

УДК 659.1

¹Н.И. Ломакин, ²М.С. Марамыгин, ³Е.Г. Черная, ⁴С.А. Кращенко

¹Волгоградский государственный технический университет, Волгоград,
email: tel9033176642@yahoo.com

²Уральский государственный экономический университет, Екатеринбург,
email: maram_m_s@mail.ru

³Волжский институт экономики, педагогики и права, Волжский, email: tchornaja@mail.ru

⁴Волгоградский филиал РЭУ им. Г.В. Плеханова, Волгоград, email: krashhenkos@bk.ru

ОПТИМИЗАЦИЯ АЛГОРИТМА ПРИМЕНЕНИЯ МОДЕЛИ ГЛУБОКОГО ОБУЧЕНИЯ «RANDOM FOREST REGRESSION» ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ ПРОГНОЗИРОВАНИЯ ЦЕНЫ ФИНАНСОВОГО АКТИВА

Ключевые слова: DL-модель, финансовая сфера, точность прогноза, фьючерсный контракт SiH4, Random Forest Regression, биржевые роботы.

В современных условиях, обусловленных широкомасштабным применением искусственного интеллекта во всех сферах человеческой деятельности, важное значение имеет успешное применение DL-моделей в финансовой сфере, в алгоритмах биржевых торговых системах. Практика показывает, что обучение с учителем имеет широкие возможности. При таком подходе используются прошлые помеченные данные, которые используются для прогнозирования будущего путем выявления закономерностей в этих данных. К такому типу обучения относятся модели линейной регрессии и логистической регрессии. Рассмотрены подходы в формировании прогноза цены финансового актива – фьючерсного контракта на доллар SiH4 с применением системы искусственного интеллекта. Актуальность исследования состоит в том, что в современных условиях все чаще используются нейросети для формирования прогнозов цен финансовых активов на биржевом рынке. Научная новизна – выдвинута и доказана гипотеза, что, меняя тайм-фрейм от 1 часового, до 15 минутного и 1 минутного можно повысить точность прогноза цены, используя DL-модель «Random Forest Regression». В ходе исследования были построены три DL-модели на разных тайм-фреймах. Практическая значимость в том, что результаты исследования могут быть использованы на практике в алгоритмах биржевых торговых роботов, что повышает результативность при проведении спекулятивных биржевых операций.

¹N.I. Lomakin, ²M.S. Maramygin, ³E.G. Chernaya, ⁴S.A. Kraschenko

¹Volgograd State Technical University, Volgograd, email: tel9033176642@yahoo.com

²Ural State Economic University, Ekaterinburg, email: maram_m_s@mail.ru

³Volzhsky Institute of Economics, Pedagogy and Law, Volzhsky, email: tchornaja@mail.ru

⁴Volgograd branch of REU named G.V. Plekhanov, Volgograd, email: krashhenkos@bk.ru

OPTIMIZATION OF THE ALGORITHM FOR APPLYING THE DEEP LEARNING MODEL “RANDOM FOREST REGRESSION” TO INCREASE THE ACCURACY OF FORECASTING THE PRICE OF A FINANCIAL ASSET

Keywords: DL model, financial sector, forecast accuracy, SiH4 futures contract, Random Forest Regression, stock robots.

In modern conditions, due to the large-scale use of artificial intelligence in all spheres of human activity, the successful use of DL models in the financial sector, in algorithms of exchange trading systems, is important. Practice shows that learning with a teacher has ample opportunities. This approach takes past labeled data and uses it to predict the future by identifying patterns in that data. This type of learning includes linear regression and logistic regression models. Approaches to forming a forecast for the price of a financial asset – a futures contract for the dollar SiH4 using an artificial intelligence system are considered. The relevance of the study lies in the fact that in modern conditions neural networks are increasingly used to generate forecasts for the prices of financial assets on the stock market. The scientific novelty lies in the fact that a hypothesis has been put forward and proven that by changing the time frame from 1 hour, to 15 minutes and 1 minute, you can increase the accuracy of the price forecast using the DL-model “Random Forest Regression”. During the study, three DL models were built on different time frames. The practical significance is that the results of the study can be used in practice in exchange trading robot algorithms, which increases the effectiveness of speculative exchange operations.

Актуальность исследования состоит в том, что в современных условиях все чаще используются нейросети для формирования прогнозов цен финансовых активов на биржевом рынке. Научная новизна – в том, что выдвинута и доказана гипотеза, что, меняя тайм-фрейм от 1 минутного, 15 минутного и до 1 часового, можно повысить точность прогноза цены, используя DL-модель «Random Forest Regression».

Практика показывает, что обучение с учителем имеет широкие возможности. При таком подходе используются прошлые помеченные данные, которые используются для прогнозирования будущего путем выявления закономерностей в этих данных. К такому типу обучения относятся модели линейной регрессии и логистической регрессии. Рассмотрены подходы в формировании прогноза цены финансового актива – фьючерсного контракта на доллар SiH4 с применением системы искусственного интеллекта.

В ходе исследования были поставлены и решены задачи:

1) исследовать тенденции применения искусственного интеллекта в финансовой сфере;

2) сформировать DL-модели для прогноза цены закрытия фьючерсного контракта SiH4 на различных тайм-фреймах: 1 минута, 15 минут, 1 час;

3) сравнить параметры, отражающие точность прогнозирования DL-модели.

Тенденции развития искусственного интеллекта в финансовой сфере свидетельствуют об увеличении масштабов его применения для формирования прогнозов, поиска закономерностей, оценки финансового риска, а также для обеспечения поддержки принятия управленческих решений в ходе цифровизации. Как показывает практика, в условиях широкомасштабной цифровизации становится актуальным изучение вопросов использования искусственного интеллекта для достижения устойчивого развития экономики, ограничения финансовых рисков при нарастающей рыночной неопределенности. Автор Абдалмуттаleb М.А. и коллеги сделали обзор современных разработок в области применения искусственного интеллекта для обеспечения стабильности финансирования и цифровых технологий [1].

В современных условиях следует отметить всплеск цифровизации бизнес-процессов, что связано с внедрением технологий «Индустрия 4.0», формированием предпосылок для внедрения нового технологического уклада «Индустрия 5.0» [2].

Работа, выполненная Зхан (Zhan) предполагает, что модели построения диаграмм могут эффективно изучать временные зависимости в данных временных рядов. Поскольку при сравнении портфелей производных от различных моделей, построения графиков зависимостей, обеспечивали увеличивающуюся доходность при низком риске и превосходили индекс S&P 500 [3].

В процессе разработки робо-эдвайзера на основе искусственного интеллекта по методу «Случайный лес» Ломакин Н.И. с коллегами предложили использовать модель машинного обучения на основе метода «Machin Learning случайный лес – Random Forest», которая представляет собой высокоэффективный нейросетевой бот-эдвайзер, обеспечивающий поддержку принятия управленческих решений, позволяя получать надежный прогноз на каждый следующий час в ходе биржевой торговли фьючерсным контрактом SiH3 [4, с. 7-21].

Деревья решений (DT) основаны на непараметрическом методе обучения с учителем, который используется для классификации и регрессии. Целью метода является создание модели, предсказывающей значение целевой переменной на основе изучения простых правил принятия решений, которые выводятся из характеристик данных. Дерево можно рассматривать как кусочно-постоянную аппроксимацию. Чем глубже дерево, тем сложнее правила принятия решений и тем точнее модель. Деревья решений используются как для задач классификации, так и для задач регрессии. Понимание важности переменных в лесах случайных деревьев представлено во многих работах, в том числе Лоупе (Louppe) [5].

Цель исследования

Цель исследования состоит в получении более высокой точности при прогнозировании цены закрытия фьючерсного контракта на доллар SiH4 с применением системы искусственного интеллекта при использовании DL-модель «Random Forest Regression» (рис.1).



Рис. 1. График динамики фьючерсного контракта на доллар SiH4 в 2024 году

Таблица 1

Датасет DL-модели «Random Forest Regression» с тайм-фреймами: 1 минута (фрагмент)

Row	Date	Open	High	Low	Close	Volume
4470	25.01.2024	90637.0	90648.0	90637.0	90638.0	122.0
4469	25.01.2024	90627.0	90637.0	90626.0	90637.0	243.0
4468	25.01.2024	90637.0	90637.0	90626.0	90626.0	114.0
4467	25.01.2024	90636.0	90637.0	90633.0	90636.0	77.0
4466	25.01.2024	90633.0	90640.0	90629.0	90637.0	358.0

```
[ ] 1 X = dataset.drop(['target'], axis=1)
    2 y = dataset['target']

[ ] 1 from sklearn.model_selection import train_test_split
    2 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20)

[ ] 1 from sklearn.tree import DecisionTreeRegressor
    2 regressor = DecisionTreeRegressor()
    3 regressor.fit(X_train, y_train)

[ ] 1 print(regressor.score(X_train, y_train))
```

Рис. 2. Скрипт программы на языке Python для DL-модели «Random Forest Regression»

Материал и методы исследования

В ходе проведенного исследования была сформирована и использована DL-модель «Random Forest Regression» для прогнозирования цены закрытия фьючерсного контракта SiH4. Датасеты формировались на основе биржевой

информации, получаемой посредством платформы QUIK с сервера Московской биржи. Обработка исходной информации, формирование и обучение DL-модели с использованием скриптов на языке Python в облачном сервисе Google Collab [6].

Таблица 2

Сравнительная характеристика датасетов DL-модели

Наименование	Тайм-фрейм		
	1 мин	15 мин	1 час
Параметр	Close	Close	Close
Count (Количество строк)	4471	6280	3059
Mean (Среднее значение)	89891,6	84717	90012,2
Std (Среднеквадратическое отклонение)	430,62	7268,08	6348,13
Min (Минимальное значение)	89020	74200	76200
25%	89601	77446	84232
50%	89853	89296	91357
75%	90177	91529	95007
Max (Максимальное значение)	90969	95718	102385

Таблица 3

Коэффициенты парной корреляции на 1 минутном тайм-фрейме

	Open	High	Low	Close	Volume	target
Open	1.000000	0.999229	0.999371	0.998888	0.043850	0.997871
High	0.999229	1.000000	0.998791	0.999416	0.064035	0.998298
Low	0.999371	0.998791	1.000000	0.999354	0.030485	0.998280
Close	0.998888	0.999416	0.999354	1.000000	0.048094	0.998828
Volume	0.043850	0.064035	0.030485	0.048094	1.000000	0.047972
target	0.997871	0.998298	0.998280	0.998828	0.047972	1.000000

Таблица 4

Параметры, отражающие качество прогнозирования DL-моделей

Наименование	Тайм-фрейм		
	1 мин	15 мин	1 час
Средняя абсолютная ошибка (Mean Absolute Error)	19,805	92,780	254,995
Отклонение, %	0	468,5	274,8
Среднеквадратическая ошибка: (Mean Squared Error)	909,4	17239,9	136992,0
Среднеквадратическая погрешность (Root Mean Squared Error)	30, 156	131,301	370,124

Таблица 5

Параметры линейной многофакторной регрессии

Наименование	Тайм-фрейм		
	1 мин	15 мин	1 час
Свободный член	105,3444	65,151	113,404
Цена открытия, Open	0,041	-0,053	-0,082
Цена максимальная, High	-0,016	0,102	0,082
Цена минимальная, Low	0,076	0,192	0,12
Цена закрытия, Close	0,898	0,758	0,879
Объем, Volume	0,0001	0,00009	-0,00006

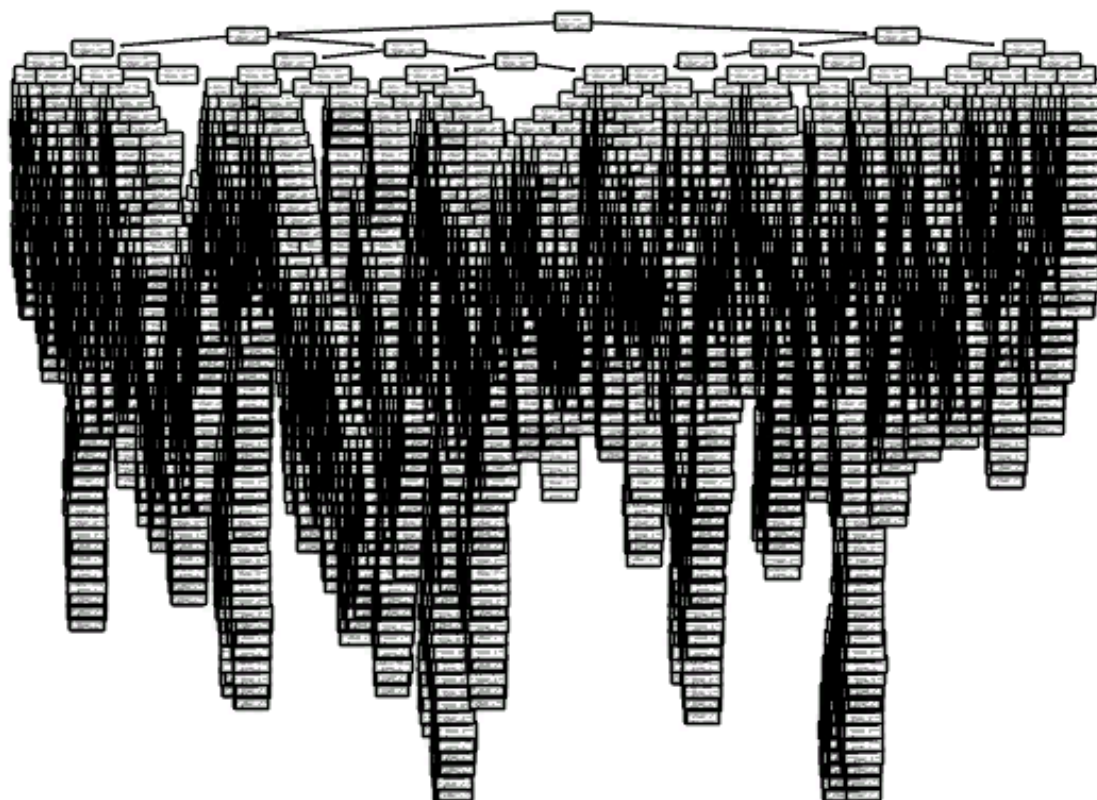


Рис. 3. Дерево решений DL-модели на 1 минутном тайм-фрейме

Были использованы три версии DL-модели, обученные на различных датасетах с тайм-фреймами: 1 минута, 15 минут, 1 час (табл. 1).

Важным шагом предварительной обработки датасета является разделение данных на обучающие и тестовые наборы. Библиотека `model_selection` Scikit-Learn содержит метод `train_test_split`, который был использован для случайного разделения данных на наборы для обучения и тестирования. Для этого был выполнен следующий код (рис. 2).

Сравнительная характеристика датасетов представлена ниже (табл. 2).

Исследования показали, что коэффициенты парной корреляции имеют высокие значения на всех трех тайм-фреймах, например, на 1 минутном (табл. 3).

Исследования показали, что точность прогноза возрастает по мере сокращения тайм-фрейма. В частности, Средняя абсолютная ошибка (MAE) увеличивается в разы по мере перехода на более продолжительные тайм-фреймы.

Так, например, MAE на 15 минутах в 4,6 раз больше чем на 1 минутном, и в 2,7 раза больше на тайм-фрейме 1 час в сравнении с 15 минутным.

Качество прогнозирования DL-моделей представлено ниже (табл. 4).

В ходе исследования при использовании библиотеки Scikit-learn, были получены параметры линейной многофакторной регрессии (табл. 5).

В целях визуализации ниже представлено лучшее дерево из ансамбля деревьев решений, DL-модели, сформированной на 1 минутном тайм-фрейме (рис. 3)

Результаты исследования и их обсуждение

Среди важнейших направлений дальнейших исследований можно выделить такие, как: применение мощных инструментов – сверточных нейронных сетей CNN, использовании NLP, Word2vec, LSTM, а также формировании когнитивных и гибридных киберфизических

систем. Повышение точности и снижение финансового риска при прогнозировании временных рядов фьючерсных контрактов SIU0 с использованием нейронной сети с помощью Word2vec Vector News, было предметом исследования коллектива авторов во главе с Ломакиным Н.И [7]. Когнитивные модели и гибридные киберфизические системы, являясь одним из перспективных направлений, находится в центре внимания многих ученых, в том числе Ломакина Н.И [8, 9]. Кроме отмеченных направлений, заслуживают внимания некоторые результаты зарубежных ученых. Так, например, результаты поиска и выявления закономерностей и факторов, которые влияют на поведение цен на биржевые активы, были получены группой ученых во главе с Сяолин Чу (Xiaoling Chu), результаты представляют собой устойчивые доказательства увеличения риска обвала цен на акции китайских фирм, занимающихся недвижимостью, которые чрезмерно полагались на заемное финансирование [10].

Заслуживают внимание результаты выявления и исследования скрытых факторов, обуславливающих изменение цен биржевых активов, как это представил Сяоин Дэн (Xiaoying Deng) с коллегами, исследуя, влияет ли индивидуальное рискованное поведение на финансирование недвижимости через теневые банки [11], а также итоги исследования Франклина Аллена (Franklin A.) касательно ценообразования доверенных кредитов,

особенно неаффилированных кредитов, которые включают в себя фундаментальные и информационные риски, что предопределяет реакцию фондового рынка [12].

Выводы

Таким образом, на основании проведенного исследования были достигнуты поставленные задачи:

1) исследовать тенденцию применения искусственного интеллекта в финансовой сфере;

2) сформировать DL-модели для прогноза цены закрытия фьючерсного контракта SiH4 на различных тайм-фреймах: 1 минута, 15 минут, 1 час;

3) сравнить параметры, отражающие точность прогнозирования DL-модели.

В результате исследования были сформулированы выводы.

Во-первых, выдвинута и доказана гипотеза, что, меняя тайм-фрейм от 1 часового, до 15 минутного и 1 минутного можно повысить точность прогноза цены, используя DL-модель «Random Forest Regression».

Во-вторых, исследования показали, что точность прогноза возрастает по мере сокращения тайм-фрейма. В частности, Средняя абсолютная ошибка (MAE) увеличивается в разы по мере перехода на более продолжительные тайм-фреймы. Так, например, MAE на 15 минутах в 4,6 раз больше чем на 1 минутном, и в 2,7 раза больше на тайм-фрейме 1 час в сравнении с 15 минутным.

Библиографический список

1. Abdalmuttaleb M.A. Artificial Intelligence for Sustainable Finance and Sustainable Technology. M. Al-Sartawi. ICGER: The International Conference On Global Economic Revolutions, LNNS. 2021. Vol. 423. P. 15-16.
2. Индустрия 4.0: Big Data, цифровизация и рост экономики [Электронный ресурс]. URL: <https://habr.com/ru/post/507822/> (дата обращения 24.01.2024)
3. Zhan N., Sun Y., Jakhar A. and Liu H. Graphical Models for Financial Time Series and Portfolio Selection. Published at ACM International Conference on AI in Finance (ICAIF '20) 2021. [Электронный ресурс]. URL: <https://arxiv.org/format/2101.09214> (дата обращения 24.01.2024)
4. Ломакин Н.И., Юрова О.В., Терехов Т.В., Шабанов Н.Т. Разработка робо-эдвайзера на основе искусственного интеллекта по методу «Случайный лес» как фактор повышения инвестиционной активности населения // п-Economy. 2023. Т. 16, № 3. С. 7-21. DOI: 10.18721/JE.16301.
5. Louppe G., Wehenkel L., Suter A. and Geurts P. Understanding variable importances in forests of randomized trees. 2020. P. 9-10 [Электронный ресурс]. URL: <https://proceedings.neurips.cc/paper/2013/file/e3796ae838835da0b6f6ea37bcf8bcb7-Paper.pdf> (дата обращения 24.01.2024).

6. Imin_Chip-bot_simple_SiU3_Model_RFRegressor [Электронный ресурс]. URL: <https://colab.research.google.com/drive/1-m7YjzytbBvqqoicwhx1IT0iBZQ172cw?usp=sharing> (дата обращения 29.01.2024).
7. Lomakin N., Kulachinskaya A., Maramygin M., Chernaya E. Improving Accuracy and Reducing Financial Risk When Forecasting Time Series of SIU0 Future Contracts Employing Neural Network with Word2vec Vector News. *Studies in Systems, Decision and Control* this link is disabled. 2022. Vol. 415. P. 281-298.
8. Ломакин Н.И., Кулачинская А., Цыганкова В.Н., Кособокова Е., Минаева О.А., Трунина В.Ф. Forecast of Stability of the Economy of the Russian Federation with the AI-System "Decision Tree" in a Cognitive Model // *International Journal of Technology (IJTech)*. 2023. Vol. 14, Is. 8. P. 1800-1809. DOI: 10.14716/ijtech.v14i8.6848.
9. Ломакин Н.И., Голодова О., Марамигин М., Кузьмина Т., Минаева О.А. Hybrid Cyber-Physical System QUIK-LUA-Random Forest for Trading on MoEx // *Creativity in Intelligent Technologies and Data Science: 5th International Conference CIT&DS 2023 (Volgograd, Russia, September 11-15, 2023): Proceedings* / eds.: A.G. Kravets, M.V. Shcherbakov, P.P. Groumpos; Volgograd State Technical University [et al.]. Cham (Switzerland): Springer Nature Switzerland AG, 2023. P. 64-79. DOI: 10.1007/978-3-031-44615-3_5.
10. Xiaoling C., Yongheng D., Desmond T. Firm Leverage and Stock Price Crash Risk: The Chinese Real Estate Market and Three-Red-Lines Policy // *The Journal of Real Estate Finance and Economics*. 2023. [Электронный ресурс]. URL: <https://link.springer.com/article/10.1007/s11146-023-09953-0> (дата обращения 24.01.2024).
11. Xiaoying D., Chong L., Seow E.O. Shadow Bank, Risk-Taking, and Real Estate Financing: Evidence from the Online Loan Market // *The Journal of Real Estate Finance and Economics*. 2023 [Электронный ресурс]. URL: <https://link.springer.com/article/10.1007/s11146-022-09936-7> (дата обращения 24.01.2024).
12. Franklin A., Qian Y., Guoqian Tu., Frank Yu. Entrusted loans: A close look at China's shadow banking system // *Journal of Financial Economics*. 2019. Vol. 133, Is. 1. P. 18-41. DOI: 10.1016/j.jfineco.2019.01.006.