

МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ НА ПРИМЕРЕ ВАЛЮТНЫХ КОТИРОВОК

TIME SERIES FORECASTING MODEL BASED ON THE EXAMPLE OF CURRENCY QUOTATIONS

**M. Khabarov
V. Lovtsov
K. Tiazhelnikova
V. Gurenko**

Summary. The article describes the actual task of developing a model of time series forecasting on the example of quotes of currency pairs. The peculiarities of working with time series are defined and the mathematical model of the subject area of the developed model is formulated. The process of dividing a time series into components is analyzed. The main approaches to the realization of time series forecasting model are studied, namely: statistical algorithms and machine learning. Necessary and sufficient conditions of stationarity of time series for realization of the model of statistical approach are considered. The initial data set for the development of time series forecasting model on the example of currency quotes based on the dataset «Currency rates: archival and current data on the value of foreign currencies against the ruble» was formed. A set of lag and aggregated informative features for the realization of models based on machine learning was formed. Practical experiments were conducted to analyze the behavior of the models in different situations of financial market stability. The results of each algorithm within the conducted experiments are presented, including MSE, MAE, MAPE and R2 metrics. Based on the results obtained, the optimal algorithm was selected to realize the task at hand. Also, the most informative features of this model for forecasting time series of currency quotes were determined.

Keywords: time series, machine learning, time series forecasting, forecasting approaches, quotes, currency pairs.

Хабаров Марк Эдуардович

Московский государственный технический
университет имени Н.Э. Баумана, г. Москва
mark.khabarov@yandex.ru

Ловцов Владимир Алексеевич

Московский государственный технический
университет имени Н.Э. Баумана, г. Москва
lovets1818@gmail.com

Тяжельникова Ксения Сергеевна

руководитель проектов, АО «ТБанк», г. Москва
kstyazhelnikova@gmail.ru

Гуренко Владимир Викторович

доцент, Московский государственный технический
университет имени Н.Э. Баумана, г. Москва
wgurenko@bmstu.ru

Аннотация. В статье описывается актуальная задача разработки модели прогнозирования временных рядов на примере котировок валютных пар. Определены особенности работы с временными рядами и сформулирована математическая модель предметной области разработанной модели. Проанализирован процесс разделения временного ряда на компоненты. Изучены основные подходы к реализации модели прогнозирования временных рядов, а именно: статистические алгоритмы и машинное обучение. Рассмотрены необходимые и достаточные условия стационарности временного ряда для реализации модели статистического подхода. Сформирован исходный набор данных для разработки модели прогнозирования временных рядов на примере валютных котировок, основанный на датасете «Валютные курсы: архивные и текущие данные о стоимости иностранных валют по отношению к рублю». Был сформирован набор лаговых и агрегированных информативных признаков для реализации моделей на основе машинного обучения. Проведены практические эксперименты с целью анализа поведения моделей в различных ситуациях стабильности финансового рынка. Представлены результаты работы каждого алгоритма в рамках проведенных экспериментов, включающие в себя метрики MSE, MAE, MAPE и R2. На основе полученных результатов для реализации поставленной задачи был выбран оптимальный алгоритм. Также были определены наиболее информативные признаки этой модели для прогнозирования временных рядов валютных котировок.

Ключевые слова: временные ряды, машинное обучение, прогнозирование временных рядов, подходы к прогнозированию, котировки, валютные пары.

Введение

В настоящее время разработка моделей прогнозирования временных рядов является одной из ключевых задач в различных областях науки и практики, включая климатологию, здравоохранение, экономику, финансы и т.д. В условиях быстро меняющегося мира, где

информация и данные становятся основными активами, способность предсказывать будущие значения на основе исторических данных приобретает особую актуальность.

Так в области экономики и финансов одной из наиболее важных задач, актуальность которой невозмож-

но переоценить, является прогнозирование валютных котировок. Эта задача стала одной из ключевых для инвесторов, трейдеров и экономистов по всему миру, ведь курсы валют не только отражают состояние экономики отдельных стран, но и влияют на международную торговлю, инвестиции и финансовую стабильность [1]. Таким образом эффективное прогнозирование изменения валютных котировок может позволить своевременно реагировать на изменения валютного рынка, являющегося одним из самых ликвидных и динамичных сегментов финансовой системы.

Модель анализа временных рядов представляет собой мощный инструмент для прогнозирования исторических данных. В случае динамики валютных котировок она позволяет учитывать исторические данные целевой валютной пары, других валютных пар, сезонные колебания и прочие факторы, оказывающие влияние на курс валюты. Используя подобную модель, становится возможным качественно оценивать финансовый рынок, а также создавать надёжные и прибыльные инвестиционные стратегии, минимизируя потенциальные убытки.

С целью создания структурированного подхода к решению задачи разработки модели прогнозирования временных рядов на примере валютных котировок необходимо сформулировать математическую модель предметной области. Определим временной ряд, как последовательность значений, отражающих динамику процесса по времени и измеренных в последовательные моменты времени через равные интервалы. Пусть последовательность действительных значений $y_0, y_1, \dots, y_t$ — это временной ряд. Тогда задача построения модели прогнозирования временных рядов на примере валютных котировок будет состоять в построении функции $F_{t,d}(y_0, y_1, \dots, y_t; w)$, то есть получении прогноза на момент времени $t + d$, где d — это горизонт прогнозирования, а w — вектор весов модели.

При этом для анализа временного ряда необходимо определение его компонент [3,5]:

- тренд (T) — плавное долгосрочное изменение уровня ряда;
- сезонность (S) — циклическое изменение уровня ряда с постоянным периодом;
- ошибка/шум (I) — непрогнозируемая случайная компонента ряда.

Иногда ещё выделяют циклическую компоненту (C) — изменение уровня ряда с переменным периодом. Если сезонность отражает регулярные, повторяющиеся с постоянной частотой, колебания внутри определённого периода, то циклическая компонента может иметь переменную длину и амплитуду и повторяться с нефиксированной периодичностью, например, как экономические циклы (бумы и спады).

Сам временной ряд можно разложить либо в сумму, либо в произведение своих компонент — это называется аддитивной, или, во втором случае мультипликативной моделью ряда [3, 4, 5]. В данном случае было принято решение работать с аддитивной моделью.

Материалы и методы решения задач, принятые допущения

Существует два основных подхода к реализации модели прогнозирования временных рядов [6-9]:

- статистический подход (подход на основе статистики и проверки гипотез): ARIMA, SARIMA, ARCH, GARCH и т.п.;
- машинное обучение (подход с использованием моделей, основанных на математических методах): регрессия, случайный лес, бустинг и т.п.

Принято решение исследовать экспериментально следующие алгоритмы прогнозирования временных рядов:

- модель сезонного авторегрессивного интегрированного скользящего среднего (Seasonal Autoregressive Integrated Moving Average, SARIMA),
- линейная регрессия (Linear Regression, LR),
- XGBoost (классический алгоритм градиентного улучшения, обладающий основными необходимыми свойствами);
- CatBoost (алгоритм градиентного улучшения, разработанный компанией Yandex, который обладает повышенной устойчивостью к переобучению, а также простотой использования и настройки параметров);
- LightGBM (алгоритм градиентного улучшения, ориентированный на скорость и эффективность работы с большими данными за счёт алгоритма построения деревьев Histogram-based).

Для исследования алгоритма SARIMA необходимо приведение временного ряда к стационарному виду, то есть должны выполняться следующие условия [6]:

- постоянное среднее значение (Среднее значение ряда должно быть стабильным и независимым от времени. Это означает, что в разные временные моменты не должно наблюдаться четкой тенденции к увеличению или уменьшению);
- постоянная вариация (Дисперсия ряда должна быть постоянной и не зависеть от времени. Это означает, что в разные моменты времени амплитуда колебаний ряда не должна изменяться);
- некоррелированность (Корреляция между значениями ряда на различных временных интервалах должна быть минимальной или полностью отсутствовать. Это указывает на отсутствие устойчивых зависимостей или повторяющихся паттернов во времени.).

Для проверки стационарности временного ряда было принято решение использовать статистический тест «критерий Дики-Фуллера», цель которого состоит в проверке наличия единичного корня во временном ряде.

Результаты

В результате анализа выбранных алгоритмов было установлено:

- SARIMA обладает простотой реализации и является самым распространённым подходом с наибольшей теоретической базой, однако имеет достаточно низкие точности в ситуациях высокой волатильности временного ряда, а также в отличие от других методов требует понимания статистического алгоритма проверки гипотез;
- алгоритм линейной регрессии обладает низкой склонностью к переобучению, а также лучшую скорость обучения, однако выбросы в данных сильно ухудшают работу, а также поддерживаются только линейные решения;
- алгоритмы градиентного улучшения имеют тенденцию к повышенной точности, так как основаны на концепции, когда большое количество деревьев обучается с учётом ошибок предыдущих, однако возможно существенное переобучение.

В данной работе был осуществлён сбор и формирование исходного набора данных для построения модели прогнозирования временных рядов на примере валютных котировок на основе датасета «Валютные курсы: архивные и текущие данные о стоимости иностранных валют по отношению к рублю», который был подготовлен проектом «Инфраструктура научно-исследовательских данных» на основе данных Центрального банка Российской Федерации [7]. Рассматриваемый датасет содержит информацию по котировкам мировых валют к рублю. В рамках проводимого исследования было принято решение разработать модель прогнозирования временного ряда валютной пары GBP/RUB (Британский фунт стерлингов/Российский рубль) на основе исторических данных этой котировки, а также валютных пар USD/RUB (Доллар США/Российский рубль) и EUR/RUB (Евро/Российский рубль).

При решении поставленной задачи прогнозирования для настройки алгоритмов машинного обучения будем использовать начальную коллекцию данных, которую разделим на два непересекающихся подмножества [8–9]:

- тренировочная (train) коллекция. Выборка, с помощью которой настраивается алгоритм;
- тестовая (test) коллекция. Выборка, с помощью которой тестируется эффективность настроенного алгоритма.

Для подбора гиперпараметров модели машинного обучения используется кросс-валидация на коллекции train, позволяющая снизить вероятность переобучения. Из-за необходимости строго контролировать соблюдение временной хронологии в данных была применена неслучайная кросс-валидация.

Существует два основных алгоритма неслучайной кросс-валидации временных рядов, в которых тестовый блок (Test) всегда после тренировочного (Train) [6, 7, 9]:

- увеличение train части и сдвиг test части в будущее;
- сдвиг train и test блоков синхронно вправо, не расширяя train (меньшее опирание алгоритма на старые данные).

Для реализации поставленной задачи был выбран первый тип кросс-валидации. Также, для применения классических моделей машинного обучения, был сформирован следующий набор информативных признаков [6, 8, 11]:

- лаги (Lags), то есть значения временного ряда в предыдущие моменты времени, которые при прогнозировании значения на завтра позволяют опираться на значения за прошлые дни. Кроме того, в данном случае в качестве лаговых признаков использовались сдвиги временных рядов котировок USD/RUB и EUR/RUB;
- агрегированные признаки по имеющемуся ряду. Примером таких признаков может быть среднее значение в месяце, медианные значения, дисперсия и т.д.

С целью анализа поведения моделей в различных ситуациях стабильности финансового рынка, в процессе решения поставленной задачи было принято решение реализовать модели для двух ситуаций:

- train (01.01.1992 — 28.11.2019) и test (29.11.2019 — 29.12.2019);
- train (01.01.1992 — 28.09.2021) и test (29.09.2021 — 29.10.2021).

В табл. 1 показаны метрики (MSE, MAE, MAPE, R2), полученные на test коллекциях для каждого алгоритма, настроенного на train выборке с помощью кросс-валидации в двух экспериментах.

Исходя из данных, представленных в табл. 1, можно сделать вывод, что в результате двух практических экспериментов предпочтительной оказалась модель на основе алгоритма CatBoost, которая показала наилучшие метрики MSE, MAE, MAPE и R2 на выборках test.

На рис. 1 представлена работа моделей на тестовой коллекции первого практического эксперимента.

Таблица 1.
Результаты практических экспериментов

Номер эксперимента	Алгоритм	Метрики			
		MSE	MAE	MAPE	R2
1	LinReg	2.543900	1.254007	0.015146	0.626767
	SARIMA	2.125751	1.216773	0.014751	0.687397
	CatBoost	<u>1.869739</u>	<u>1.000337</u>	<u>0.012041</u>	<u>0.723202</u>
	LightGBM	1.949264	1.085842	0.013130	0.711849
	XGBoost	1.953474	1.111856	0.013503	0.711064
2	LinReg	2.078580	1.164289	0.011467	0.558066
	SARIMA	2.110416	1.232422	0.012171	0.544957
	CatBoost	<u>1.739513</u>	<u>1.008278</u>	<u>0.009911</u>	<u>0.627225</u>
	LightGBM	2.023863	1.119409	0.010987	0.566179
	XGBoost	1.870440	1.074756	0.010564	0.597143

На рис. 2 представлена работа моделей на тестовой коллекции первого практического эксперимента.

В результате проведённых практических экспериментов можно заметить, что наилучшие результаты в различных ситуациях финансового рынка показала

модель прогнозирования временных рядов валютных котировок на основе алгоритма CatBoost.

Также была произведена оценка информативности признаков посредством анализа их весов для модели на базе выбранного наилучшего алгоритма. Самые значимые полученные веса для признаков представлены в табл. 2.

Исходя из данных, представленных в табл. 2, можно сделать вывод, что пятью наиболее информативными признаками для модели являются: сдвиг временного ряда на одно значение, три значения, среднее значение временного ряда по месяцу в рамках годов, а также сдвиг ряда USD на два и сдвиг целевого ряда на четыре значения.

Обсуждение полученных результатов

Необходимо понимать, что для получения более точной модели прогнозирования временных рядов на примере валютных котировок, возможно использование новостных данных, которые смогут учитывать влияние различного новостного фона, формирующего ситуативные факторы. Также, в качестве экзогенного фактора, возможно, например, использование нефтяных котировок. Возможность учёта описанных данных ляжет в основу дальнейших исследований.

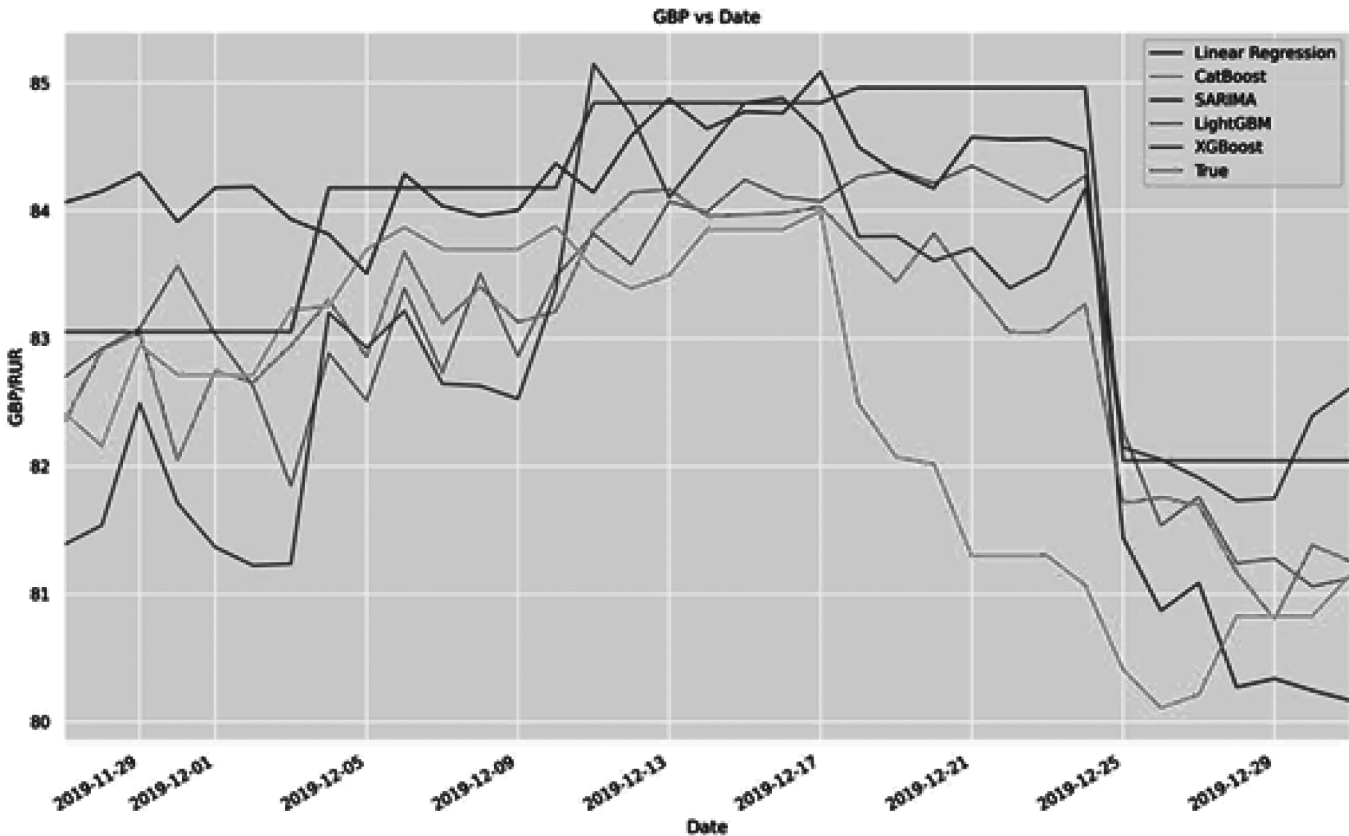


Рис. 1. Пример работы моделей в первом практическом эксперименте

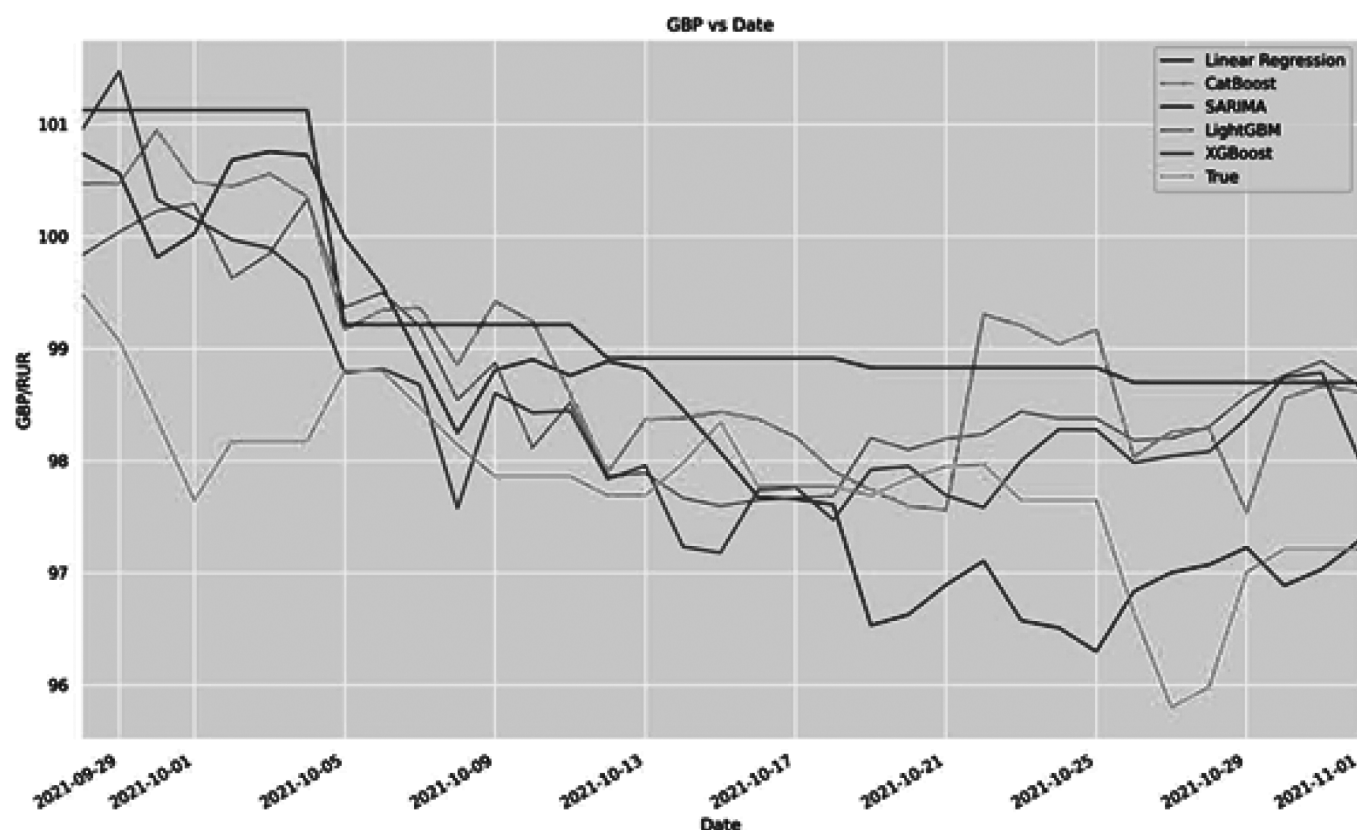


Рис. 2. Пример работы моделей во втором практическом эксперименте

Таблица 2.
Полученные веса признаков для наилучшей модели

Признак	Вес признака в модели
Month average	-0.0153
Lag 3 USD	-0.0026
Lag 5	0.0168
Year average	0.0195
Lag 2 EUR	0.0209
Lag 6	0.0217
Lag 1 EUR	0.0441
Lag 4	0.0579
Lag 2 USD	0.6371
Month_year average	0.7860
Lag 3	0.8713
Lag 1	0.9670

Заключение

В рамках данной работы была решена задача разработки модели прогнозирования временных рядов на примере котировок валютных пар. Был проведен сбор

и подготовка данных, связанных с котировками валютной пары GBP/RUB. Произведён анализ алгоритмов прогнозирования временных рядов котировок на финансовом рынке (SARIMA, Linear Regression, XGBoost, CatBoost, LightGBM) и на их основе осуществлена настройка моделей с помощью неслучайной кросс-валидации, произведено сравнение эффективности и качества их работы с помощью набора метрик (MSE, MAE, MAPE, R2) в рамках двух практических экспериментов в различных состояниях финансового рынка. В результате экспериментов установлено, что наиболее предпочтительным вариантом является модель прогнозирования временных рядов валютных котировок на основе алгоритма CatBoost. Определены дальнейшие направления развития исследования и разработки. Также на основе весов признаков выбранной модели был сделан вывод о наиболее информативных признаках в решении поставленной задачи. Ими оказались (в порядке убывания информативности):

- сдвиг временного ряда на одно значение;
- сдвиг временного ряда на три значения;
- среднее значение временного ряда по месяцу в рамках годов;
- сдвиг временного ряда USD/RUB на два значения;
- сдвиг временного ряда на четыре значения.

ЛИТЕРАТУРА

1. Time Series Econometrics: Basic Terms of Time Series Analysis and Forecasting / S.V. Kurovsky, D.A. Mishin, D.S. Lezhnin [et al.] // Innovation & Investment. — 2024. — No. 8. — P. 497–502. — EDN DFEKOL.
2. Shishkov E.M., Pronichev A.V. & Savelyev A.A. (2022). Forecasting time series via machine learning methods based on the power output schedule of an electric power station. International Research Journal, 2(116). <https://doi.org/10.23670/IRJ.2022.116.2.009>.
3. Наумов В.Н. Прогнозирование валютных курсов на основе сравнительного анализа динамики временных рядов / В.Н. Наумов, С.А. Буров // Управленческое консультирование. — 2015. — № 2(74). — С. 55–64. — EDN TLQRWT.
4. Лебедев Я.Б. Прогнозирование курсов валют методом ARIMA / Я.Б. Лебедев // Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем: материалы Всероссийской конференции с международным участием, Москва, 17–21 апреля 2023 года. — Москва: Российский университет дружбы народов (РУДН), 2023. — С. 358–361. — EDN MHJWFW.
5. Garg R., Barpanda S., "Machine Learning Algorithms for Time Series Analysis and Forecasting", Art. no. arXiv:2211.14387, 2022. doi:10.48550/arXiv.2211.14387.
6. Ivanyuk V., Tsvirkun A. Time series analysis and financial asset forecasting methods // Proceedings of 2021 14th International Conference Management of Large-Scale System Development, MLSD 2021. — 2021. — Т. -, №. -. — С. -. doi: 10.1109/MLSD52249.2021.9600198
7. Degtiarev Konstantin. (2005). Прогнозирование валютных котировок с использованием модифицированного стационарного метода, основанного на нечетких временных рядах. 10.13140/RG.2.2.21406.84803.
8. Ворокова Н.Х. Анализ временных рядов в прогнозировании (на примере валютного курса доллара к рублю) / Н.Х. Ворокова, Ж.И. Аблаева, К.И. Куник // Вестник Алтайской академии экономики и права. — 2024. — № 7–2. — С. 266–272. — DOI 10.17513/vaael.3591. — EDN RWVVKT.
9. Ярушкина Н.Г. Моделирование трафика терминал-сервера на основе анализа нечетких тенденций временных рядов / Н.Г. Ярушкина, Т.Р. Юнусов, Т.В. Афанасьева // Программные продукты и системы. — 2007. — № 4. — С. 38. — EDN JVSSFD.
10. A. Minitaeva. «Multimodel Approach to Forecasting Nonlinear Nonstationary Processes in Optimal Control Problems.» *Proceedings of the 2024 6th International Youth Conference on Radio Electronics, Electrical and Power Engineering, REEPE 2024*, 2024, doi: 10.1109/REEPE60449.2024.10479796
11. V. Kukartsev, A. Borodulin, S. Samolotskii. «Use of Analytical Methods in Variant Forecasting Tasks.» *Lecture Notes in Networks and Systems*, vol. 1000, 2024, pp. 63–70. doi: 10.1007/978-981-97-3289-0_5
12. V. Ivanyuk, K. Shuvalov, G. Akhobadze «Development of Modern Forecasting Models.» *Proceedings of 2023 16th International Conference Management of Large-Scale System Development, MLSD 2023*, 2023, doi: 10.1109/MLSD58227.2023.10303848
13. Gorlacheva E.N., Omelchenko I.N., Drogovoz P.A. et al. 2022, 'Using a Box-Jenkins model for production factor forecasting in the radio-electronic industry', *AIP Conference Proceedings*, vol. 2383, no. -, pp. -. doi: 10.1063/5.0074528
14. Zubov N., Misrikhanov M., Ryabchenko V. et al. 2020, 'Prediction of time series of overhead lines failure rate with chaotic indicators', *E3S Web of Conferences*, vol. 216, no. -, pp. -. doi: 10.1051/e3sconf/202021601016
15. Kong Wen & Lu Yifan & Hou Linyuan. (2023). A Review of Chaotic Time Series Prediction Methods. 134–137. doi: 10.1109/CCDC58219.2023.10326789.
16. Rakitskiy Anton. (2022). Efficient Algorithms for Time Series Prediction Method. 120–123. doi: 10.1109/SIBIRCON56155.2022.10016954.
17. Qiang Chen & Xin Qi. (2019). Research on Trend Analysis and Prediction Algorithm Based on Time Series. 73–76. doi: 10.1109/EITCE47263.2019.9095181.
18. Li Zhenwei & Han Jing & Song Yuping. (2020). On the forecasting of high frequency financial time series based on ARIMA model improved by Deep Learning. Journal of Forecasting. 39. 10.1002/for.2677.
19. Kelany Omnia & Aly Sherin & Ismail, Mohamed. (2020). Deep Learning Model for Financial Time Series Prediction. 120–125. 10.1109/IIT50501.2020.9299063.
20. Wanchoo Karan. (2019). Retail Demand Forecasting: A Comparison between Deep Neural Network and Gradient Boosting Method for Univariate Time Series. 1–5. 10.1109/I2CT45611.2019.9033651.

© Хабаров Марк Эдуардович (mark.khabarov@yandex.ru); Ловцов Владимир Алексеевич (lovets1818@gmail.com); Тяжелникова Ксения Сергеевна (kstyazhelnikova@gmail.ru); Гуренко Владимир Викторович (wgurenko@bmstu.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»