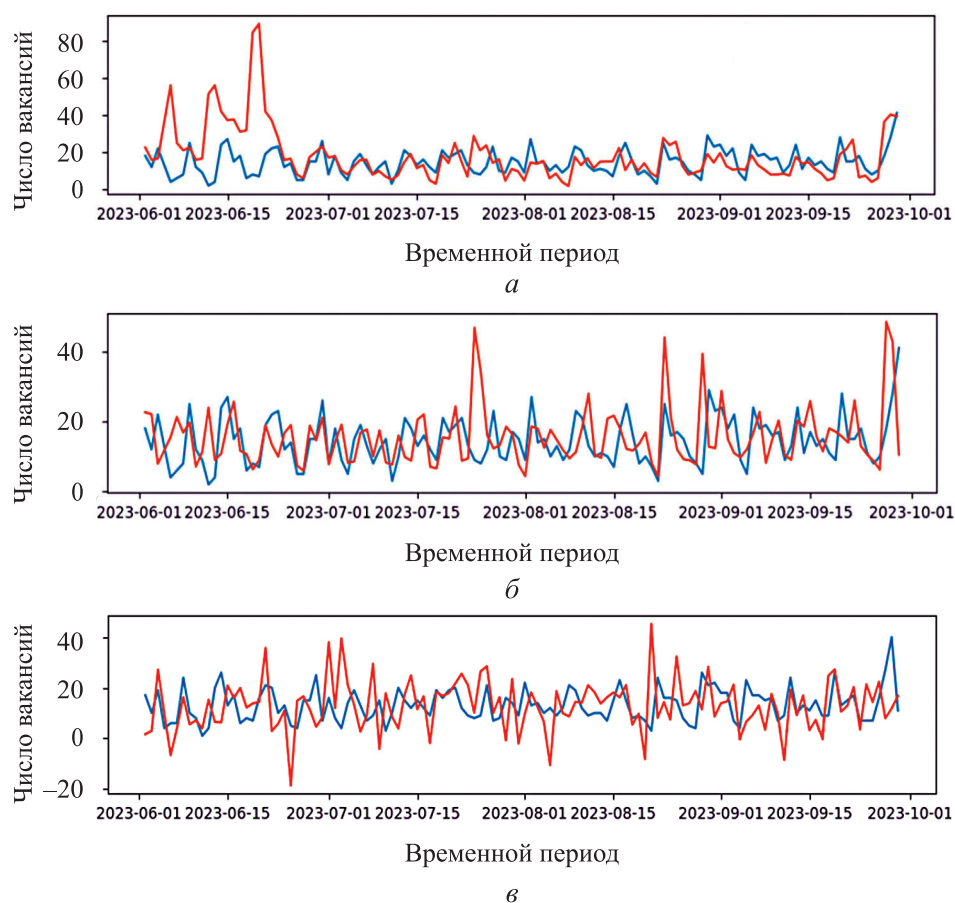


Рис. 1. Графики значений, предсказанных (—) с помощью SSA (а), гибридного метода (б) и TimeLLM (в), и реальных значений (—) для вакансий ИТ-специальностей

Графики значений, также предсказанных с помощью SSA, гибридного метода и TimeLLM, и реальных значений для вакансий руководящих должностей, приведены на рис. 2. Получены следующие значения ошибок предсказаний: MAE = 42 и MAPE = 575 — SSA; MAE = 19 и MAPE = 50 — гибридный метод; MAE = 28 и MAPE = 0,25 — TimeLLM.

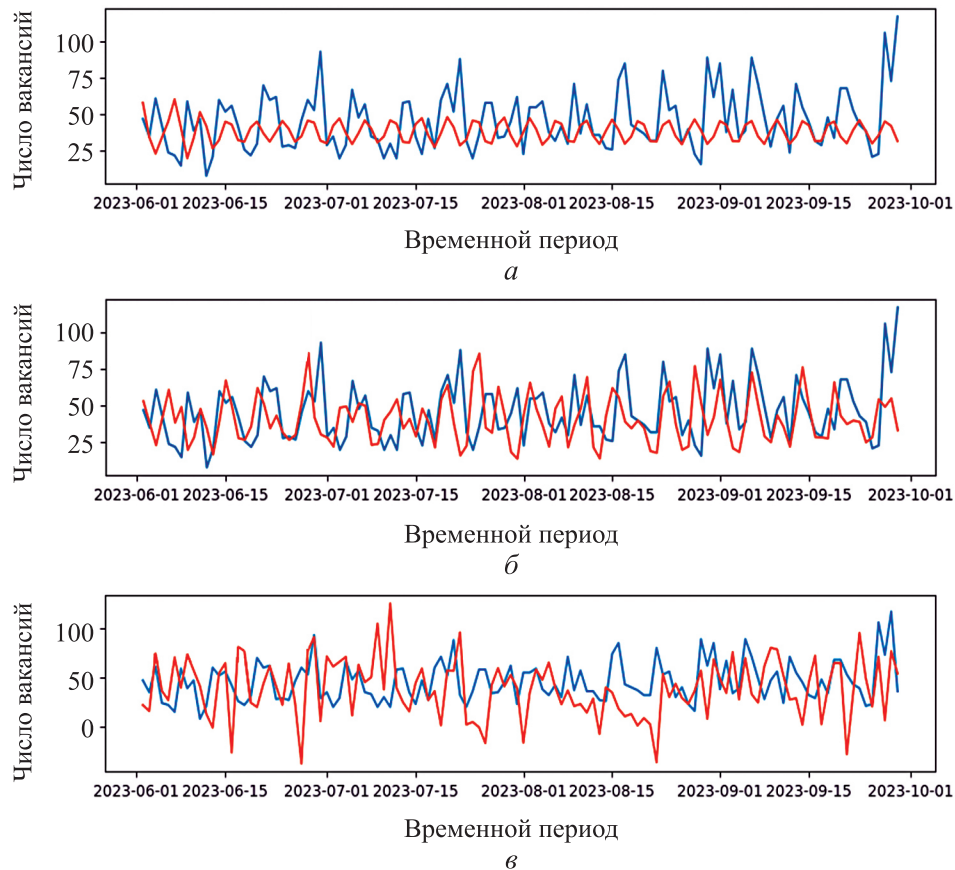


Рис. 2. Графики значений, предсказанных (—) с помощью SSA (а), гибридного метода (б) и TimeLLM (в), и реальных значений (—) для вакансий руководящих должностей

Обсуждение результатов. Согласно полученным результатам, гибридный метод и TimeLLM по качеству значительно лучше классического метода [16]. Однако языковая модель может давать отрицательные значения, что не входит в область допустимых значений для решаемой задачи [17]. Отметим, что в экспериментах использована языковая модель со сравнительно малым числом параметров, что может влиять на качество прогнозов [18]. Таким образом, для разработчика прогнозной системы встает выбор между скоростью ее работы и качеством прогноза. Тем не менее целе-

вой функцией для такого рода предсказаний является метрика MAPE, так как она лучше показывает ошибку в масштабах исходных данных. Основываясь на этой метрике, можно сделать вывод, что наилучшие результаты показал метод TimeLLM [19].

Заключение. Апробированы классические, гибридные методы, а также методы, основанные на применении больших языковых моделей. Полученные результаты свидетельствуют о высоком потенциале применения гибридных методов, на основе LLM на современном рынке труда России. Использование этих методов позволяет существенно повысить точность прогнозирования временных рядов, описывающих динамику вакансий, что указывает на высокую эффективность методов. Показано, что с помощью гибридных методов и методов на основе LLM можно добиться высокой прогностической точности, которая превосходит классические и нейросетевые методы, при этом методы, основанные на LLM, способны выходить за пределы допустимых значений. Данный недостаток можно нивелировать с помощью замены экстремальных значений значениями, полученными с помощью других методов. Кроме того, TimeLLM является гибким инструментом, который позволяет использовать для прогноза экзогенные переменные, а также настраивать входной промпт в зависимости от специфики входного временного ряда [20].

ЛИТЕРАТУРА

- [1] Makridakis S., Spiliotis E., Assimakopoulos V. The M4 Competition: 100,000 time series and 61 forecasting methods. *Int. J. Forecast.*, 2020, vol. 36, no. 1, pp. 54–71. DOI: <https://doi.org/10.1016/j.ijforecast.2019.04.014>
- [2] Hyndman R.J., Athanasopoulos G. *Forecasting: principles and practice*. Melbourne, OTexts, 2018.
- [3] Lim B., Arik S.Ö., Loeff N., et al. Time-series forecasting with deep learning: a survey. *Philos. Trans. R. Soc. A*, 2021, vol. 379, no. 2194, art. 20200209. DOI: <https://doi.org/10.1098/rsta.2020.0209>
- [4] Hamilton J.D. *Time series analysis*. Princeton, Princeton University Press, 1994.
- [5] Box G.E.P., Jenkins G.M., Reinsel G.C., et al. *Time series analysis*. Hoboken, John Wiley & Sons, 2015.
- [6] Shumway R.H., Stoffer D.S. *Time series analysis and its applications*. New York, Springer, 2017.
- [7] Diebold F.X. *Elements of forecasting*. Cincinnati, South-Western, 2001.
- [8] Harvey A.C. *Forecasting, structural time series models and the Kalman filter*. Cambridge, Cambridge University Press, 1990.

- [9] Cowpertwait P.S.P., Metcalfe A.V. Introductory time series with R. New York, Springer, 2009.
- [10] Bengio Y., Goodfellow I., Courville A. Deep learning. Cambridge, MIT Press, 2016.
- [11] Hochreiter S., Schmidhuber J. Long short-term memory. *Neural Comput.*, 1997, vol. 9, no. 8, pp. 1735–1780. DOI: <https://doi.org/10.1162/neco.1997.9.8.1735>
- [12] Vaswani A., Shazeer N., Parmar N., et al. Attention is all you need. *NIPS*, 2017, vol. 30. URL: <https://arxiv.org/pdf/1706.03762>
- [13] Kingma D.P., Ba J. Adam: a method for stochastic optimization. *arXiv:1412.6980*. DOI: <https://doi.org/10.48550/arXiv.1412.6980>
- [14] Cleveland R.B., Cleveland W.S., McRae J.E., et al. STL: a seasonal-trend decomposition procedure based on Loess. *Journal of Official Statistics*, 1990, vol. 6, no. 1, pp. 3–73.
- [15] Brownlee J. Machine learning mastery with Python. Vermont, Machine Learning Mastery, 2016.
- [16] Seni G., Elder J.F. Ensemble methods in data mining. Berlin, Springer, 2010.
- [17] Dietterich T.G. Ensemble methods in machine learning. In: *Multiple classifier systems*. Springer, 2000, pp. 1–15. DOI: https://doi.org/10.1007/3-540-45014-9_1
- [18] Zhou Z.H. Ensemble methods. Foundations and algorithms. London, CRC Press, 2012.
- [19] Jin M., Shiyu W., Ma L., et al. Time-LLM: time series forecasting by reprogramming large language models. *arXiv:2310.03506*. DOI: <https://doi.org/10.48550/arXiv.2310.01728>
- [20] Laptev N., Yosinski J., Lippi M., et al. Time series extreme event forecasting with neural networks at uber. In: *International Conference on Machine Learning*, 2017, vol. 34, pp. 1–5.

Пролетарский Андрей Викторович — д-р техн. наук, декан факультета «Информатика и системы управления» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Березкин Дмитрий Валерьевич — канд. техн. наук, доцент кафедры «Компьютерные системы и сети» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Цаплин Сергей Тимофеевич — ассистент кафедры «Компьютерные системы и сети» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Просьба ссылаться на эту статью следующим образом:

Пролетарский А.В., Березкин Д.В., Цаплин С.Т. Гибридная интеллектуальная система прогнозирования потребности экономики России в квалифицированных кадрах. *Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение*, 2026, № 2 (155), с. 80–91. EDN: CUNSTQ

HYBRID INTELLIGENT SYSTEM FOR FORECASTING THE RUSSIAN ECONOMY'S NEED FOR QUALIFIED PERSONNEL

A.V. Proletarsky

D.V. Berezkin

S.T. Tsaplin

pav@bmstu.ru

berezkind@bmstu.ru

tsaplin@bmstu.ru

BMSTU, Moscow, Russian Federation

Abstract

The article discusses a hybrid intelligent system for forecasting the Russian economy's demand for skilled personnel, with an emphasis on the analysis of time series of vacancies. In a dynamically changing labor market, accurate forecasts allow employers and job seekers to plan their strategies more effectively. The study covers the period from June 2022 to November 2023 and uses data from leading job sites. The work uses both classical statistical methods and modern machine learning technologies, including Singular Spectrum Analysis (SSA), XGBoostRegressor, and an innovative TimeLLM approach based on Large Language Models (LLM). The hybrid approach combining SSA and XGBoostRegressor, as well as the TimeLLM method, showed high forecasting accuracy, especially for IT specialties and executive positions. The quality of forecasts was assessed using the Mean Absolute Error and Mean Absolute Percentage Error metrics. The results demonstrate the advantages of hybrid and LLM methods over classical approaches, although they point to possible limitations associated with going beyond the permissible values. It is noted that hybrid approaches and LLM-based methods allow achieving high predictive accuracy, but require taking into account the specificity of the data and possible correction of extreme values. It is recommended to use TimeLLM as a flexible tool capable of integrating exogenous variables and adjusting to specific forecasting tasks

Keywords

Time series, SSA, LLM, artificial intelligence, open data, hybrid algorithms, machine learning

Received 30.10.2025

Accepted 21.05.2026

© Author(s), 2026

REFERENCES

- [1] Makridakis S., Spiliotis E., Assimakopoulos V. The M4 Competition: 100,000 time series and 61 forecasting methods. *Int. J. Forecast.*, 2020, vol. 36, no. 1, pp. 54–71. DOI: <https://doi.org/10.1016/j.ijforecast.2019.04.014>

- [2] Hyndman R.J., Athanasopoulos G. Forecasting: principles and practice. Melbourne, OTexts, 2018.
- [3] Lim B., Arik S.Ö., Loeff N., et al. Time-series forecasting with deep learning: a survey. *Philos. Trans. R. Soc. A*, 2021, vol. 379, no. 2194, art. 20200209. DOI: <https://doi.org/10.1098/rsta.2020.0209>
- [4] Hamilton J.D. Time series analysis. Princeton, Princeton University Press, 1994.
- [5] Box G.E.P., Jenkins G.M., Reinsel G.C., et al. Time series analysis. Hoboken, John Wiley & Sons, 2015.
- [6] Shumway R.H., Stoffer D.S. Time series analysis and its applications. New York, Springer, 2017.
- [7] Diebold F.X. Elements of forecasting. Cincinnati, South-Western, 2001.
- [8] Harvey A.C. Forecasting, structural time series models and the Kalman filter. Cambridge, Cambridge University Press, 1990.
- [9] Cowpertwait P.S.P., Metcalfe A.V. Introductory time series with R. New York, Springer, 2009.
- [10] Bengio Y., Goodfellow I., Courville A. Deep learning. Cambridge, MIT Press, 2016.
- [11] Hochreiter S., Schmidhuber J. Long short-term memory. *Neural Comput.*, 1997, vol. 9, no. 8, pp. 1735–1780. DOI: <https://doi.org/10.1162/neco.1997.9.8.1735>
- [12] Vaswani A., Shazeer N., Parmar N., et al. Attention is all you need. *NIPS*, 2017, vol. 30. Available at: <https://arxiv.org/pdf/1706.03762>
- [13] Kingma D.P., Ba J. Adam: a method for stochastic optimization. *arXiv:1412.6980*. DOI: <https://doi.org/10.48550/arXiv.1412.6980>
- [14] Cleveland R.B., Cleveland W.S., McRae J.E., et al. STL: a seasonal-trend decomposition procedure based on Loess. *Journal of Official Statistics*, 1990, vol. 6, no. 1, pp. 3–73.
- [15] Brownlee J. Machine learning mastery with Python. Vermont, Machine Learning Mastery, 2016.
- [16] Seni G., Elder J.F. Ensemble methods in data mining. Berlin, Springer, 2010.
- [17] Dietterich T.G. Ensemble methods in machine learning. In: *Multiple classifier systems*. Springer, 2000, pp. 1–15. DOI: https://doi.org/10.1007/3-540-45014-9_1
- [18] Zhou Z.H. Ensemble methods. Foundations and algorithms. London, CRC Press, 2012.
- [19] Jin M., Shiyu W., Ma L., et al. Time-LLM: time series forecasting by reprogramming large language models. *arXiv:2310.03506*. DOI: <https://doi.org/10.48550/arXiv.2310.01728>
- [20] Laptev N., Yosinski J., Lippi M., et al. Time series extreme event forecasting with neural networks at uber. In: *International Conference on Machine Learning*, 2017, vol. 34, pp. 1–5.

Proletarsky A.V. — Dr. Sc. (Eng.), Dean of the Faculty of Computer Science and Control Systems, BMSTU (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Berezkin D.V. — Cand. Sc. (Eng.), Assoc. Professor, Department of Computer Systems and Networks, BMSTU (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Tsaplin S.T. — Assistant, Department of Computer Systems and Networks, BMSTU (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Please cite this article in English as:

Proletarsky A.V., Berezkin D.V., Tsaplin S.T. Hybrid intelligent system for forecasting the Russian economy's need for qualified personnel. *Herald of the Bauman Moscow State Technical University, Series Instrument Engineering*, 2026, no. 2 (155), pp. 80–91 (in Russ.). EDN: CUNSTQ