

61. Penatti O.A.B., Nogueira K., Santos J.A. dos. Do Deep Features Generalize from Everyday Objects to Remote Sensing and Aerial Scenes Domains?, *2015 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015, pp. 44-51. DOI: 10.1109/CVPRW.2015.7301382.
62. Chen T., Kornblith S., Norouzi M., Hinton G. A Simple Framework for Contrastive Learning of Visual Representations, *arXiv*, 2020. arXiv:2002.05709. DOI: 10.48550/arXiv.2002.05709.
63. Xu J., Tang H., Ren Y., Peng L., Zhu X., He L. Multi-level Feature Learning for Contrastive Multi-view Clustering, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 16030-16039. DOI: 10.1109/CVPR52688.2022.01558.
64. Wang Y.; Albrecht C.M.; Ait Ali Braham N.; Mou L.; Zhu X.X. Self-Supervised Learning in Remote Sensing: A Review, *IEEE Geoscience and Remote Sensing Magazine*, 2022, Vol. 10, No. 4, pp. 213-247. DOI: 10.1109/MGRS.2022.3198244.
65. Ayuba D. L.; Guillemaut J.-Y.; Marti-Cardona B.; Mendez Maldonado O. SpecBPP: A Self-Supervised Learning Approach for Hyperspectral Representation and Soil Organic Carbon Estimation, *arXiv*, 2025. arXiv:2507.19781. DOI: 10.48550/arXiv.2507.19781.
66. Liu B., Yu X., Zhang P. A Semi-Supervised Convolutional Neural Network for Hyperspectral Image Classification, *Remote Sensing Letters*, 2017, Vol. 8, No. 9, pp. 839-848. DOI: 10.1080/2150704X.2017.1331053.
67. Wang L., Hao S., Wang Q., Wang Y. Semi-Supervised Classification for Hyperspectral Imagery Based on Spatial-Spectral Label Propagation, *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, Vol. 97, pp. 123-137. DOI: 10.1016/j.isprsjprs.2014.08.016.
68. Zhang D., Ren Y., Liu C., Han Z., Wang J. Weighted Contrastive Prototype Network for Few-Shot Hyperspectral Image Classification with Noisy Labels, *Remote Sensing*, 2024, Vol. 16, No. 18, Art. 3527. DOI: 10.3390/rs16183527.
69. Ait Ali Braham N., Wang Y., Albrecht C.M., Mou L., Zhu X.X. Self-Supervised Learning for Few-Shot Hyperspectral Image Classification, *arXiv*, 2022. arXiv:2206.12117. DOI: 10.48550/arXiv.2206.12117.
70. Kumar V., Singh R.S., Rambabu M., Dua Y. Deep learning for hyperspectral image classification: A survey, *Computer Science Review*, 2024, Art. 100658. DOI: 10.1016/j.cosrev.2024.100658.

Филонова Марьяна Андреевна – Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова; e-mail: filonovamaryana@ynadex.ru; г. Новочеркасск, Россия; тел.: +79185767418; кафедра «Информационные и измерительные системы и технологии»; ассистент; аспирант.

Широбокова Светлана Николаевна – Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова; e-mail: shirobokova_sn@mail.ru; г. Новочеркасск, Россия; тел.: +79518372637; кафедра «Информационные и измерительные системы и технологии»; к.э.н.; доцент.

Filonova Maryana Andreevna – Platov South-Russian State Polytechnic University (NPI); e-mail: filonovamaryana@ynadex.ru; Novocherkassk, Russia; phone: +79185767418; the Department «Information and measuring systems and technologies»; master's student.

Shirobokova Svetlana Nikolaevna – Platov South-Russian State Polytechnic University (NPI); e-mail: shirobokova_sn@mail.ru; Novocherkassk, Russia; phone: +79518372637; the Department «Information and measuring systems and technologies»; cand. of econ. sc.; associate professor.

УДК 004.942

DOI 10.18522/2311-3103-2026-3-208-219

Л.Э. Хайруллина, З.Н. Хакимов, Д.И. Галиев, А.Н. Хайруллина

ПРОГНОЗИРОВАНИЕ ДВИЖЕНИЯ ЦЕНЫ ОБЛИГАЦИИ С ПОМОЩЬЮ ГИБРИДНОГО МЕТОДА НА ОСНОВЕ XGBOOST И ГЕНЕТИЧЕСКОГО АЛГОРИТМА

Представлен гибридный метод к прогнозированию направления движения цены облигаций, сочетающий метод машинного обучения XGBoost с оптимизацией гиперпараметров посредством генетического алгоритма. Исследование направлено на решение задачи бинарной классификации направления цены облигации ОАО «РЖД» на следующий торговый день. Методология исследования включает формирование расширенного признакового пространства из 18 технических индикаторов, рассчитанных на основе дневных OHLCV-данных. Для настройки гиперпараметров

XGBoost реализован генетический алгоритм с использованием библиотеки DEAP. Исследование проводилось на трех временных горизонтах: 01.01.21-31.10.25, 01.01.23-31.10.25, с 01.01.24-31.10.25. В результате показана существенная зависимость эффективности модели от временного горизонта обучающих данных. Наилучшее качество продемонстрировала модель, обученная на данных за 2024–2025 годы, с точностью на тестовой выборке 64.4%, сбалансированными метриками *precision* и *recall*, а также высокими оценками *F1-score* для обоих классов. Модели, обученные на более длинных периодах (2021–2025 и 2023–2025 гг.), показали снижение обобщающей способности, что свидетельствует о превалировании актуальности данных над их объемом в условиях изменения денежно-кредитной политики России в 2021–2025 гг. Для поддержания прогностической силы модели в меняющихся рыночных условиях рекомендовано использование скользящего окна обучения продолжительностью 1.5–2 года. Проведенное сравнение со стратегией «Buy & Hold» подтвердило эффективность предложенного гибридного подхода.

XGBoost; генетический алгоритм; гибридный метод; облигации; финансовый ряд; DEAP.

L.E. Khairullina, Z.N. Khakimov, D.I. Galiev, A.N. Khairullina

PREDICTING BOND PRICE MOVEMENTS USING A HYBRID METHOD BASED ON XGBOOST AND A GENETIC ALGORITHM

The article presents a hybrid method for predicting the direction of bond price movements, combining the XGBoost machine learning method with hyperparameter optimization using a genetic algorithm. The research is aimed at solving the problem of binary classification of the direction of the price of the Russian Railways bond on the next trading day. The research methodology includes the formation of an expanded feature space of 18 technical indicators calculated on the basis of daily OHLCV data. To configure XGBoost hyperparameters, a genetic algorithm is implemented using the DEAP library. The study was conducted on three time horizons: 01.01.21-31.10.25, 01.01.23-31.10.25, from 01.01.24-31.10.25. As a result, a significant dependence of the effectiveness of the model on the time horizon of the training data is shown. The best quality was demonstrated by a model trained on data from 2024–2025, with an accuracy of 64.4% in the test sample, balanced precision and recall metrics, as well as high *F1-score* scores for both classes. Models trained over longer periods (2021–2025 and 2023–2025) showed a decrease in generalizing ability, which indicates that the relevance of the data prevails over its volume in the context of changes in Russia's monetary policy in 2021–2025. To maintain the predictive power of the model in changing market conditions, it is recommended to use a sliding learning window of 1.5–2 years. The comparison with the "Buy & Hold" strategy confirmed the effectiveness of the proposed hybrid approach.

XGBoost; genetic algorithm; hybrid method; bonds; financial series; DEAP.

Введение. Прогнозирование колебаний цен облигаций остается одной из ключевых задач в финансовой аналитике, поскольку рынок долговых инструментов все более подвержен влиянию макроэкономических факторов, новостных событий и сдвигов в ликвидности. В последние годы классические эконометрические модели, такие как ARIMA или GARCH, часто демонстрируют ограниченную способность к захвату нелинейных и динамичных паттернов в ценовых движениях. В связи с этим растет интерес к гибридным подходам на базе машинного обучения [1–3].

Одним из наиболее эффективных в прогнозировании динамики финансовых временных рядов является алгоритм XGBoost. XGBoost (eXtreme Gradient Boosting) – это масштабируемый и эффективный алгоритм ансамблевого машинного обучения, основанный на градиентном бустинге над деревьями решений. Он минимизирует регуляризованную функцию потерь, включающую как ошибку модели, так и штрафы за сложность деревьев (L1- и L2-регуляризация), что обеспечивает высокую обобщающую способность и устойчивость к переобучению [4]. Оптимизация производится итеративно с использованием градиента и гессиана функции потерь, а архитектура алгоритма допускает параллелизацию и эффективное использование памяти. XGBoost эффективно работает на сравнительно небольших выборках, устойчив к шуму, а регуляризованная функция потерь позволяет лучше обобщать закономерности на новых данных. На сегодняшний день применение алгоритма XGBoost в различных прикладных задачах достаточно широко освещено в работах отечественных и зарубежных авторов и интерес к нему не утихает. Так, в работе [5] данный алгоритм используется для создания модели, предсказывающей, завершит ли пользователь онлайн-курс, а в статье [6] – для классификации кандидатов по вакантным работам.

Значительное количество работ посвящено сравнительному анализу алгоритма XGBoost с другим методами. Исследования [7] подтвердили, что XGBoost может достигать большей точности прогнозирования цен облигаций, чем традиционные модели. В работе [8] установлено, что в задаче прогнозирования волатильности акций Сбербанка XGBoost демонстрирует несколько лучшие результаты по сравнению с LSTM. Бакуменко Л.П., Васильева Н.С. [9] постулируют, что LSTM-модель может быть предпочтительной для более долгосрочных прогнозов, а модель XGBoost - для краткосрочных. В статье авторов Zhang, Bian и др. [10] XGBoost показал наилучшую точность, превзойдя методы ARIMA, LSTM, Prophet и GBDT. В работе [11] авторы заключают, что в задаче прогнозирования спроса на продукцию модели SARIMA и XGBoost дополняют друг друга и предлагают создание гибридного метода для улучшения точности прогнозов. Отметим, что в последние годы интерес к подобным гибридным методам прогнозирования, дополняющих XGBoost, только растет. Например, в работе [12] авторы получили интегрированную модель: ARIMA (для линейных компонент), LSTM (для долгосрочных временных зависимостей) и XGBoost (для нелинейных паттернов). Гибридная модель ARIMA-LSTM-XGBoost продемонстрировала превосходство над отдельными алгоритмами по точности прогнозирования. Odhiambo с соавторами [13] разработали гибридную модель ARIMA-XGBOOST для моделирования динамики мобильных денежных операций в Кении. В статье [14] для прогнозирования временного ряда также используются несколько методов, среди которых комбинация XGBoost с LSTM. В работах [15–18] отмечается, что XGBoost в ансамбле с генетическим алгоритмом для отбора признаков и настройки параметров делает модель более устойчивой в таких задачах, как обнаружение мошенничества или финансовом прогнозировании.

Данная статья имеет своей целью реализацию гибридного метода к прогнозированию направления движения цены облигации на основе комбинации технического анализа, машинного обучения (XGBoost) и генетического алгоритма (ГА) с дальнейшим сравнением его эффективности на различных временных диапазонах.

Объектом исследования является временной ряд дневных цен облигации ОАО «РЖД» (тикер «РЖД Б01Р2R»).

Предметом исследования являются прогностические способности гибридного метода XGBoost+ГА в задаче бинарной классификации направления движения цены облигации на следующий торговый день.

1. Постановка задачи. Пусть p_i – цена закрытия облигации ОАО «РЖД» в торговый день i , x_i – вектор признаков, сформированный на момент времени i на основе 18 технических индикаторов (трендовые, импульсные, осцилляторы, индикаторы волатильности), рассчитанных по дневным OHLCV-данным (Open, High, Low, Close, Volume) облигации ОАО «РЖД». Описание всех индикаторов приведено в разделе 3.

Определим бинарную целевую переменную $y_{i+1} \in \{0,1\}$, кодирующую направление движения цены на следующий торговый день:

$$y_{i+1} = \begin{cases} 1, & p_{i+1} > p_i \\ 0, & p_{i+1} \leq p_i \end{cases}$$

где $y_{i+1} = 1$, если ожидается рост цены, $y_{i+1} = 0$, если ожидается снижение цены или цена остается без изменения.

Требуется построить классификатор направления движения цены облигации на следующий торговый день $\hat{y}_{i+1} = f(x_i; \theta)$, где θ – гиперпараметры XGBoost (максимальная глубина дерева, скорость обучения, количество деревьев в ансамбле и др.). Поиск оптимальных θ будем выполнять с помощью генетического алгоритма. В качестве фитнес-функции генетического алгоритма выберем среднюю точность модели на валидационных данных.

В рамках исследования будем рассматривать три временных горизонта: 2021–2025 гг., 2023–2025 гг. и 2024–2025 гг., на которых проверим работоспособность гибридного метода XGBoost+ГА, и выявим временной интервал, на котором метод продемонстрирует наилучшие прогностические способности.

2. Гибридный метод XGBoost+ГА. Предлагаемый гибридный метод объединяет XGBoost и генетический алгоритм: последний автоматически подбирает гиперпараметры XGBoost, оптимизируя точность классификации. Схематично работа данного метода приведена на рис. 1.

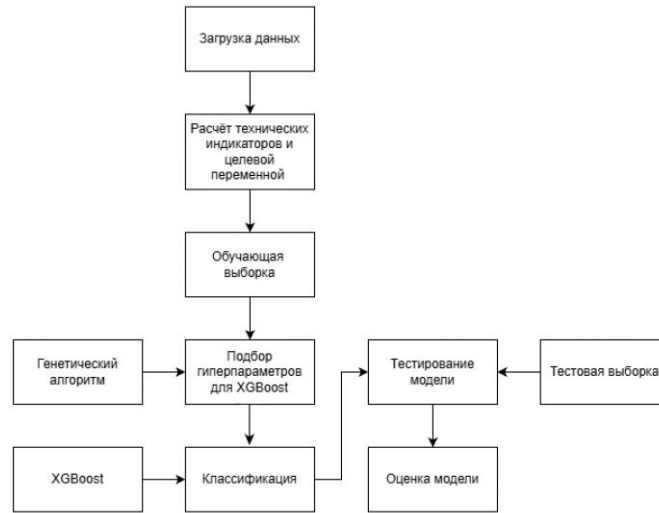


Рис. 1. Схема гибридного метода XGBoost+ГА

Опишем отдельно каждый алгоритм.

Алгоритм XGBoost. Пусть у нас есть обучающие данные:

$$D_{train} = \{(x_i, y_i)\}_{i=1}^N,$$

где x_i – вектор из 18 технических индикаторов, рассчитанных на день i , $y_i \in \{0,1\}$ – целевая переменная, N – количество наблюдений в обучающей выборке.

Градиентный бустинг – это техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений:

$$y_i = \varphi(x_i) = \sum_{k=1}^K f_k(x_i), \quad f_k \in F,$$

где F – множество всех возможных деревьев решений, K – число деревьев, f_k – k -е дерево.

Обучение ансамбля проводится последовательно. На каждой итерации вычисляются отклонения предсказаний уже обученного ансамбля на обучающей выборке. Следующая модель, которая будет добавлена в ансамбль, будет предсказывать эти отклонения. Таким образом, добавив предсказание нового дерева к предсказаниям обученного ансамбля мы можем уменьшить среднее отклонение модели, которое и является целью оптимизационной задачи [5].

XGBoost (Extreme Gradient Boosting) – это конкретная реализация градиентного бустинга, но с рядом важных усовершенствований, которые делают его быстрее, точнее и устойчивее к переобучению, чем классический градиентный бустинг. Впервые алгоритм XGBoost был предложен Тяньчи Чэнь (Tianqi Chen) и Карлосом Гестрином (Carlos Guestrin) в статье [4]. В отличие от классического бустинга, XGBoost минимизирует регуляризованную функцию потерь:

$$L(\varphi) = \sum_{i=1}^n l(y_i, y_i) + \sum_{k=1}^K \Omega(f_k),$$

где

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda |\omega|^2$$

– регуляризационный член, T – количество листьев в дереве, ω – веса листьев (выходные значения), $\gamma, \lambda \geq 0$ – гиперпараметры, управляющие сложностью модели, l – функция потерь.

В алгоритме XGBoost существуют два основных класса гиперпараметров:

1. Параметры, управляющие сложностью модели (например, максимальная глубина дерева, минимальная сумма весов в листе);

2. Параметры, управляющие процессом обучения (скорость обучения, параметры регуляризации, сэмплирование).

Опираясь на [4], практические руководства от разработчиков (например, [19]) и общепринятую практику, в работе определены диапазоны изменения каждого гиперпараметра XGBoost следующими границами:

- ♦ глубина дерева (max_depth): [3, 10];
- ♦ скорость обучения (learning_rate): [0.001, 0.1];
- ♦ количество деревьев (n_estimators): [100, 1000];
- ♦ доля объектов (строк), используемых при построении каждого дерева (subsample): [0.6, 1.0];
- ♦ доля признаков (колонок), используемых при построении каждого дерева (colsample_bytree): [0.6, 1.0];
- ♦ минимальное уменьшение функции потерь для разбиения узла (gamma): [0, 5];
- ♦ минимальная сумма гессианов (вторых производных) в листе (min_child_weight): [1, 10];
- ♦ L1-регуляризация (reg_alpha): [0, 10];
- ♦ L2-регуляризация (reg_lambda): [0, 10].

Настройка гиперпараметров алгоритма XGBoost осуществлялась с помощью генетического алгоритма.

Генетический алгоритм. Генетический алгоритм имитирует процесс естественного отбора и схематично представлен на рис. 2.

Генетический алгоритм был реализован с применением специализированной библиотеки DEAP (Distributed Evolutionary Algorithms in Python) языка программирования Python [20]. Выбор данного инструментария обусловлен его гибкостью, наличием готовых операторов селекции, скрещивания и мутации, а также проверенной эффективностью при решении задач оптимизации.

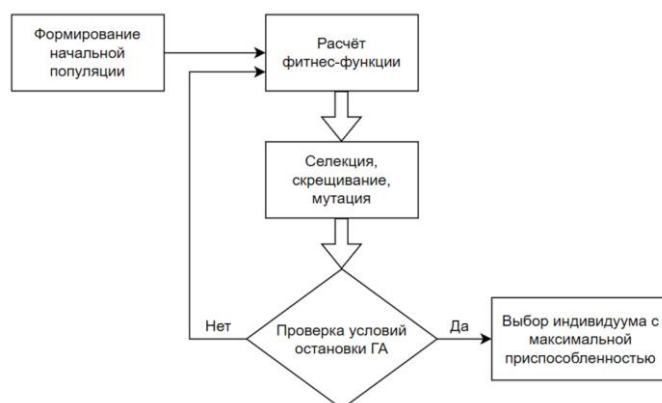


Рис. 2. Схема функционирования генетического алгоритма

Ниже приведен псевдокод алгоритма XGBoost с подбором гиперпараметров на основе генетического алгоритма:

```

1: Формирование начальной популяции  $P \leftarrow \{\theta^{(1)}, \dots, \theta^{(P)}\}$  со случайными гиперпараметрами
2: for each individual  $\theta \in P$  do
3:    $\theta.\text{fitness} \leftarrow \text{Evaluate}(\theta, \mathcal{D}_{\text{train}})$ 
4: end for
5:  $\mathcal{H} \leftarrow \text{HallOfFame}(1)$ 
6: for  $g = 1$  to  $G$  do
7:    $P' \leftarrow \emptyset$ 
8:   while  $|P'| < |P|$  do
9:      $\theta_a, \theta_b \leftarrow \text{TournamentSelect}(P, t=3)$ 
10:     $\theta_c \leftarrow \text{Crossover}(\theta_a, \theta_b; \alpha=0.2)$ 
11:     $\theta_c \leftarrow \text{Mutate}(\theta_c; \text{bounds}, p_{\text{mut}}=0.2)$ 
12:     $P'.\text{append}(\theta_c)$ 
13:   end while
14:    $P \leftarrow P'$ 
15:   for each  $\theta \in P$  do
16:      $\theta.\text{fitness} \leftarrow \text{Evaluate}(\theta, \mathcal{D}_{\text{train}})$ 
17:   end for
19:    $\mathcal{H}.\text{update}(P)$ 
20: end for
21:  $\theta^* \leftarrow \mathcal{H}[0]$ 
22:  $\mathcal{M}^* \leftarrow \text{TrainXGBoost}(\theta^*, \mathcal{D}_{\text{train}})$ 
23: return  $\mathcal{M}^*$ 

```

На вход алгоритма подается начальная популяция хромосом P – векторов из 9 гиперпараметров XGBoost; каждый ген хромосомы кодирует один из гиперпараметров. Популяция – это набор хромосом, которые существуют в одном поколении; в нашем алгоритме размер популяции выбран $P = 30$ особей. Каждая особь популяции представляет собой потенциальное решение задачи, попытку настроить XGBoost.

Фитнес-функция оценивает, насколько хорошо определенная хромосома работает в решении данной задачи. Мы определяем фитнес-функцию как среднюю точность модели XGBoost, обученной с данными гиперпараметрами на валидационных подвыборках. Соответственно, чем выше точность, тем лучше приспособленность. Хромосома с самым высоким показателем пригодности считается лучшей. Копирование лучших особей (хромосом) в новую популяцию происходит с помощью турнирной селекции с размером турнира $t = 3$.

Скрещивание (кроссовер) – обмен генами между двумя родителями для создания потомков. Пусть $p1$ и $p2$ – родительские хромосомы, а ch – их потомок. Каждый ген потомка вычисляется по формуле:

$$ch = (1 - \beta) \cdot p1 + \beta \cdot p2,$$

где $\beta \in [-\alpha, 1 + \alpha]$, α – параметр скрещивания. Для скрещивания мы использовали оператор `sxBlend` из библиотеки DEAP с $\alpha = 0,2$.

Мутация – это оператор, который вводит случайные изменения в один или несколько генов хромосомы для получения новых подмножеств признаков. Для каждого гена хромосомы с фиксированной вероятностью $p_{\text{mut}} = 0,2$ принимается решение, подвергнется ли он мутации. Если ген подлежит мутации, его текущее значение заменяется на новое, случайным образом выбранное из заданного допустимого диапазона для соответствующего гиперпараметра. Такой подход позволяет избежать преждевременной сходимости алгоритма к локальному оптимуму.

Для нового решения вновь рассчитывается фитнес-функция. Отметим, что лучшая особь популяции при этом вносится в специальный контейнер `HallOfFame` («Зал Славы», `hof`) библиотеки DEAP. `HallOfFame` автоматически отслеживает всех особей, проходящих через алгоритм, и сохраняет лучшую. При появлении новой лучшей особи переменная

hof автоматически обновляется. Таким образом, hof всегда содержит одну особь – ту, у которой самое высокое значение фитнес-функции среди всех, что были оценены за всё время. Применение HallofFame гарантирует, что глобальный оптимум, найденный в ходе эволюции, не будет утерян даже при смене поколений.

Критерий останова алгоритма. Алгоритм выполняется фиксированное число поколений (в нашем коде 20), что обеспечивает баланс между качеством оптимизации и вычислительной сложностью. На выходе получим набор гиперпараметров XGBoost, доставляющий наилучшую точность модели.

3. Входные данные. В качестве исходных данных были использованы дневные OHLCV-записи (Open, High, Low, Close, Volume) по облигации ОАО «РЖД» с биржевым тикером «РЖД Б01Р2R» (RU000A0JXQ44) за несколько периодов наблюдения: 1) с 01.01.2021 по 31.10.2025, 2) с 01.01.23 по 31.10.2025, 3) с 01.01.24 по 31.10.2025 (рис. 3).

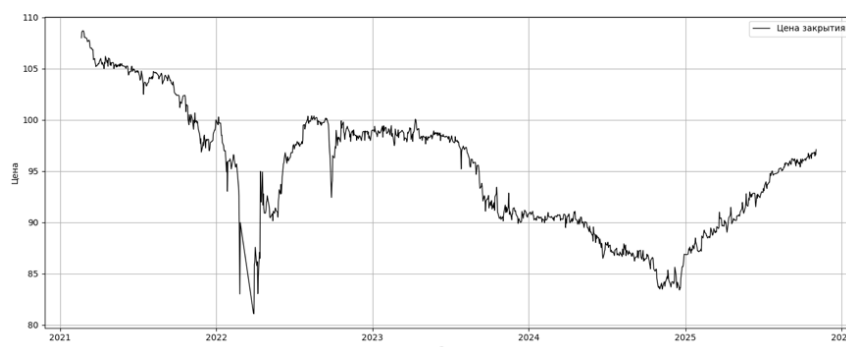


Рис. 3. Движение цены облигации «РЖД Б01Р2R» за период с 01.01.2021 по 31.10.2025

Для получения структурированной информации, выявления скрытых закономерностей, трендов, импульсов и точек разворота на финансовых рынках был сформирован расширенный набор признаков, включающий 18 технических индикаторов, рассчитанных на основе дневных OHLCV-данных. А именно:

- ◆ простые скользящие средние: десятидневные SMA(10), тридцатидневные SMA(30);
- ◆ 12-дневная экспоненциальная скользящая средняя: EMA(12);
- ◆ индекс относительной силы: RSI(14);
- ◆ индекс товарного канала: CCI(20);
- ◆ индикатор «Полосы Боллинджера»: верхняя полоса BB_High, нижняя полоса BB_Low, ширина полосы Боллинджера BB_Width;
- ◆ индикаторы тренда и импульса: MACD_Line, MACD_Signal, MACD_Hist;
- ◆ индикатор волатильности: ATR;
- ◆ сила тренда: ADX;
- ◆ осцилляторы перекупленности/перепроданности: Stoch_K, Stoch_D;
- ◆ доходность: процентное изменение цены за 1 день Return_1;
- ◆ процентное изменение цены за 5 дней: Return_5;
- ◆ волатильность доходности: Volatility_10.

Все индикаторы являются стандартными и подробно описаны в специализированной литературе по техническому анализу [21].

4. Результаты. Эксперимент проводился в соответствии со схемой, представленной на рис. 1. Для каждого из трёх временных горизонтов (2021–2025, 2023–2025, 2024–2025 гг.) данные разделялись на обучающую (80%) и тестовую (20%) выборки с сохранением хронологического порядка (внутри обучающей выборки дополнительно выделялись валидационные подвыборки для оценки фитнес-функции генетического алгоритма). На обучающей выборке запускался генетический алгоритм, находились оптимальные гиперпараметры XGBoost, далее модель оценивалась на тестовой выборке. В ходе эксперимента для каждого временного горизонта фиксировались наилучшая особь (оптимальный набор гиперпараметров) и наиболее важные технические индикаторы. Фиксация наилучшей

особи позволяет проанализировать, какие настройки XGBoost оказались предпочтительными для каждого временного интервала. Фиксация наиболее важных индикаторов даёт возможность интерпретировать модель и выявить, какие технические сигналы вносят основной вклад в прогноз.

Для оценки качества бинарной классификации использовались метрики Accuracy, Precision, Recall и F1-score:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1_score = 2 \frac{Precision \times Recall}{Precision + Recall}$$

Здесь TP, FP, FN, TN являются элементами матрицы ошибок и описаны в табл. 1

Таблица 1

Элементы матрицы ошибок

	Фактически: движение вверх	Фактически: движение вниз или без изменения
Прогноз: движение вверх	TP (модель предсказала рост и цена фактически выросла)	FP (модель предсказала рост, но цена упала или не изменилась)
Прогноз: движение вниз или без изменения	FN (модель предсказала движение вниз или отсутствие роста, но цена выросла)	TN (модель предсказала движение вниз или отсутствие роста и цена действительно не выросла)

Сравнение показателей эффективности алгоритма XGBoost+ГА на различных диапазонах входных данных приведено в табл. 2.

Таблица 2

Сравнение результатов вычислений

Диапазон	01.01.2021 - 31.10.2025	01.01.23 - 31.10.2025	01.01.24 - 31.10.2025
Лучшая особь	[4.45, 0.06, 620, 0.62, 0.91, 3.46, 4.98, 1.74, 2.52]	[3.88, 0.02, 369, 0.72, 0.66, 1.42, 9.01, 0.81, 4.28]	[4.16, 0.05, 485, 0.67, 0.99, 2.20, 4, 2.06, 5]
Наиболее важные технические индикаторы	SMA(10), Return_1, Return_5, MACD_Hist, CCI	Stoch_K, Return_1, Return_5, RSI, MACD_Hist	Return_5, Return_1, Stoch_D, MACD_Hist, Stoch_K
Точность на обучающей выборке (accuracy)	0,579	0,573	0,559
Точность на тестовой выборке (accuracy)	0,575	0,523	0,644
precision (прогноз на рост)	0,57	0,49	0,68
precision (прогноз на спад или без изменения)	0,58	0,81	0,63
recall (прогноз на рост)	0,35	0,95	0,42
recall (прогноз на спад или без изменения)	0,77	0,17	0,83
F1-score (прогноз на рост)	0,43	0,64	0,52
F1-score (прогноз на спад или без изменения)	0,60	0,26	0,72

Полученные результаты показали существенную зависимость качества прогнозной модели от временного диапазона данных при решении задачи бинарной классификации направления движения цены облигации. Модель, обученная на данных за последние 2 года, демонстрирует наилучшее качество на тестовой выборке (64,4%), при этом и баланс между классами значительно улучшен при достаточно высокой F1-score. Модели, обученные на более длинных окнах (2021–2025 и 2023–2025), показывают снижение обобщающей способности. Такой дисбаланс в поведении модели на разных окнах может быть связан с периодами изменения режима денежно-кредитной политики в России в 2021–2025 гг. Модели, обученные на смешанных режимах, усредняют паттерны, теряя чувствительность к текущему состоянию рынка. В то же время данные только за 2024–2025 гг. отражают актуальный режим, что позволяет модели адекватно выявлять как восходящие, так и нисходящие тренды и обеспечивать сбалансированное поведение.

Результаты экспериментов показали, что гибридный метод XGBoost+ГА проявляет адаптивность к структуре временных рядов: на разных временных горизонтах модель автоматически через встроенные механизмы важности признаков XGBoost выбирает различные типы технических индикаторов. Так, на данных 2021–2025 гг. приоритет получают трендовые индикаторы SMA(10), MACD_Hist, осциллятор для поиска точек локального разворота CCI(20). На данных 2023–2025 важными оказались осцилляторы Stoch_K, RSI(14), как сигналы к смене направления. На горизонте 2024–2025 гг. доминируют индикаторы краткосрочного импульса Return_1, Return_5, Stoch_D, улавливающие текущий импульс, недельный тренд и точки разворота. Индикаторы доходности Return_1 и Return_5 входят в число наиболее важных для всех трёх временных горизонтов (см. табл. 2). Это объясняется их фундаментальной ролью в краткосрочном прогнозировании: изменение цены за последний день (или неделю) является наиболее прямым и оперативным сигналом для предсказания движения на следующий день.

Что касается настроек гиперпараметров XGBoost, то модель, обученная на данных 2024–2025 гг., использует умеренную глубину деревьев ($\text{max_depth} \approx 4$) и среднюю скорость обучения ($\text{learning_rate} = 0,05$). Она обладает умеренной регуляризацией, достаточной для подавления шума. Модель использует почти все технические индикаторы при построении каждого дерева ($\text{colsample_bytree} = 0,99$), что подтверждает высокую информативность признаков пространства. Горизонт 2021–2025 гг. характеризуется наибольшей сложностью (количество деревьев $n_estimators = 620$) и высокой регуляризацией разбиений ($\text{gamma} = 3,46$). Это соответствует задаче обучения на разнородных данных, охватывающих несколько рыночных режимов: требуется достаточная ёмкость модели, но с ограничением на создание излишних узлов. Горизонт 2023–2025 демонстрирует самую низкую скорость обучения ($\text{learning_rate} = 0,02$) и наибольшее значение $\text{min_child_weight} = 9,01$, что свидетельствует о консервативной стратегии регуляризации: модель намеренно ограничивает свою сложность, избегая создания листьев с малым весом, чтобы не запоминать случайные или неустойчивые паттерны.

Отметим, что модельная стратегия XGBoost+ГА показала лучший результат и по сравнению с одной из самых простых и распространённых стратегий инвестирования на фондовом рынке Buy&Hold, при которой инвестор покупает финансовые активы, чтобы удерживать их в течение длительного времени с целью добиться роста стоимости, несмотря на колебания рынка (рис. 4).

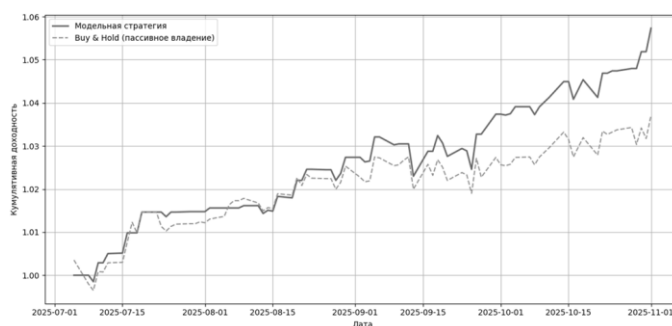


Рис. 4. Сравнение стратегий XGBoost+ГА и Buy&Hold

Заключение. Предложенный в работе гибридный метод сочетает в себе преимущества ансамблевого обучения и эволюционной оптимизации, что обеспечивает ему ряд существенных преимуществ перед существующими аналогами в задаче прогнозирования движения цен ценных бумаг. В отличие от классических эконометрических моделей (ARIMA, GARCH), метод способен улавливать сложные нелинейные зависимости и адаптивно подстраиваться под структурные сдвиги в рыночных данных, не требуя предположений о стационарности ряда. По сравнению с нейросетевыми подходами предложенное решение демонстрирует сопоставимую или более высокую точность при существенно меньших вычислительных затратах на обучение, а также сохраняет интерпретируемость результатов за счёт механизма оценки важности признаков. Ключевой особенностью предложенного метода является автоматизированная настройка гиперпараметров посредством генетического алгоритма, которая исключает субъективный фактор ручного подбора и обеспечивает нахождение глобально оптимальной конфигурации модели. Механизм элитизма (hof) гарантирует, что найденное лучшее решение не будет утеряно в процессе эволюции. Это позволяет получить модель, максимально адаптированную к текущей структуре рынка. Наконец, регуляризованная функция потерь гарантирует устойчивость модели к переобучению на зашумлённых финансовых данных.

Исследование было проведено на рынке корпоративных облигаций – активе, отличающемся от акций более прямой зависимостью от денежно-кредитной политики и невысокой амплитудой дневных ценовых движений. Успешное применение гибридного метода XGBoost+ГА на таком инструменте подтверждает его устойчивость, а также свидетельствуют о потенциале метода для прогнозирования и других классов активов со схожей структурой неопределённости. Анализ, проведенный в работе, выявил, что в условиях современного российского рынка данный метод оптимален для полтора-двухлетнего окна обучения – наибольшую прогностическую силу показала модель, обученная на данных 2024-2025 гг. Увеличение объёма обучающих данных за счёт включения более ранних периодов (2021–2023 гг.) приводит к снижению обобщающей способности модели. В дальнейшем планируется исследовать возможность применения метода для других классов активов, а также адаптировать его для многодневного прогнозирования.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Gu Shihao & Kelly Bryan & Xiu Dacheng. Empirical Asset Pricing via Machine Learning // *The Review of Financial Studies*. – 2020. – 33. – P. 2223-2273. – 10.1093/rfs/hhaa009.
2. Cabrol Axel & Drobetz Wolfgang & Otto Tizian & Puhan Tatjana. Predicting Corporate Bond Illiquidity via Machine Learning // *SSRN Electronic Journal*. – 2023. – 10.2139/ssrn.4489504.
3. Mishin A.A., Vakulenko O.S. Integrated approach to forecasting government bond prices based on event analysis and machine learning // *Economics and management: problems, solutions*. – Moscow, 2025. – Vol. 3, No. 7. – P. 165-174.
4. Chen Tianqi & Guestrin Carlos. XGBoost: A Scalable Tree Boosting System. – 2016. – P. 785-794. – 10.1145/2939672.2939785.
5. Макаров Д.А. Использование алгоритма XGBoost для предсказания завершения курса обучающимся // *StudNet*. – 2021. – № 1. – URL: <https://cyberleninka.ru/article/n/ispolzovanie-algoritma-xgboost-dlya-predskazaniya-zaversheniya-kursa-obuchayushimsya>.
6. Syahputra Akhmal & Saputro Rujianto. Application of the XGBoost Model with Hyperparameter Tuning for Industry Classification for Job Applicants. – 2024. – *SinkrOn* 8. – P. 1920-1931. – 10.33395/sinkron.v8i3.13840.
7. Yan Q., Hu H. XGBoost Machine Learning Model and Bond Pricing // *SSRN*. – 2024. – URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4846755.
8. Салазов А.В. Сравнительный анализ эффективности XGboost и LSTM для прогнозирования волатильности акций Сбербанка // *Auditorium*. – 2025. – № 2 (46). – URL: <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-effektivnosti-xgboost-i-lstm-dlya-prognozirovaniya-volatilnosti-aksiy-sberbanka>.
9. Бакуменко Л.П., Васильева Н.С. Сравнительный анализ прогностических моделей развития криптовалютного рынка (LSTM и XGBOOST) на примере биткоина // *Учет и статистика*. – 2023. – № 4 (72). – URL: <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-prognosticheskikh-modeley-razvitiya-kriptovalyutnogo-rynka-lstm-i-xgboost-na-primere-bitkoina>.
10. Zhang Lingyu & Bian Wenjie & Qu Wenyi & Tuo Liheng & Wang Yunhai. Time series forecast of sales volume based on XGBoost // *Journal of Physics: Conference Series*. – 2021. – 1873. – 012067. – 10.1088/1742-6596/1873/1/012067.

11. *Thejovathi Murari and Rao, Dr. M V P Chandra Sekhara and Priyadharsini E.J. and Siddi Someshwar and Karthik B. and Abbas Syed Hauider.* Optimizing Product Demand Forecasting with Hybrid Machine Learning and Time Series Models: A Comparative Analysis of XGBoost and SARIMA (November 15, 2024) // Proceedings of the 3rd International Conference on Optimization Techniques in the Field of Engineering (ICOFE-2024).
12. *Nichani Rabia & Laid Gasmi & Laiche Nabil & Salheddine Kabou.* Optimizing financial time series predictions with hybrid ARIMA, LSTM, and XGBoost Models // Studies in Engineering and Exact Sciences. – 2024. – 5. e11188. – 10.54021/seesv5n2-582.
13. *Odhambo Stephen & Nyakundi Cornelious & Waititu Hellen.* Developing a Hybrid ARIMA-XGBOOST Model for Analysing Mobile Money Transaction Data in Kenya // Asian Journal of Probability and Statistics. – 2024. – 26. – P. 108-126. – 10.9734/ajpas/2024/v26i10662.
14. *Gond B. PHM-MAX-XGBoost-LSTM-SVM: A Hybrid Model for Time Series Forecasting – A Comparative Analysis with Traditional and Modern Approaches.* – 2025. – 10.31224/4754.
15. *Adil Mehday, Abdellah Chehri, Abdeslam Jakimi, Rachid Saadane.* Hyperparameter Optimization with Genetic Algorithms and XGBoost: A Step Forward in Smart Grid Fraud Detection // Sensors. – 2024. – URL: <https://www.mdpi.com/1424-8220/24/4/1230>.
16. *Lingeswari R., Brindha S.* Online payments fraud prediction using optimized genetic algorithm based feature extraction and modified loss with XGboost algorithm for classification // Swarm and Evolutionary Computation. – Vol. 95. – <https://doi.org/10.1016/j.swevo.2025.101934>.
17. *Kyung Keun Yun, Sang Won Yoon, Daehan Won.* Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process // Expert Systems with Applications. – 2021. – Vol. 186, 115716. – ISSN 0957-4174. – <https://doi.org/10.1016/j.eswa.2021.115716>.
18. *Yun K.K., et al.* Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process // Expert Systems with Applications. – 2021. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0957417421010988>.
19. *A Guide on XGBoost hyperparameters tuning.* – Режим доступа: <https://www.kaggle.com/code/prashant111/a-guide-on-xgboost-hyperparameters-tuning>.
20. *DEAP documentation.* – URL: <https://deap.readthedocs.io/en/master/>.
21. *Швагер Д.* Технический анализ: полный курс: перевод с английского. – М.: Альпина Паблишер, 2021. – 804 с.

REFERENCES

1. *Gu Shihao & Kelly Bryan & Xiu Dacheng.* Empirical Asset Pricing via Machine Learning, *The Review of Financial Studies*, 2020, 33, pp. 2223-2273. 10.1093/rfs/hhaa009.
2. *Cabrol Axel & Drobetz Wolfgang & Otto Tizian & Puhon Tatjana.* Predicting Corporate Bond Illiquidity via Machine Learning, *SSRN Electronic Journal*, 2023. 10.2139/ssrn.4489504.
3. *Mishin A.A., Vakulenko O.S.* Integrated approach to forecasting government bond prices based on event analysis and machine learning, *Economics and management: problems, solutions*. Moscow, 2025, Vol. 3, No. 7, pp. 165-174.
4. *Chen Tianqi & Guestrin Carlos.* XGBoost: A Scalable Tree Boosting System, 2016, pp. 785-794. 10.1145/2939672.2939785.
5. *Makarov D.A.* Ispol'zovanie algoritma XGBoost dlya predskazaniya zaversheniya kursa obuchayushchimsya [Using the XGBoost algorithm to predict course completion for students], *StudNet*, 2021, No. 1. Available at: <https://cyberleninka.ru/article/n/ispolzovanie-algoritma-xgboost-dlya-predskazaniya-zaversheniya-kursa-obuchayushchimsya>.
6. *Syahputra Akhmal & Saputro Rujianto.* Application of the XGBoost Model with Hyperparameter Tuning for Industry Classification for Job Applicants, 2024. *Sinkron* 8, pp. 1920-1931. 10.33395/sinkron.v8i3.13840.
7. *Yan Q., Hu H.* XGBoost Machine Learning Model and Bond Pricing, *SSRN*, 2024. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4846755.
8. *Salazov A.V.* Sravnitel'nyy analiz effektivnosti XGboost i LSTM dlya prognozirovaniya volatil'nosti aktsiy Sberbanka [Comparative analysis of the effectiveness of XGBoost and LSTM for predicting the volatility of Sberbank shares], *Auditorium*, 2025, No. 2 (46). Available at: <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-effektivnosti-xgboost-i-lstm-dlya-prognozirovaniya-volatilnosti-aktsiy-sberbanka>.
9. *Bakumenko L.P., Vasil'eva N.S.* Sravnitel'nyy analiz prognosticheskikh modeley razvitiya kriptovalyutnogo rynka (LSTM i XGBOOST) na primere bitkoina [Comparative analysis of predictive models of the development of the cryptocurrency market (LSTM and XGBOOST) on the example of bitcoin], *Uchet i statistika* [Accounting and Statistics], 2023, No. 4 (72). Available at: <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-prognosticheskikh-modeley-razvitiya-kriptovalyutnogo-rynka-lstm-i-xgboost-na-primere-bitkoina>.

10. Zhang Lingyu & Bian Wenjie & Qu Wenyi & Tuo Liheng & Wang Yunhai. Time series forecast of sales volume based on XGBoost, *Journal of Physics: Conference Series*, 2021, 1873, 012067. 10.1088/1742-6596/1873/1/012067.
11. Thejovathi Murari and Rao, Dr. M V P Chandra Sekhara and Priyadharsini E.J. and Siddi Someshwar and Karthik B. and Abbas Syed Hauider. Optimizing Product Demand Forecasting with Hybrid Machine Learning and Time Series Models: A Comparative Analysis of XGBoost and SARIMA (November 15, 2024), *Proceedings of the 3rd International Conference on Optimization Techniques in the Field of Engineering (ICOFE-2024)*.
12. Nichani Rabia & Laid Gasmi & Laiche Nabil & Salheddine Kabou. Optimizing financial time series predictions with hybrid ARIMA, LSTM, and XGBoost Models, *Studies in Engineering and Exact Sciences*, 2024, 5. e11188. 10.54021/seesv5n2-582.
13. Odhiambo Stephen & Nyakundi Cornelious & Waititu Hellen. Developing a Hybrid ARIMA-XGBOOST Model for Analysing Mobile Money Transaction Data in Kenyam, *Asian Journal of Probability and Statistics*, 2024, 26, pp. 108-126. 10.9734/ajpas/2024/v26i10662.
14. Gond B. PHM-MAX-XGBoost-LSTM-SVM: A Hybrid Model for Time Series Forecasting – A Comparative Analysis with Traditional and Modern Approaches, 2025. 10.31224/4754.
15. Adil Mehdary, Abdellah Chehri, Abdeslam Jakimi, Rachid Saadane. Hyperparameter Optimization with Genetic Algorithms and XGBoost: A Step Forward in Smart Grid Fraud Detection, *Sensors*, 2024. Available at: <https://www.mdpi.com/1424-8220/24/4/1230>.
16. Lingeswari R., Brindha S. Online payments fraud prediction using optimized genetic algorithm based feature extraction and modified loss with XGboost algorithm for classification, *Swarm and Evolutionary Computation*, Vol. 95. Available at: <https://doi.org/10.1016/j.swevo.2025.101934>.
17. Kyung Keun Yun, Sang Won Yoon, Daehan Won. Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process, *Expert Systems with Applications*, 2021, Vol. 186, 115716. ISSN 0957-4174. Available at: <https://doi.org/10.1016/j.eswa.2021.115716>.
18. Yun K.K., et al. Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process, *Expert Systems with Applications*, 2021. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0957417421010988>.
19. A Guide on XGBoost hyperparameters tuning. Available at: <https://www.kaggle.com/code/prashant111/a-guide-on-xgboost-hyperparameters-tuning>.
20. DEAP documentation. Available at: <https://deap.readthedocs.io/en/master/>.
21. Shvager D. Tekhnicheskii analiz: polnyy kurs: perevod s angliyskogo [Technical analysis: a complete course]. Moscow: Al'pina Publisher, 2021, 804 p.

Хайруллина Лилия Эмитовна – Казанский (Приволжский) федеральный университет; e-mail: lxayrullina@yandex.ru; г. Казань, Россия; тел.: +79272464003; к.ф.-м.н.; доцент кафедры информационных систем.

Хакимов Зульфат Ниязович – Казанский (Приволжский) федеральный университет; e-mail: znkhakimov@stud.kpfu.ru; г. Казань, Россия; тел.: +79586249636; аспирант.

Галиев Данияр Ильнурович – Казанский (Приволжский) федеральный университет; e-mail: daniyargalievi@gmail.com; г. Казань, Россия; тел.: +79393902476; аспирант.

Хайруллина Азалия Ниязовна – Казанский (Приволжский) федеральный университет; e-mail: azhajrullina@gmail.com; г. Казань, Россия; тел.: +79586265959; студентка.

Khairullina Lilia Emitovna – Kazan (Volga Region) Federal University; e-mail: lxayrullina@yandex.ru; Kazan, Russia; phone: +79272464003; cand. of phys. and math. sc.; associate professor of the Department of Information Systems.

Khakimov Zulfat Niyazovich – Kazan (Volga Region) Federal University; e-mail: znkhakimov@stud.kpfu.ru; Kazan, Russia; phone: +79586249636; graduate student.

Galiev Daniyar Ilmurovich – Kazan (Volga Region) Federal University; e-mail: daniyargalievi@gmail.com; Kazan, Russia; phone: +79393902476; graduate student.

Khairullina Azalia Niyazovna – Kazan (Volga Region) Federal University; e-mail: azhajrullina@gmail.com; Kazan, Russia; phone: +79586265959; student.