

СТАТЬИ

УДК 004.85:004.896

DOI 10.17513/snt.40434

**ИССЛЕДОВАНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ
ДЛЯ ПРОГНОЗИРОВАНИЯ СОСТОЯНИЯ
ВИБРАЦИИ КОМПРЕССОРНОЙ УСТАНОВКИ**¹Вотякова Л.Р., ²Горячев А.С.¹ФГБОУ ВО «Казанский национальный исследовательский технологический университет»,
Казань, e-mail: vot5lil5rad5@gmail.com;²Нижнекамский химико-технологический институт (филиал)
ФГБОУ ВО «Казанский национальный исследовательский
технологический университет», Нижнекамск

Алгоритмы машинного обучения предоставляют широкий спектр возможностей для предиктивного анализа работы технологического оборудования. Они позволяют выявлять скрытые закономерности и прогнозировать будущие события на основе имеющихся данных. Целью данного исследования является анализ и сравнение методов машинного обучения для реализации задач прогнозирования состояния вибрации компрессорной установки. Каждый из методов имеет свои преимущества и ограничения, поэтому выбор конкретного алгоритма зависит от решаемой задачи применительно к характеристикам оборудования и данных. В данной работе исследованы методы машинного обучения для решения задачи регрессии: в случае прогнозирования одномерных рядов использованы методы авторегрессии и авторегрессии со скользящей средней; для прогнозирования многомерных рядов использованы методы линейной регрессии, деревьев решений, случайный лес и градиентный бустинг. Проведен сравнительный анализ методов машинного обучения. Из результатов анализа видно, что для задачи регрессии наилучшими методами являются ансамблевые методы, такие как случайный лес и XGBoost. Применение XGBoost значительно улучшает качество прогнозов, особенно при работе с большим объемом данных. Для моделей AutoReg и ARIMA необходим стационарный временной ряд для получения точных и интерпретируемых результатов. Важно экспериментировать и настраивать параметры для каждого конкретного временного ряда, учитывая такие факторы, как наличие сезонности, тренда и выбросов в данных. При выборе метода прогнозирования необходимо учитывать не только его математические характеристики, но и особенности конкретных данных, на которых он будет применен.

Ключевые слова: вибрация, компрессор, прогнозирование, авторегрессия, регрессия, деревья решений, случайный лес, градиентный бустинг

**RESEARCH OF MACHINE LEARNING ALGORITHMS
FOR VIBRATION CONDITION PREDICTION
COMPRESSOR UNIT**¹Votyakova L.R., ²Goryachev A.S.¹Kazan National Research Technological University, Kazan,
e-mail: vot5lil5rad5@gmail.com;²Nizhnekamsk Institute of Chemical Technology (branch)
of the Kazan National Research Technological University, Nizhnekamsk

Machine learning algorithms provide a wide range of possibilities for predictive analysis of technological equipment operation. They allow you to identify hidden patterns and predict future events based on available data. The purpose of this study is to analyze and compare machine learning methods for the realization of tasks of predicting the vibration state of a compressor unit. Each method has its advantages and limitations, so the choice of a specific algorithm depends on the problem being solved in relation to the characteristics of the equipment and data. In this paper, machine learning methods are studied to solve the regression problem: in the case of predicting one-dimensional series, autoregressive and moving average autoregressive methods are used; methods of linear regression, decision trees, random forest and gradient boosting are used to predict multidimensional series. A comparative analysis of machine learning methods has been carried out. It can be seen from the analysis results that ensemble methods such as random forest and XGBoost are the best methods for the regression problem. Using XGBoost significantly improves the quality of predictions, especially when working with large amounts of data. The AutoReg and ARIMA models require a stationary time series to produce accurate and interpretable results. It is important to experiment and adjust the parameters for each specific time series, taking into account factors such as seasonality, trend, and outliers in the data. When choosing a forecasting method, it is necessary to take into account not only its mathematical characteristics, but also the specifics of the specific data on which it will be applied.

Keywords: vibration, compressor, forecasting, autoregression, regression, decision trees, random forest, gradient boosting

Введение

Одна из главных тенденций современного производства – это концепция, которая предполагает цифровизацию производства, внедрение в него предиктивной (прогнозной) аналитики, работа с большими данными, создание цифровых двойников [1–3]. Данная работа посвящена анализу применения методов машинного обучения для решения задач предиктивного анализа при работе технологического оборудования, задействованного в технологическом процессе переработки и производства нефтепродуктов [4–6]. Своевременное предсказание эффективной работы оборудования и выявление аномалий в его работе существенно увеличит экономическую эффективность и безопасность производства, так как уменьшит вероятность нарушения в обеспечении проведения непрерывных технологических процессов.

Цель исследования – провести анализ и сравнение методов машинного обучения для реализации задач прогнозирования состояния вибрации компрессорной установки.

Материалы и методы исследования

Для построения моделей был использован язык Python, среда разработки Jupyter Notebooks. В разработке программ по исследованию алгоритмов машинного обучения использование библиотек Python обусловлено их функциональными возможностями и особенностями, которые позволяют эффективно обрабатывать и анализировать данные [7, 8]. NumPy служит основой для численных вычислений, Pandas упрощает манипуляцию табличными данными, Matplotlib и Seaborn обеспечивают возможности визуализации данных, Statsmodels позволяет оценивать и анализировать различные статистические модели, Scikit-learn предоставляет разнообразные алгоритмы машинного обучения.

Объект исследования – компрессорная установка модели С-85.6, управляемая с помощью программируемого логического контроллера, данные с которого отправляются в программный пакет для диспетчерского управления и сбора данных, который представляет собой распределенную систему управления CENTUM VP от компании Yokogawa Electric Corporation, предназначенной для мониторинга и управления технологическим процессом. Система резервирует все компоненты системы, имеет распределенную архитектуру «клиент-сервер», хранит данные технологической линии, которые и понадобятся для исследования. Для проведения анализа был использован

табличный формат данных – .csv. Для прогнозирования технического состояния компрессора методом машинного обучения проведены исследования исторических данных значений с датчиков вибрации, давления, падения штока (износ сальника) поршневой компрессорной установки. Полученные для анализа значения с датчика вибрации снимаются с периодичностью в 1 мин. Количество наблюдений 43 165 строк. В исследовании выполняется прогнозирование вибрации корпуса компрессора (датчик поз.1V-VIA1967.PV).

Для совершения оперативных прогнозов с помощью авторегрессионных моделей временной ряд должен быть стационарным [9–11]. По стационарному ряду строится прогноз, так как его будущие статистические характеристики не отличаются от наблюдаемых текущих. Проверка ряда на стационарность показала, что полученное значение p -value 0,018859 меньше 0,05 – значит, ряд стационарен. Распределение целевой переменной (вибрация) показало, что оно близко к нормальному, с легким перекосом вправо, в основном значения варьируются в диапазоне от 7 до 10 мм/с. Корреляционная матрица показала высокую корреляцию между переменными давления ($P_{\text{нагнет}}$ и $P_{\text{всас}}$) и целевой переменной и низкую корреляцию между сдвигами штоков и вибрацией. Наиболее коррелирующие признаки с вибрацией: $P_{\text{нагнет_4}}$, $P_{\text{нагнет_3}}$, $P_{\text{всас_4}}$, $P_{\text{всас_3}}$, $P_{\text{нагнет_2}}$. Визуализация показала линейную зависимость между вибрацией и давлениями, что подтверждает значимость этих признаков для модели. В данном исследовании, в случае прогнозирования одномерных рядов, использованы методы авторегрессии и авторегрессии со скользящей средней [12]; для прогнозирования многомерных рядов использованы методы линейной регрессии, деревьев решений [13], случайный лес [14] и градиентный бустинг [15]. Во всех моделях обучающая выборка включает данные, использованные для обучения модели прогнозирования. Тестовая выборка содержит данные, которые не использовались при обучении модели, но применялись для проверки ее точности. Прогон представляет собой данные, обработанные моделью после завершения обучения. Доверительный интервал указывает на уровень уверенности в прогнозах, показывая верхнюю и нижнюю границы возможных значений.

Результаты исследования и их обсуждение

Авторегрессионные модели (AR-модели) используют прошлые значения ряда для прогнозирования его будущих зна-

чений. Эта модель предполагает, что текущее значение ряда зависит от его предыдущих значений. В данном исследовании были использованы модели AutoReg и ARIMA. На графике, представленном на рис. 1, отображены прогнозируемые значения вибрации компрессора. Из графика видно, что модель прогнозирует вибрацию компрессора только в краткосрочном периоде.

Комбинация авторегрессионной модели порядка p и модели скользящего среднего порядка q дает смешанную авторегрессионную модель скользящего среднего (ARMA (p, q)). ARMA, как и AR модель, используется только для стационарных рядов. Для работы с нестационарными рядами формируется модель ARIMA, тем, что в модели ARMA включается оператор взятия простых разностей порядка d .

На графике, представленном на рис. 1, отображены прогнозируемые значения вибрации компрессора с использованием модели ARIMA. Результат показал, что модель прогнозирует вибрацию компрессора также в краткосрочном периоде.

На рис. 2 представлены прогнозы модели, обученной на методе линейной регрессии. Также на рис. 2 приведено огра-

ниченное отображение данных для улучшения читаемости. Значения полученных метрик при выполнении программы приведены в таблице сравнения с другими алгоритмами.

Деревья решений (Decision Trees) – один из наиболее эффективных инструментов интеллектуального анализа данных и предсказательной аналитики. На рис. 3 представлен прогноз модели, обученной на методе деревьев решений. Значения полученных метрик приведены в таблице анализа (таблица).

При прогнозировании алгоритм «Случайный лес» (Random Forest) объединяет результаты всех деревьев либо путем голосования (для задач классификации), либо путем усреднения (для задач регрессии). Этот процесс совместного принятия решений, основанный на выводах нескольких деревьев, обеспечивает стабильные и точные результаты. Случайные леса уменьшают переобучение и обеспечивают надежные прогнозы в различных условиях. На рис. 3 представлен прогноз модели, обученной на методе «Случайный лес». Значения полученных метрик приведены в таблице анализа.

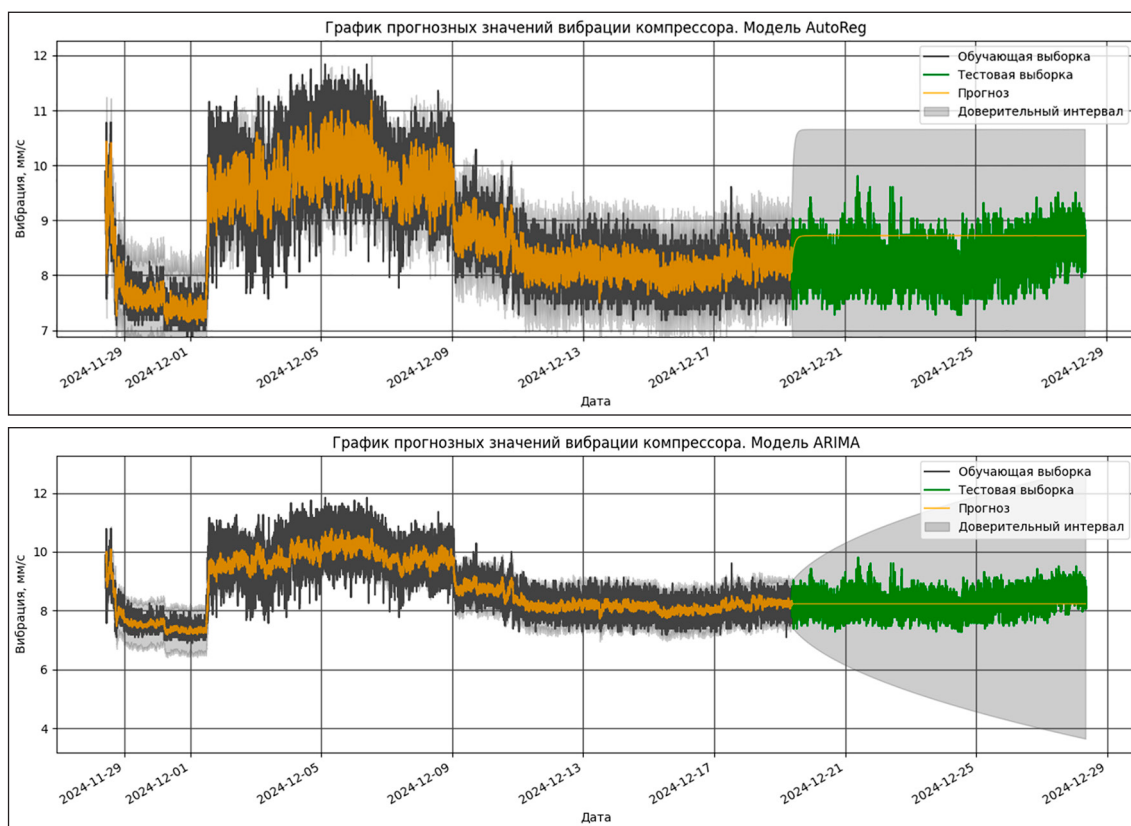


Рис. 1. График предсказания с помощью модели AR и ARIMA
Источник: составлено авторами по результатам данного исследования

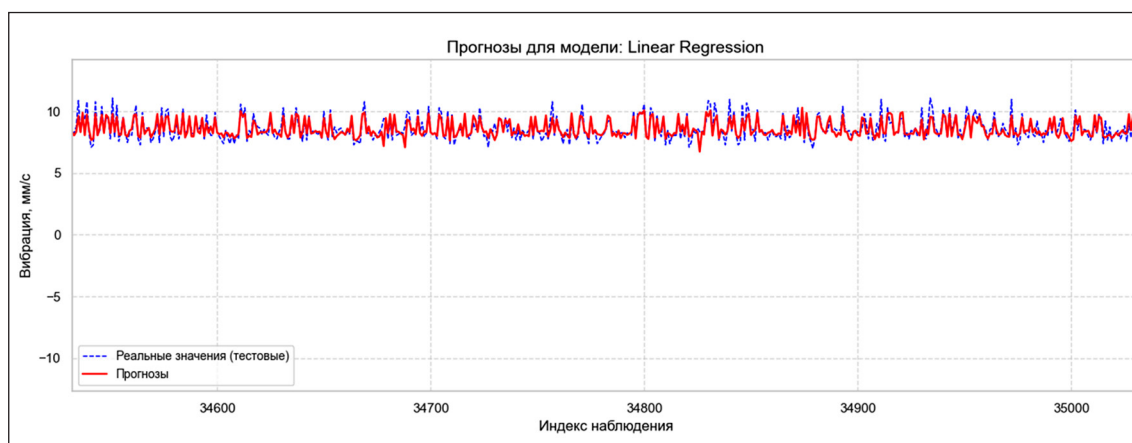


Рис. 2. Графическое представление прогноза модели линейной регрессии
Источник: составлено авторами по результатам данного исследования

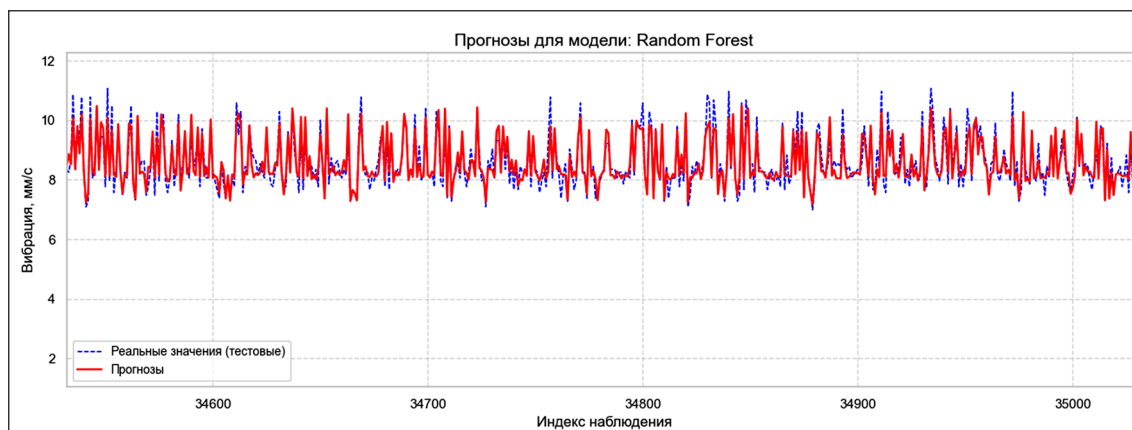
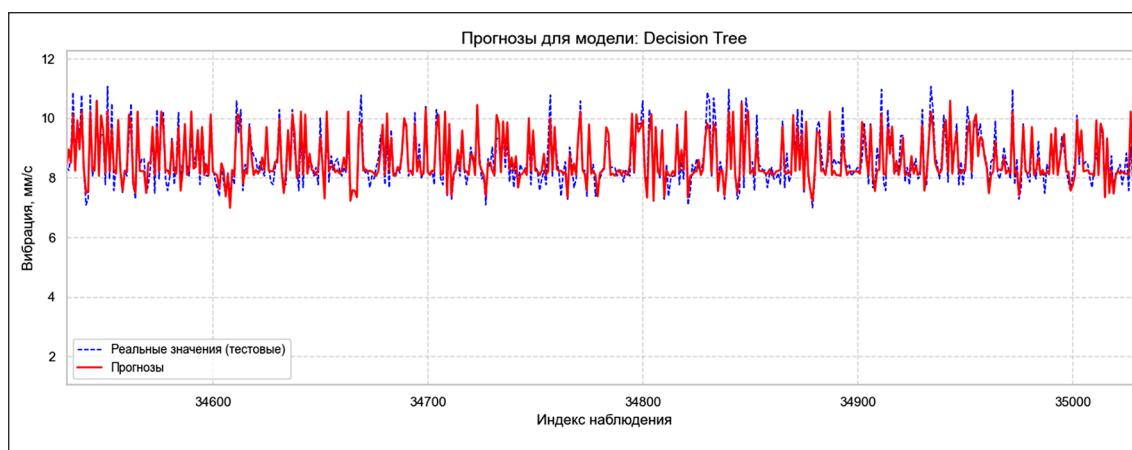


Рис. 3. Графическое представление прогноза модели дерева решений и Random Forest
Источник: составлено авторами по результатам данного исследования

В рамках исследования была использована модель градиентного бустинга из библиотеки XGBoost, которая предлагает высокоэффективную реализацию градиентного бустинга на основе решающих деревьев. Прогнозирующая модель

также состоит из ансамбля слабых предсказывающих моделей – деревьев решений. Подготовка данных для предсказаний с использованием XGBoost аналогична подготовке данных для модели линейной регрессии.

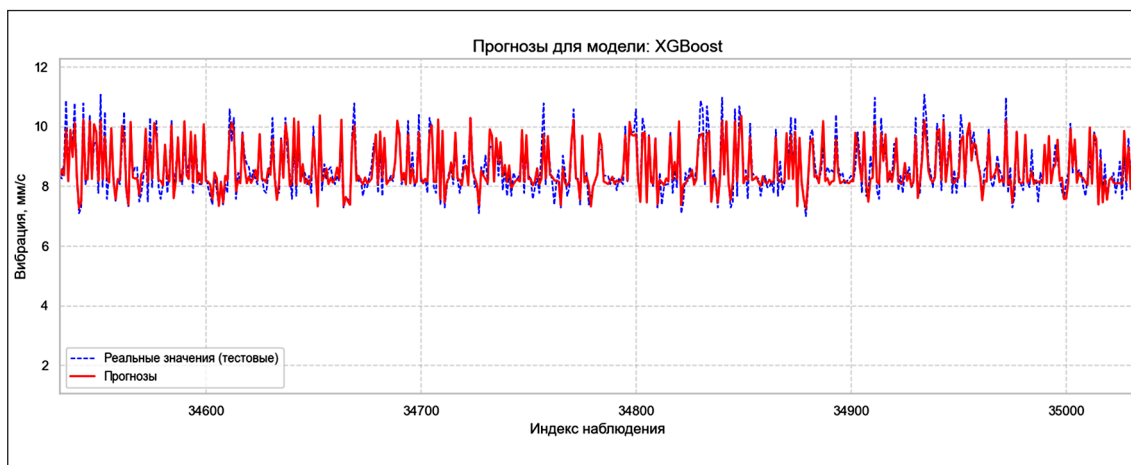


Рис. 4. Графическое представление прогноза модели XGBoost
 Источник: составлено авторами по результатам данного исследования

Метрики качества моделей

	Model	R ²	MSE	MAPE
1	AutoReg	0,980000	0,347148	6,355062
2	ARIMA	0,818801	0,138474	3,233626
3	Linear Regression	0,606385	0,309984	4,675680
4	Decision Trees	0,801051	0,156679	3,408400
5	Random Forest	0,827071	0,136187	3,223779
6	XGBoost	0,823468	0,139025	3,267489

Источник: составлено авторами на основе полученных данных в ходе исследования.

Главная ценность библиотеки XGBoost заключается в ее высокопроизводительной реализации. Благодаря различным оптимизациям, таким как эффективное управление пропущенными значениями, поиск порога только среди персентилей, оптимизированная работа с кэшированием и поддержка распределенного обучения, производительность возрастает в десятки или даже сотни раз по сравнению с примитивными реализациями. На рис. 4 представлены прогнозы модели, обученной на методе градиентного бустинга. Значения полученных метрик приведены в таблице анализа.

Информация об исследованных моделях с полученными метриками (коэффициент детерминации, среднеквадратичная ошибка, средняя абсолютная ошибка в процентах) их качества представлена в сводной таблице.

В результате проведенного статистического анализа были получены значимые метрики качества аппроксимации. Количественная оценка авторегрессионной модели (AutoReg) свидетельствует

о том, что вариабельность целевой переменной на 98 % обусловлена изменениями предикторной переменной вибрации, что подтверждает высокую прогностическую способность модели. Дополнительно проведенный анализ среднеквадратичной ошибки (MSE), составившей 0,35, указывает на незначительную дисперсию отклонений наблюдаемых значений от расчетных. Данный факт свидетельствует о высоком качестве подгонки модели к исходному набору данных.

Коэффициент детерминации интегрированной модели авторегрессии скользящего среднего (ARIMA) указывает на то, что модель объясняет примерно 82 % вариации данных. Это довольно хороший результат, свидетельствующий о том, что модель достаточно точно описывает динамику временного ряда. Низкое значение MSE указывает на умеренную ошибку прогноза, но при этом она все еще остается приемлемо низкой. В среднем прогнозы отличаются от истинных значений на 3,23 %, это очень низкий уровень ошибки. Совокуп-

ный анализ всех диагностических метрик позволяет сделать обоснованный вывод о высокой эффективности модели ARIMA в контексте прогнозирования временных рядов. Полученные результаты подтверждают ее пригодность для решения задач прогностического моделирования с обеспечением высокого уровня точности предсказаний.

Линейная регрессия (Linear Regression) показала наихудшее качество предсказания с метрикой $R^2 = 0,606$. Это свидетельствует о том, что зависимость между признаками и целевой переменной нелинейная, и простая линейная модель не может адекватно описать данные.

Проведенное сравнительное исследование эффективности алгоритмов машинного обучения выявило существенные преимущества модели дерева решений (Decision Trees) над классической линейной регрессией в контексте качества аппроксимации данных. Коэффициент детерминации ($R^2 = 0,801$) демонстрирует значительное превосходство данной модели, что свидетельствует о ее способности эффективно выявлять и моделировать нелинейные взаимосвязи в исследуемом наборе данных. Однако относительно высокий MSE (0,156) может указывать на потенциальную проблему переобучения модели. Данный факт требует дополнительного внимания при дальнейшей настройке гиперпараметров и оптимизации архитектуры модели для достижения оптимального баланса между способностью к обобщению и точностью аппроксимации.

Модель случайного леса (Random Forest) показала лучшее качество предсказания с $R^2 = 0,827$ и снижением ошибки MSE = 0,136. Использование ансамбля деревьев помогло уменьшить переобучение и повысить обобщающую способность модели.

Градиентный бустинг (XGBoost) показал близкое качество к случайному лесу с $R^2 = 0,823$ и MSE = 0,139. Модель также хорошо справилась с нелинейностями и сложными зависимостями, а встроенные методы регуляризации уменьшили вероятность переобучения.

Заключение

В результате проведенного исследования собраны данные в виде временных рядов для разработки предиктивной модели, которые отображают основные параметры в работе компрессора. Был проведен сравнительный анализ методов машинного обучения для прогнозирования состояния вибрации компрессорной установки.

Для задачи регрессии одномерных рядов лучше всего подошли модели авторегрессии. Для моделей AutoReg и ARIMA необходим стационарный временной ряд для получения точных и интерпретируемых результатов. При этом модель AutoReg имеет более высокие показатели качества прогнозирования. Для задачи регрессии многомерных рядов лучше всего подходят ансамблевые методы, такие как случайный лес и XGBoost. Случайный лес можно рекомендовать как основную модель из-за чуть более высокого R^2 и простоты настройки гиперпараметров. XGBoost может быть полезен для дальнейшего улучшения качества предсказаний, особенно на большом количестве данных. Важно экспериментировать и настраивать параметры для каждого конкретного временного ряда, учитывая такие факторы, как наличие сезонности, тренда и выбросов в данных. При выборе метода прогнозирования необходимо учитывать не только его математические характеристики, но и особенности конкретных данных, на которых он будет применен.

Список литературы

1. Тарасов И.В. Индустрия 4.0: Понятие, концепция, тенденции развития // Стратегии бизнеса. 2018. № 5. С. 43–49. DOI: 10.17747/2311-7184-2018-5-43-49.
2. Афанасьев А.А. Индустрия 4.0: к вопросу о перспективах цифровой трансформации промышленности в России // Вопросы инновационной экономики. 2023. Т. 13. № 3. С. 1427–1446. DOI: 10.18334/vinec.13.3.117880.
3. Шваб Клаус. Четвертая промышленная революция / Пер. с английского. М.: Эксмо, 2023. 208 с. [Электронный ресурс]. URL: <https://content.img-gorod.ru/pim/products/images/35/5c/0195ff2d-5d36-7d39-9447-15acba3c355c.pdf> (дата обращения: 29.04.2025).
4. Шигабутдинов А.К., Пресняков В.В., Ананд Субраманиан В., Новиков М.А., Храмов А.А., Минигулов Р.Р., Идрисов М.Р., Коновнин А.А., Уразайкин А.С., Фхунов Р.Н. Физико-химические характеристики остатка комбинированного гидрокрекинга гудрона, получаемого, при промышленной эксплуатации установки VCC на КГПТО АО «ТАИФ-НК» // Вестник технологического университета. 2022. Т. 25. № 7. С. 55–59. URL: <https://www.knitu.ru/servlet/contentblob?id=428340> (дата обращения: 29.04.2025).
5. Lathong K., Wisaeng K. An innovative hybrid machine learning techniques for predicting construction cost estimates // International Journal for Computational Civil and Structural Engineering. 2024. Vol. 20, Is. 3. P. 69–83. DOI: 10.22337/2587-9618-2024-20-3-69-83.
6. Шлыков С.В. Применение методов машинного обучения для автоматизации процессов в нефтегазовой отрасли // Транспорт и хранение нефтепродуктов и углеводородного сырья. 2023. № 2. С. 46–53. DOI: 10.24412/0131-4270-2023-2-46-53.
7. Вотякова Л.Р., Зотин А.В. Исследование идентификации и реализации многопараметрического контроллера на языке Python // Современные наукоемкие технологии. 2024. № 7. С. 16–21. DOI: 10.17513/snt.40079.
8. Вотякова Л.Р. Программная реализация интеллектуальной системы поддержки принятия решений для подбора

персонала в IT-сфере // Научно-технический вестник Поволжья. 2024. № 5. С. 89–91. URL: <https://ntvprt.ru/ru/archive-vypuskov> (дата обращения: 01.06.2025).

9. Кувайскова Ю.Е. Применение ансамблевых методов машинного обучения для прогнозирования технического состояния объекта // Известия Самарского научного центра Российской академии наук. 2021. Т. 23. № 1 (99). С. 111–114. DOI: 10.37313/1990-5378-2021-23-1-111-114.

10. Фирюлина М.А., Каширина И.Л. Описание процесса прогнозирования проблемных состояний с применением ансамблевых методов машинного обучения // Инженерный вестник Дона. 2022. № 4 (88). С. 34–46. URL: <http://www.ivdon.ru/ru/magazine/archive/n4y2022/7589> (дата обращения: 01.06.2025).

11. Zheng Zhang, Zhihui Lai, Yong Xu, Ling Shao, Jian Wu, Guosen Xie. Discriminative elastic-net regularized linear regression // IEEE Transactions on Image Processing. 2017. Т. 26. № 3. С. 1466–1481. DOI: 10.1109/TIP.2017.2651396.

12. Альчаков В.В., Крамарь В.А. Оценка методов машинного обучения для прогнозирования сезонных временных рядов // Известия ЮФУ. Технические науки. 2023. № 2 (232). С. 250–263. DOI: 10.18522/2311-3103-2023-2-250-263.

13. Yang Ch., Wu G., Ding K., Shi T., Li Q., Wang J. Improving land use/land cover classification by integrating pixel unmixing and decision tree methods // Remote Sensing. 2017. Т. 9. № 12. С. 1222. DOI: 10.3390/rs9121222.

14. Krauss C., Do X.A., Huck N. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500 // European Journal of Operational Research. 2016. Т. 259. № 2. С. 689–702. DOI: 10.1016/j.ejor.2016.10.031.

15. Lamara B., Notton G., Fouilloy A., Voyant C., Rabah D. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components // Renewable energy. 2019. Т. 132. С. 871–884. DOI: 10.1016/j.renene.2018.08.044.