

МАТЕМАТИЧЕСКИЕ МЕТОДЫ И ИНСТРУМЕНТЫ ЭКОНОМИЧЕСКОГО АНАЛИЗА

ИССЛЕДОВАТЕЛЬСКАЯ СТАТЬЯ

<https://doi.org/10.31063/AlterEconomics/2025.22-3.6>

УДК 336.6

JEL C53, E37, L16



Применение методов машинного обучения для прогнозирования нефтяных котировок¹

Варвара В. НАЗАРОВА ¹⁾ , Борис А. ЛОДЯГИН ²⁾ , Фёдор А. КРУГЛОВ ³⁾,
Арсений В. КРУГЛОВ ⁴⁾

^{1, 2, 3, 4)} Национальный исследовательский университет «Высшая школа экономики»,
г. Санкт-Петербург, Российская Федерация

Для цитирования: Назарова, В. В., Лодягин, Б. А., Круглов, Ф. А., Круглов, А. В. (2025). Применение методов машинного обучения для прогнозирования нефтяных котировок. *AlterEconomics*, 22(3), 482–502. <https://doi.org/10.31063/AlterEconomics/2025.22-3.6>

Аннотация. Статья посвящена сравнительному анализу точности краткосрочного прогнозирования цен на нефть с использованием традиционных статистических моделей (ARIMA, SARIMAX) и модели глубинного обучения LSTM. Актуальность исследования обусловлена высокой волатильностью нефтяного рынка и необходимостью повышения точности прогнозов в условиях нестабильной макроэкономической среды, где нефтяные котировки зависят не только от собственной динамики, но и от внешних факторов. Исследование восполняет недостаток работ, напрямую сравнивающих модели статистического и нейросетевого прогнозирования при единых условиях, а также оценивающих влияние конкретных экзогенных переменных. В качестве прогнозируемой переменной рассматривается цена на нефть марки WTI. В выборку включены ежедневные данные за 2015–2019 годы — период с относительно стабильными макроэкономическими условиями без выраженных экзогенных кризисов. В модели SARIMAX и LSTM использованы экзогенные макроэкономические и финансовые переменные-предикторы, отражающие состояние фондового, финансового и сырьевого рынков США: индексы Dow Jones, Wilshire 5000, DXY, танкерный индекс BDTI, спред Brent–WTI, а также объемы добычи, переработки, запасов и потребления нефти. Результаты показывают, что модель LSTM с включением индекса Dow Jones демонстрирует наименьшую среднеквадратичную ошибку прогноза (RMSE = 1,5), что подтверждает превосходство методов машинного обучения над авторегрессионными подходами в рассмотренных условиях и при наличии нелинейных взаимосвязей. Включение экзогенных факторов приводит к улучшению точности прогноза как для SARIMAX, так и для LSTM, однако эффект для последней модели оказывается умеренным. Полученные результаты позволяют заключить, что применение нейросетевых моделей с учетом рыночного контекста повышает точность краткосрочного прогнозирования. В качестве направлений будущих исследований предлагается тестирование других архитектур машинного обучения (Random Forest, XGBoost), использование альтернативных наборов переменных и адаптация моделей к условиям высокой волатильности (*regime-switching models*).

Ключевые слова: прогнозирование цены на нефть, авторегрессионные модели, машинное обучение, ARIMA, SARIMAX, LSTM, нефтяной рынок США

¹ © Назарова В. В., Лодягин Б. А., Круглов Ф. А., Круглов А. В. Текст. 2025.

RESEARCH ARTICLE

Application of AI for Oil Price Forecasting

Varvara V. NAZAROVA¹⁾  , Boris A. LODIAGIN²⁾ , Fedor A. KRUGLOV³⁾,
Arseniy V. KRUGLOV⁴⁾

^{1, 2, 3, 4)} National Research University “Higher School of Economics”, St. Petersburg, Russian Federation

For citation: Nazarova, V. V., Lodiagin, B. A., Kruglov, F. A., & Kruglov, A. V. (2025). Application of AI for Oil Price Forecasting. *AlterEconomics*, 22(3), 482–502. <https://doi.org/10.31063/AlterEconomics/2025.22-3.6>

Abstract. This paper examines methods for forecasting oil prices, comparing traditional autoregressive models (ARIMA, SARIMAX) with machine learning approaches (LSTM). The target variable is the price of WTI crude oil. The dataset covers 2015–2019 and includes both WTI price data and a set of exogenous variables: the Wilshire 5000, Dow Jones, and DXY indices; the Baltic Dirty Tanker Index (BDTI); the Brent–WTI price spread; volumes of crude oil produced and refined in the US; the number of active drilling rigs; US crude oil inventories; US petroleum product consumption; US net oil imports; and the number of days US crude oil inventories can cover demand, excluding the strategic petroleum reserve. Forecasts were generated over a five-day horizon. The results indicate that machine learning outperforms autoregressive models. The LSTM model incorporating the Dow Jones index as an exogenous variable showed the best performance (RMSE: 1.5). Autoregressive models performed less effectively, with RMSE values of 1.668 for ARIMA and 1.51 for SARIMAX. Even the LSTM model without exogenous variables (RMSE: 1.51) outperformed ARIMA, underscoring the relative advantage of machine learning in short-term oil price forecasting.

Keywords: oil price forecasting, autoregressive models, machine learning, ARIMA, SARIMAX, LSTM, US oil market

1. Введение

Нефть занимает центральное место среди ключевых ресурсов современной цивилизации, играя важную роль в экономике, промышленности и повседневной жизни людей. Как основной источник энергии и сырье для производства широкого спектра товаров — от топлива до пластмасс — нефть оказывает значительное влияние на глобальную экономику. Ее стоимость непосредственно влияет на многие аспекты человеческой деятельности, начиная от уровня инфляции (Sek et al., 2015) и заканчивая стоимостью транспортных услуг (Gross et al., 2012).

Существует ряд работ, рассматривающих влияние цены на нефть на мировую экономику, рынки стран и отдельных регионов, отрасли промышленности. Например, Р.А. Ратти и соавторы (Ratti & Vespignani, 2014) приходят к выводу, что существует тесная взаимосвязь мировых цен на нефть и глобальных финансовых условий с основными экономическими показателями, такими как реальный объем производства и инфляция.

Помимо влияния на мировую экономику, цены на нефть способны оказывать существенное экономическое влияние на экономики отдельных стран. К их числу можно отнести как страны, зависимые от импорта нефти, так и страны, занимающиеся её экспортом. Для последних это влияние иногда называют термином «ресурсное проклятие». Доходы таких государств и их ВВП часто зависят от экспорта нефтяных ресурсов, потому что колебание цен на этот ресурс способно оказывать большое влияние на их экономики, подчас дестабилизируя их. Например, в статье Т.Б. Фам и соавторов (Pham & Sala, 2019) рассматривается критическое влияние изменения цен на нефть на экономику Вьетнама — страны, являющейся чистым экспортером нефти, одновременно с этим импортирующим до 70 % продуктов нефтепереработки. Есть и другие работы, изучающие влияние цен на сырую нефть на малые экономики.

Цены на нефть имеют решающее значение и для больших экономик, таких как китайская экономика, не имеющая достаточного количества нефтяных ресурсов. Согласно данным Главного таможенного управления, КНР в 2023 г. импортировал до 75 % от всей потребляемой страной нефти. Исследователи указывают на негативную краткосрочную связь между ценой на нефть и объемом производства в Китае: увеличение цены на нефть на 1 % снижает выпуск продукции примерно на 0,38 % (Tang et al., 2010). Этот факт свидетельствует о прямом влиянии изменений цен на нефть на китайскую экономику. Существует ряд исследований, также подтверждающих наличие этого влияния для стран, зависимых от импорта нефтепродуктов.

В связи с большим значением нефти для мировой экономики точное прогнозирование цен на нефть стало одной из важнейших задач для правительств, компаний и инвесторов.

Инвесторы и трейдеры могут использовать прогнозирование цены на нефть для осуществления торговли и извлечения прибыли. Компании за счет прогноза цены на нефть могут рассчитывать изменение транспортных расходов или изменение стоимости сырья и материалов, производимых из нефти. То есть прогнозы стоимости нефти могут быть необходимы для построения финансовых планов. Прогнозирование цен на нефть особенно важно для государств-экспортеров нефти, т. к. их государственный бюджет рассчитывается на основе прогноза цены на нефть. Для государств-импортеров нефти её цена также важный показатель, поскольку на его основе можно определить ожидаемый объем производства в стране. Область прогнозирования имеет долгую историю, существуют сотни методов прогнозирования. Их можно подразделить на несколько видов: статистические модели и модели, использующие машинное обучение. Классификация методов представлена в таблице 1.

Таким образом, цель исследования — сравнение статистических прогнозирования и моделей с машинным обучением, выявление модели, наиболее точно предсказывающей цену на нефть.

Статистические модели иногда называют традиционными. Они включают в себя модели ARIMA, SARIMA, GARCH, VAR, VECM и др. Одни модели хорошо ра-



Рис. 1. Динамика цены на нефть марки WTI с 1983 по 2024 гг. (источник: сост. по: Infotables. (n.d.). Цена НЕФТИ марки Brent WTI Urals за баррель по годам (Таблица график). <https://infotables.ru/statistika/95-tseny-tarify/1325-tsena-na-neft-tablitsa#wti> (дата обращения: 15.05.2025))

Fig. 1. Dynamics of WTI Oil Prices from 1983 to 2024

Таблица 1

Классификация методов прогнозирования цены на нефть

Table 1

Classification of Oil Price Forecasting Methods

Статистические модели	Модели машинного обучения
ARIMA (Ariyanti & Yusnitasari, 2023)	LSTM (Natarajan & Ashok, 2018)
SARIMA (Wang, 2023)	Bi-LSTM (Vo et al., 2020)
SARIMAX (Agrawal et al., 2024)	EEMD-LSTM (Wu et al., 2019)
GARCH (Mohammadi & Su, 2010)	LSTM-CNN (Li et al., 2021)
VECM (Ratti & Vespignani, 2014)	ANN (Gupta & Nigam, 2020)

Источник: сост. авторами.

ботаю́т с линейными зависимостями, например, ARIMA. Другие могут учитывать сезонность (SARIMA). Третьи способны моделировать волатильность (GARCH). Однако эти модели работают только для стационарных рядов и не учитывают экзогенные переменные. Модель SARIMAX способна учитывать экзогенные факторы при прогнозировании, однако также плохо справляется с нелинейными зависимостями. Регрессионная модель VAR также работает только со стационарными рядами, но, в отличие от ARIMA, учитывает взаимное влияние нескольких переменных. Кроме того, для данной регрессионной модели существует модификация VARX, способная включать в модель экзогенные переменные. Подробнее некоторые из указанных моделей рассматриваются в разделах Обзор литературы и Методы.

Помимо традиционных моделей прогнозирования существуют модели с использованием машинного обучения. Данные модели более сложные в реализации, менее прозрачные в работе с данными, однако этот вид моделей решает целый фронт задач: позволяет работать с нелинейными зависимостями, улавливает сложные паттерны в данных. К моделям этого вида можно отнести SVR, Random Forest, XGBoost, LSTM с её модификациями и др. Недостатком моделей с машинным обучением является их сложность, долгое обучение, требовательность к объёму вычислительных мощностей. Часть этих моделей требуют постоянной перенастройки. Однако все перечисленные затруднения позволяют качественно повысить точность прогнозирования.

Большинство исследований на тему прогнозирования занимается разработкой новых модификаций моделей и новых моделей. Одна из целей данной работы — показать, что экзогенные переменные способны позитивно влиять на качество прогноза для любых видов моделей. В работе используется широкий набор факторов, способных влиять на цену нефти: индекс Wilshire 5000, индекс доллара, индекс Dow Jones, чистый импорт нефти в США, потребление нефтепродуктов в США, количество активных буровых установок в США, индекс танкерных перевозок BDTI, спред Brent–WTI, объём добываемой нефти в США, количество дней, в течение которых запасы сырой нефти в США могут покрывать спрос, исключая стратегический резерв нефти (SPR), объём перерабатываемой нефти в США, запасы нефти в США. В процессе настройки моделей выделяются только наиболее значимые экзогенные переменные.

На основе этих данных строятся следующие модели: ARIMA, SARIMAX и LSTM. Для последней модели создаются две модификации: только с историческими данными о цене на нефть марки WTI и с включением вышеперечисленных экзогенных переменных.

Классификация моделей прогнозирования

Одной из задач данного исследования является сравнение моделей этих двух видов с точки зрения точности даваемого ими прогноза. В таблице 2 представлен краткий обзор литературы, использующей различные модели прогнозирования и сравнивающей их.

Таблица 2

Классификация исследований о прогнозировании цены на нефть

Table 2

Classification of Studies on Oil Price Forecasting

Статья	Тема исследования	Заключение
Natarajan & Ashok (2018)	Исследование применяет нейронные сети (LSTM) для многомерного прогнозирования спотовых цен на нефть с учетом макроэкономических и рыночных факторов	Модель LSTM показала более высокую точность прогнозов по сравнению с традиционными методами (ARIMA, GARCH) благодаря способности улавливать сложные нелинейные зависимости в данных
Li et al. (2019)	Прогнозирование цен на нефть с использованием методов глубокого обучения (LSTM и word2vec) на основе анализа текстовых данных (новостей, отчетов)	Текстовая модель на основе LSTM превзошла традиционные подходы (ARIMA, GARCH) за счет способности извлекать смысловые паттерны из новостей, что особенно эффективно в периоды высокой волатильности
Gupta & Nigam (2020)	Прогнозирование цен на сырую нефть с использованием искусственной нейронной сети (ИНС) и сравнение её эффективности с традиционными статистическими методами	ИНС продемонстрировала более высокую точность прогнозирования по сравнению с линейными моделями (ARIMA) благодаря способности анализировать сложные нелинейные зависимости в данных, особенно в условиях высокой волатильности рынка
Li et al. (2021)	Прогнозирование доходности и волатильности фьючерсов на нефть с использованием методов глубокого обучения (LSTM, CNN) на основе декомпозиции данных и анализа тональности новостей	Гибридная модель (ICEEMDAN + LSTM-CNN) значительно превзошла традиционные подходы (GARCH, HAR) благодаря способности выделять сложные временные зависимости и учитывать влияние новостного сентимента, особенно в периоды рыночных шоков
Morgado-Dias et al. (2023)	Систематический обзор применения анализа тональности (<i>sentiment analysis</i>) в нефтяном секторе, включая методы обработки текстовых данных и их влияние на прогнозирование цен	Анализ тональности новостей и социальных медиа значительно улучшает точность прогнозов в нефтяном секторе, особенно при комбинации с машинным обучением (LSTM, BERT), превосходя традиционные экономические модели
Kristjanpoller & Minutolo (2016)	Прогнозирование волатильности цен на нефть с использованием гибридной модели, сочетающей искусственную нейронную сеть (ИНС) и GARCH-подход	Гибридная модель ANN-GARCH продемонстрировала более высокую точность в прогнозировании волатильности по сравнению со стандартными GARCH-моделями благодаря способности ИНС улавливать нелинейные зависимости и сложные паттерны в данных
Mohammadi & Su (2010)	Анализ динамики цен на сырую нефть на международных рынках с применением комбинированных моделей ARIMA-GARCH для учета временных зависимостей и волатильности	Модель ARIMA-GARCH показала высокую эффективность в описании динамики нефтяных цен, превосходя простые ARIMA-модели за счет учета кластеризации волатильности и долгосрочных ценовых трендов

Окончание табл. 2 на след. стр.

Окончание табл. 2

Статья	Тема исследования	Заключение
Wu et al. (2019)	Прогнозирование цен на сырую нефть с использованием улучшенного метода декомпозиции (EEMD) и долгосрочной краткосрочной памяти (LSTM) для учета сложных временных паттернов	Комбинированная модель EEMD-LSTM значительно превзошла традиционные методы (ARIMA, GARCH) и простые LSTM-модели благодаря способности EEMD выделять скрытые компоненты данных, что улучшило точность прогнозирования
Li et al. (2019)	Прогнозирование цен на нефть марки Brent с использованием двунаправленной LSTM-сети (Bi-LSTM), учитывающей временные зависимости в данных	Модель Bi-LSTM продемонстрировала более высокую точность прогнозов по сравнению с обычными LSTM и традиционными методами (ARIMA, SVR) благодаря способности анализировать временные ряды в обоих направлениях (прямом и обратном), что улучшило учет долгосрочных зависимостей
Ariyanti & Yusnitasari (2023)	Сравнение эффективности моделей ARIMA и SARIMA для прогнозирования цен на сырую нефть с учетом сезонных и несезонных компонент временных рядов	Модель SARIMA показала более высокую точность прогнозирования по сравнению с классической ARIMA за счет учета сезонных паттернов, что особенно важно для нефтяных цен с их циклической динамикой
Wang (2023)	Прогнозирование цен на нефть с использованием усовершенствованной SARIMA-модели, модифицированной для повышения точности за счет оптимизации параметров и учета внешних факторов	Улучшенная SARIMA-модель превзошла стандартные ARIMA и SARIMA по точности прогнозирования благодаря адаптивному подбору параметров и интеграции дополнительных регрессоров, что особенно эффективно для нестационарных нефтяных рядов
Agrawal et al. (2024)	Эмпирическое исследование прогнозирования цен на сырую нефть с использованием расширенной SARIMAX модели, интегрирующей экзогенные переменные для повышения точности	SARIMAX продемонстрировала превосходство над базовой SARIMA за счет учета внешних факторов (геополитика, макроэкономические показатели), что позволило улучшить точность прогнозов в условиях нестабильности рынка
Zhang & Hong (2022)	Исследование применяет LSTM-нейронные сети для прогнозирования цен на нефть, сравнивая их эффективность с традиционными статистическими методами (например, ARIMA) в условиях нестабильности рынка	LSTM показала более высокую точность по сравнению с ARIMA благодаря способности улавливать сложные нелинейные зависимости и долгосрочные временные паттерны в данных, особенно в периоды резких ценовых колебаний
Foroutan & Lahmiri (2024)	Сравнение широкого спектра моделей глубокого и машинного обучения (LSTM, BiLSTM, GRU, CNN-гибриды, TCN, LightGBM, Random Forest и др.) для прогнозирования цен на нефть (WTI, Brent) и драгоценные металлы в условиях волатильности 2020–2022 гг.	TCN показала наилучшую точность (минимальный MAE), LightGBM продемонстрировал сопоставимые результаты. LSTM и другие рекуррентные сети уступают более современным архитектурам, что ставит под сомнение их преимущество в прогнозировании цен на нефть

Источник: сост. авторами.

Перечисленные исследования по-разному подходят к проблеме прогнозирования: одни используют анализ волатильности, другие сентимент-анализ, третьи анализируют сезонность.

На основе обзора литературы и цели работы была сформулирована следующие гипотеза:

H1: Прогнозирование цены на нефть с помощью машинного обучения демонстрирует более высокую предсказательную способность по сравнению с регрессией.

Для проверки гипотезы мы ограничиваемся четырьмя моделями: тремя статистическими (ARIMA, SARIMA, SARIMAX) и одной с использованием машинного обучения (LSTM).

Одной из широко используемых статистических моделей для прогнозирования временных рядов является ARIMA (Autoregressive Integrated Moving Average). Данная модель сочетает в себе авторегрессию (AR), интегрирование (I) и скользящее среднее (MA). Наличие интегрирования позволяет использовать в качестве данных нестационарные ряды после дифференцирования.

Модификацией ARIMA с учетом фактора сезонности является модель SARIMA (Wang, 2023). Данная модель требует подбора дополнительных четырех параметров, связанных с фактором сезонности, что несколько усложняет ее использование. Тем не менее исследователи В. Ариянти и Т. Юснитасари (Ariyanti & Yusnitasari, 2023) пришли к выводу, что данная модель не дает значимого улучшения прогноза на ежедневных данных о цене на нефть, о чем говорят и другие исследования. Это может означать, что фактор сезонности либо не имеет значения при прогнозировании цены на нефть, либо выявляется на более длинных временных промежутках. В связи с этим чаще всего в исследованиях пренебрегают данной моделью в пользу простоты и строят прогноз с помощью ARIMA.

Другой модификацией модели ARIMA выступает модель SARIMAX, которая учитывает не только фактор сезонности, но и различные внешние факторы, т. е. включает в себя экзогенные переменные. Главным преимуществом данной модели является то, что она учитывает не только исторические данные о зависимой переменной, но и релевантные дополнительные данные: геополитические события (санкции, войны), макроэкономические показатели (ВВП, инфляция, спрос на нефть), финансовые факторы (курс доллара, биржевые индексы), природные и технологические факторы (ураганы, остановки добычи). В исследовании А. Агравал и соавторы (Agrawal et al., 2024) утверждают, что данная черта SARIMAX позволяет учитывать внешние шоки и делает эту модель очень полезной в периоды высокой волатильности.

Кроме описанных статистических моделей в исследовании используется модель с машинным обучением LSTM. Рассматриваются две модификации данной модели: с использованием только исторических данных о цене на нефть и с добавлением в модель экзогенных переменных.

Важно отметить, что в литературе чаще всего встречаются исследования с модификациями модели LSTM. Например, И. Ли и соавторы (Li et al., 2021) используют модель VMD-BiLSTM, которая включает анализ настроений инвесторов. Авторы приходят к выводу, что добавление новостей о рынке нефти улучшает прогноз и делает модель более эффективной, чем некоторые эконометрические модели и модели с машинным обучением. Кроме того, исследователи отмечают, что их модель лучше реагирует на шоки.

Исследователи Ю.-С. У и др. (Wu et al., 2019) предлагают метод, сочетающий в себе метод декомпозиции временных рядов EEDM (Ensemble Empirical Mode Decomposition) и классическую LSTM. EEDM в данной модели позволяет лучше учитывать «шумы», что делает эту модель более точной в прогнозировании, нежели модели ARIMA, SVM и обычную LSTM. К Гуан и Х. Гонг (Guan & Gong, 2023) предложили модель Mod-VMD-BiLSTM, в которой вариационное модальное разложение (VMD) позволяет выделить сглаженные подсерии, а двунаправленный LSTM обучается на их комбинациях; показано статистически значимое снижение ошибок прогноза на месячных ценах на нефть по сравнению с бенчмарками. Этот результат указывает, что предварительная фильтрация шума способна повышать устойчивость нейросетевого прогноза.

Однако не все исследования усложняют исходную модель включением в нее дополнительных методов обработки временных рядов. К. Джанг и М. Хонг (Zhang & Hong, 2022) используют классическую LSTM для анализа цены на нефть на разных временных промежутках. В работе использовались данные с 1986 по 2021 гг. Авторы пришли к нескольким важным выводам. Во-первых, модель LSTM обладает сильной обобщающей способностью и применима для прогнозирования цен на нефть с разными временными горизонтами. Во-вторых, по сравнению с моделями ANN и ARIMA модель LSTM показала более высокую точность прогнозирования цен на нефть на любых временных горизонтах. В-третьих, для модели LSTM чем длиннее временная шкала прогнозируемых цен, тем ниже точность прогноза. По этой причине авторы предполагают, что при длительном прогнозировании цен на нефть необходимо включать в модель другие факторы для сохранения высоких показателей точности.

В связи с этим в исследовании используется как LSTM на основе исторических цен на нефть марки WTI, так и LSTM с включением экзогенных переменных. Предполагается, что добавление этих переменных способно улучшить прогноз не только на длинных временных интервалах, но и при прогнозировании на 5 дней вперед.

Кроме того, использование LSTM в сочетании с различными факторами является как вариантом улучшения данной модели, так и способом ответить на второй важнейший исследовательский вопрос данной работы: приводит ли включение в модели экзогенных переменных к повышению точности прогноза относительно моделей с одной переменной? Исходя из этого исследовательского вопроса формулируется вторая гипотеза.

H2: Включение экзогенных переменных в модель приводит к увеличению точности прогноза по сравнению с моделями, использующими только данные о цене нефти.

2. Данные

Исследование проводится на временном интервале с 1 января 2015 г. по 31 декабря 2019 г. Указанный период характеризуется относительной стабильностью динамики цен на нефть: существенных и резких падений или стремительных ростов в этот временной промежуток не наблюдалось. В рассматриваемые годы также отсутствовали крупные экономические и политические кризисы, которые могли бы оказать заметное воздействие на нефтяной рынок. Данный интервал представляет собой последний продолжительный «спокойный» период: до него имел место

нефтяной кризис 2014 г., а после последовали кризисы 2020 г. и 2022 г., которые вновь нарушили стабильность.

Прогнозирование цен на нефть в условиях относительной макроэкономической стабильности требует применения подходов, отличных от тех, что используются во время кризисных периодов, когда рынок подвержен шокам и нестабильности (Tian et al., 2023). В данной работе внимание сосредоточено именно на прогнозировании цен в относительно «бескризисные» годы. Все данные, собранные за указанный период, были последовательно обработаны и разделены на обучающую и тестовую выборки в соотношении 80:20. Подобный метод разделения используется и в работе (Amulya et al., 2021), что обеспечивает сопоставимость исследовательских процедур.

Прогнозируемой переменной в исследовании выступает спотовая цена нефти марки West Texas Intermediate (WTI). Выбор данной марки нефти не является случайным. Во-первых, именно WTI широко применялась в предшествующих академических исследованиях, посвящённых вопросам моделирования и прогнозирования нефтяного рынка (Li et al., 2019). Во-вторых, нефть этой марки добывается на территории США, в штате Техас, что облегчает доступ к статистическим показателям, отражающим национальный рынок нефти. В-третьих, США являются одним из крупнейших потребителей нефти в мире, а также обладают богатой системой официальной статистики, что делает данный рынок особенно удобным для анализа. Таким образом, выбор спотовой цены WTI обусловлен как академической традицией, так и практическими преимуществами доступности и значимости данных.

Источником информации о спотовых ценах WTI является ресурс Investing.com¹, с которого были выгружены данные за указанный временной промежуток с ежедневной частотой. При выгрузке данных выявились пропуски, вызванные отсутствием торгов в выходные и праздничные дни. Для устранения этих пропусков применялась двухэтапная процедура: первое пропущенное значение в последовательности заменялось с помощью линейной интерполяции, тогда как все последующие значения внутри одного пропуска — с помощью интерполяции полиномом третьей степени (кубической). Такая процедура применялась для обеспечения целостности временного ряда. После завершения данной обработки ценовые данные были признаны пригодными для использования в прогнозных моделях.

Экзогенные переменные, включённые в исследование, были разделены на четыре основные группы: показатели, отражающие состояние нефтяного рынка; индикаторы, характеризующие финансовый рынок; индикаторы промышленного сектора; показатели, описывающие фондовый рынок.

Для характеристики состояния фондового рынка США использовался индекс Wilshire 5000. Этот индекс представляет собой наиболее всеобъемлющий показатель фондового рынка, охватывающий широкий спектр компаний — как промышленных, так и относящихся к сфере услуг и банковскому сектору. В предыдущих исследованиях для прогнозирования цен на нефть чаще применялись другие индексы фондового рынка, в частности индекс S&P 500 (An et al., 2019). Однако включение именно Wilshire 5000 представляется оправданным, т. к. данный показатель отражает совокупное состояние рынка акций США. Состояние фондового рынка может оказывать косвенное влияние на нефтяной рынок, поскольку рост котировок акций связан с ожиданиями повышения выручки и прибыли компаний. Поскольку нефть

¹ Investing.com — stock market quotes & financial news. (n. d.). Investing.com. <https://www.investing.com/> (дата обращения: 12.02.2025).

влияет практически на любую сферу бизнеса (через энергетику, транспортную логистику и ценообразование в смежных отраслях), рост фондового индекса может быть сопряжён с ростом совокупного потребления нефти, а значит, и с повышением её цены. Данные по индексу Wilshire 5000 были выгружены с ресурса Investing.com¹ с ежедневной частотой. Пропуски, возникшие из-за выходных и праздничных дней, были заполнены по описанной выше схеме: линейная интерполяция для первого пропуска и интерполяция третьей степени для всех последующих.

Для учёта состояния промышленного сектора использовался индекс Dow Jones, который применялся и в других исследованиях (Li et al., 2019). Значимость данного индекса для анализа нефтяного рынка объясняется тем, что предприятия промышленного сектора напрямую зависят от цен на нефть и нефтепродукты, а также от стоимости энергоресурсов и логистических расходов. Таким образом, динамика данного индекса оказывает влияние на спрос на нефть и её цену. Источником данных по индексу Dow Jones также выступал ресурс Investing.com², частота наблюдений — ежедневная. Пропуски были устранены по той же методике, что и в случае с индексом Wilshire 5000.

В качестве индикатора состояния финансового сектора использовался индекс доллара США (DXY). Включение данного индекса в модель прогнозирования цен на нефть имеет как теоретические, так и эмпирические основания. Ранее он также применялся исследователями (Li et al., 2019). Его значимость обусловлена, во-первых, тем, что мировые цены на нефть номинированы в долларах США, следовательно, колебания курса доллара непосредственно отражаются на международных потоках импорта и экспорта нефти. Во-вторых, потребление нефти связано с общей экономической активностью, которая зависит от денежно-кредитной политики Федеральной резервной системы США. Изменение процентных ставок и иных параметров денежной политики оказывает влияние на курс доллара по отношению к другим валютам, а через него — на потребление и цену нефти. Данные по индексу доллара были получены с ресурса Investing.com³ за рассматриваемый период, частота — ежедневная. Пропущенные значения также были восстановлены по единой методике: линейная интерполяция для первого значения и интерполяция третьей степени для последующих.

Для анализа состояния нефтяного рынка использовался набор показателей, включающий: чистый импорт нефти США, потребление нефтепродуктов в США, объём добычи нефти на территории страны, объём переработки нефти, количество активных буровых установок, уровень запасов нефти, количество дней, на которое этих запасов достаточно, спред цен между марками Brent и WTI, а также индекс танкерных перевозок BDTI (Baltic Dirty Tanker Index). Для большинства показателей (все, кроме спреда Brent–WTI и индекса BDTI) доступны еженедельные данные, которые были выгружены из базы данных Управления энергетической информации (EIA)⁴. Спред между Brent и WTI, а также индекс BDTI представлены

¹ Investing.com — stock market quotes & financial news. (n. d.). Investing.com. <https://www.investing.com/> (дата обращения: 12.02.2025).

² Там же.

³ Там же.

⁴ U.S. Energy Information Administration (EIA). (2025). <https://www.eia.gov/> (дата обращения: 12.02.2025).

в виде ежедневных данных, которые были получены с ресурса Investing.com¹. Для всех рядов применялась стандартная процедура восстановления пропусков.

Индекс BDTI отражает стоимость танкерных перевозок нефти. Его рост может изменять поведение покупателей и влиять на спрос, что в конечном счёте отражается на цене нефти. Поэтому данный индекс обоснованно включён в модель прогнозирования. Включение спреда Brent–WTI основано на результатах исследования (Fattouh, 2009), где показано, что расхождения в ценах между сортами нефти стимулируют арбитражные операции: трейдеры продают нефть WTI и покупают Brent, что усиливает давление на цену WTI. Кроме того, разрыв между Brent и WTI отражает относительный спрос на эти сорта нефти, а значит, может оказывать долгосрочное влияние на ценовую динамику.

Переменные спроса включают: потребление нефтепродуктов, объём переработки нефти, уровень запасов, количество дней обеспечения запасами и чистый импорт. Снижение объёмов переработки нефти ведет к уменьшению спроса на сырую нефть, тогда как их рост — к увеличению спроса. Запасы отражают баланс спроса и предложения: их рост сигнализирует о превышении предложения, что может вести к снижению цен, а сокращение запасов, напротив, указывает на усиление спроса и рост цен. Чистый импорт также является индикатором спроса: его снижение может указывать на сокращение потребления нефти, что оказывает давление на цену, а рост — на расширение спроса. Показатель количества дней, на которые хватает текущих запасов сырой нефти, также имеет важное значение: его уменьшение свидетельствует о росте спроса и повышательном давлении на цены, а увеличение — об избытке предложения и потенциальном снижении цен.

Переменные предложения представлены двумя индикаторами: объёмом добычи нефти и числом активных буровых установок. Число установок отражает долгосрочные тенденции в добыче: их сокращение, как правило, предвещает снижение объёмов добычи нефти в будущем.

Все используемые показатели имеют различные единицы измерения. Так, индексы Wilshire 5000, Dow Jones, DXY и BDTI выражены в базисных пунктах. Объёмы добычи нефти, потребления нефтепродуктов, чистого импорта и переработки измеряются в миллионах баррелей. Спред Brent–WTI выражается в долларах США за баррель. Показатель «количество дней обеспечения запасами» измеряется в сутках. При этом данные о запасах нефти на территории США представлены в виде изменений относительно уровня предыдущей недели, измеряемых в миллионах баррелей.

3. Методы

Для проверки стационарности временного ряда был проведён тест Дики–Фуллера. Полученное значение *p-value* составило 0,297, что превышает пороговое значение 0,05, следовательно, исходный ряд является нестационарным. В целях устранения нестационарности было проведено интегрирование первого порядка, после чего вновь применён тест Дики–Фуллера. Для преобразованного ряда значение *p-value* оказалось равным 0,00, что позволяет сделать вывод о стационарности преобразованного ряда.

Для построения модели ARIMA требовалось определить оптимальные значения её параметров: порядок авторегрессии, порядок интегрирования и порядок сколь-

¹ Investing.com — stock market quotes & financial news. (n. d.). Investing.com. <https://www.investing.com/> (дата обращения: 12.02.2025).

зующего среднего. Порядок интегрирования был установлен на основе теста Дики–Фуллера и равен единице. Значения параметров авторегрессии и скользящего среднего были определены с использованием критерия Акаике (AIC), при этом оптимальными считаются те значения, при которых AIC минимален. В результате были получены следующие параметры: порядок авторегрессии — 4, порядок скользящего среднего — 4. На основе этих параметров была построена модель. Все параметры оказались статистически значимыми, т. к. их p -value < 0,05. Значение критерия Акаике для модели составило 4694,395. Для оценки точности прогноза использовалась среднеквадратическая ошибка (RMSE), равная 1,668.

Модель SARIMA является расширением модели ARIMA, позволяющим учитывать сезонность временного ряда. Её использование целесообразно в случае выявления сезонных компонентов в данных. Для проверки наличия сезонности были применены три метода: декомпозиция временного ряда, анализ автокорреляционной функции и преобразование Фурье. Результаты анализа Фурье, выполненного с помощью языка Python, выявили возможные периоды сезонности: 182, 608, 203, 365, 456, 261, 183, 152, 304. Однако указанные значения не совпали с результатами декомпозиции ряда, а автокорреляционная функция не показала выраженной сезонной структуры. На основании совокупного анализа было сделано заключение об отсутствии сезонности, что исключает необходимость использования модели SARIMA.

Модель SARIMAX, являющаяся обобщением ARIMA с учётом экзогенных переменных, строилась на базе ранее полученных параметров ARIMA и результатов анализа сезонности. Параметры сезонности были приравнены к нулю ввиду отсутствия сезонных компонентов в ряде, остальные параметры соответствовали модели ARIMA. Критерий Акаике показал, что данный набор параметров является оптимальным и для модели SARIMAX. Следующим этапом стало определение экзогенных переменных, значимых для прогноза. На основе статистической значимости (p -value < 0,05) были отобраны следующие переменные: индекс Wilshire 5000, индекс Dow Jones, индекс доллара (DXY), спред котировок нефти марок Brent–WTI, а также запасы нефти на территории США. Значение критерия Акаике для построенной модели SARIMAX составило 4462,02.

Построение модели LSTM требовало предварительной нормализации данных. Далее подбирались параметры нейронной сети: количество эпох обучения (*epochs*), размер батча (*batch size*), коэффициент регуляризации (*dropout rate*), метод оптимизации (*optimizer*), функция активации (*activation function*), функция потерь (*loss function*), количество слоёв сети (*num layers*), число скрытых единиц (*hidden size*) и длина входной последовательности (*sequence length*).

Длина входной последовательности определяет число временных шагов, учитываемых моделью при обучении, т. е. интервал, в течение которого сеть «видит» динамику котировок. Количество скрытых единиц определяет размерность внутреннего представления данных в слое сети: чем выше этот показатель, тем более сложные зависимости может выучить модель. Однако рост числа скрытых нейронов увеличивает вычислительные затраты и риск переобучения. Количество слоёв влияет на способность сети выявлять более сложные паттерны. Увеличение числа слоёв повышает глубину модели, но одновременно ведёт к росту требований к вычислительным ресурсам и может вызвать переобучение.

Метод оптимизации определяет процедуру обновления весов нейронной сети на основе ошибки прогноза. Среди наиболее распространённых методов выделя-

ются Adam (Adaptive Moment Estimation), RMSprop (Root Mean Square Propagation), SGD with Momentum (Stochastic Gradient Descent) и AdaGrad (Adaptive Gradient Algorithm). В настоящем исследовании был использован оптимизатор Adam, продемонстрировавший высокую эффективность на сравнительно небольших выборках и обладающий преимуществом адаптивной корректировки скорости обучения.

Функция потерь обеспечивает количественную оценку отклонения прогноза от фактических значений. Для LSTM применяются следующие функции: среднеквадратическая ошибка (MSE), средняя абсолютная ошибка (MAE), функция Хубера (Huber loss), а также квантильная ошибка (Quantile loss). В данной работе использовалась функция MSE, отличающаяся простотой интерпретации и высокой чувствительностью к выбросам.

На основе предварительного анализа был выбран следующий набор параметров для LSTM: функция активации — гиперболический тангенс, оптимизатор — Adam, функция потерь — MSE, количество слоёв — 2, коэффициент *dropout* — 0,2. Прогнозирование осуществлялось на 5 дней методом расширяющегося окна. Остальные гиперпараметры были подобраны экспериментально с целью минимизации RMSE. Оптимальным оказался набор: количество эпох — 200, размер батча — 16, число скрытых нейронов — 50 (в каждом слое), количество слоёв LSTM — 2, длина входной последовательности — 90. Данный набор параметров обеспечил значение $RMSE = 1,47$. На основе спрогнозированных значений был построен график (рис. 2), позволяющий визуальное сопоставить прогноз с фактическими значениями целевой переменной.

Дополнительно была построена модель LSTM с учётом экзогенных переменных. Подбор гиперпараметров также осуществлялся на основе минимизации RMSE. На первом этапе проверялся набор гиперпараметров, использованный в модели без экзогенных переменных. Данная модель показала $RMSE = 1,88$, что выше, чем в предыдущем случае. Затем были протестированы альтернативные наборы параметров. Оптимальным оказался следующий набор: количество слоёв LSTM — 2, число скрытых нейронов — 100 (в каждом слое), длина входной последовательности — 90, количество эпох — 70, размер батча — 16. Данный вариант обеспечил $RMSE = 1,50$. После получения прогноза было проведено обратное преобразование



Рис. 2. LSTM без экзогенных переменных (источник: сост. авторами)

Fig. 2. LSTM without Exogenous Variables



Рис. 3. LSTM без экзогенных переменных (источник: сост. авторами)

Fig. 3. LSTM without Exogenous Variables

данных, а на основе результатов построен график (рис. 3), визуализирующий прогноз и позволяющий сопоставить его с фактическими значениями.

4. Результаты

На основе модели ARIMA был построен график прогноза, сопоставленный с фактическими значениями прогнозируемой переменной — цены на нефть марки WTI. Для тестовой выборки был рассчитан показатель точности прогноза RMSE, равный 1,668. Этот результат в дальнейшем использовался для сравнения эффективности различных моделей прогнозирования.

График модели ARIMA (рис. 4) показывает умеренное соответствие прогноза фактическим данным, однако наблюдаются заметные отклонения в периоды высокой волатильности. Это свидетельствует о том, что модель недостаточно точно прогнозирует резкие колебания цены на нефть. Данный недостаток особенно критичен при прогнозировании на длительные горизонты, поскольку приводит к существенным отклонениям прогнозных значений от фактических. Таким образом,

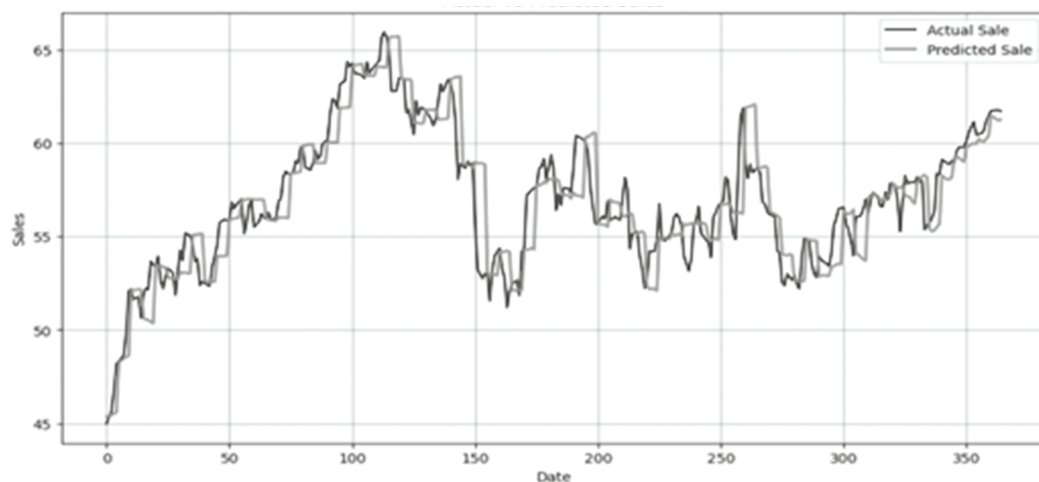


Рис. 4. ARIMA (4,1,4) (источник: сост. авторами)

Fig. 4. ARIMA (4,1,4)

результаты настоящей работы подтверждают выводы предшествующих исследований: модель ARIMA удовлетворительно работает в краткосрочном прогнозировании, но плохо улавливает резкие изменения цен даже на небольших интервалах.

Следует отметить, что ARIMA использует для прогнозирования исключительно один временной ряд — исторические данные о цене на нефть WTI. Это означает, что модель не учитывает влияние внешних факторов. По этой причине в исследовании применялась модификация ARIMA, позволяющая включать экзогенные переменные, — модель SARIMAX.

Для построения модели SARIMAX были протестированы различные комбинации экзогенных переменных. Наилучший результат прогнозирования ($RMSE = 1,51$) показала модель, включающая индексы Wilshire 5000, Dow Jones, индекс доллара (DXY), спред Brent–WTI, а также данные о запасах нефти на территории США. Подобный набор переменных не встречался в аналогичном составе в предыдущих исследованиях, однако использование рыночных индексов для прогнозирования цен на нефть фиксировалось, например, в работе Х. Ли и соавторов (Li et al., 2019), где применялся индекс S&P 500. Тот факт, что включение дополнительных переменных улучшило результат прогноза по сравнению с моделью без экзогенных факторов, подтверждает вторую гипотезу настоящего исследования.

Выбор указанных переменных для модели SARIMAX был обусловлен их статистической значимостью ($p\text{-value} < 0,05$). Включение фондовых индексов Wilshire 5000 и Dow Jones положительно отразилось на точности прогноза, поскольку эти показатели отражают экономическую активность и совокупный спрос на нефть. Снижение фондовых индексов указывает на спад деловой активности, что влечёт за собой сокращение потребления энергоресурсов, логистических услуг и продукции нефтехимии, а значит, и снижение спроса на нефть. Кризисные явления в реальном секторе экономики сопровождаются как падением фондовых индексов, так и снижением цен на нефть, что усиливает их взаимосвязь. Так, резкое снижение индекса Dow Jones в 2019–2020 гг. предшествовало обвалу нефтяных цен, и модель SARIMAX с этой переменной смогла лучше предсказать падение. При этом важно учитывать, что обвалы на фондовом рынке происходят не только в периоды кризисов, но также могут быть вызваны внезапными новостями и ожиданиями.

Индекс доллара (DXY) также оказался значимой экзогенной переменной. Это объясняется тем, что мировая торговля нефтью ведётся в долларах США. Укрепление доллара относительно других валют приводит к снижению долларовой цены на нефть, поскольку делает её менее доступной для стран-импортёров. Дополнительно индекс доллара отражает общую деловую активность в США, а его динамика зависит от денежно-кредитной политики Федеральной резервной системы. Как ключевой ресурс нефть чувствительно реагирует на колебания данного индекса, что обосновывает включение этой переменной в модель.

Влияние на цену нефти оказывает и спред между котировками Brent и WTI. Его эффект объясняется несколькими факторами. Во-первых, финансовым: при значительном расхождении спреда трейдеры продают нефть WTI и покупают Brent, что усиливает давление на котировки WTI. Во-вторых, логистическим: удорожание транспортировки нефти WTI снижает её привлекательность и спрос, что ведёт к увеличению спреда. В-третьих, геополитическим: обострение ситуации на Ближнем Востоке может привести к перебоям поставок нефти Brent, что вызывает рост

спроса на нефть WTI, рост её цен и сокращение спреда. Таким образом, включение данной переменной положительно влияет на точность прогнозирования.

Запасы нефти в США являются одним из ключевых индикаторов, отражающих фактор предложения. Увеличение запасов сигнализирует о переизбытке нефти на рынке и оказывает понижающее давление на цены. Добавление временного ряда, характеризующего запасы нефти, также повысило прогностическую способность модели SARIMAX.

В целом использование модели SARIMAX позволяет повысить качество прогнозов за счёт учёта экзогенных переменных. Однако, несмотря на улучшение прогноза, модель остаётся авторегрессионной и, как и ARIMA, плохо справляется с резкими колебаниями цен и медленно адаптируется к изменениям тренда.

В соответствии с гипотезой 1, применение методов машинного обучения должно улучшать точность прогнозирования цен на нефть. Для проверки данной гипотезы в исследование была включена модель LSTM.

Первый вариант модели строил прогноз исключительно на основе исторических данных о ценах на нефть марки WTI. Использование LSTM для прогнозирования обосновано её способностью выявлять сложные нелинейные зависимости в данных. Результаты показали, что данная модель обеспечивает более высокую точность по сравнению с ARIMA: RMSE составил 1,51 против 1,668. Для минимизации ошибки были подобраны оптимальные гиперпараметры: два слоя LSTM для обработки последовательностей, длина входной последовательности — 90, количество скрытых нейронов — 50 (в каждом слое), количество эпох обучения — 300, размер батча — 16.

Следует отметить, что модель LSTM демонстрирует лучшую адаптацию к краткосрочным колебаниям цены, а также более высокое визуальное соответствие фактическим данным по сравнению с моделью ARIMA, особенно на нелинейных участках. Вместе с тем использование данной архитектуры сопряжено с определёнными сложностями: обучение модели требует значительных ресурсов, плохо подходит для анализа малых временных интервалов, а также требует тщательной настройки гиперпараметров. Кроме того, при одинаковом наборе гиперпараметров результаты прогнозирования могут незначительно варьироваться, что свидетельствует о стохастическом характере обучения нейронных сетей.

Второй вариант модели LSTM был построен с учётом как основной переменной (спотовая цена нефти WTI), так и экзогенных переменных. Этот вариант является усложнением предыдущего, поскольку требует подбора гиперпараметров и переобучения модели под каждый набор экзогенных факторов.

В данном исследовании наилучший результат был достигнут при использовании прогноза на основе спотовой цены нефти WTI и индекса Dow Jones в качестве экзогенной переменной. Подбор гиперпараметров также осуществлялся с ориентиром на минимизацию RMSE. Оптимальным оказался следующий набор параметров: два слоя LSTM для обработки последовательностей, длина входной последовательности — 90, количество скрытых нейронов — 100 (в каждом слое), количество эпох обучения — 70, размер батча — 16.

Визуальное представление прогнозов данной модели приведено на рисунке 5 и рисунке 6.

Таким образом, добавление экзогенных переменных позволило улучшить качество прогноза, что подтверждает сделанное предположение о преимуществах



Рис. 5. LSTM с экзогенной переменной (источник: сост. авторами)

Fig. 5. LSTM with an Exogenous Variable

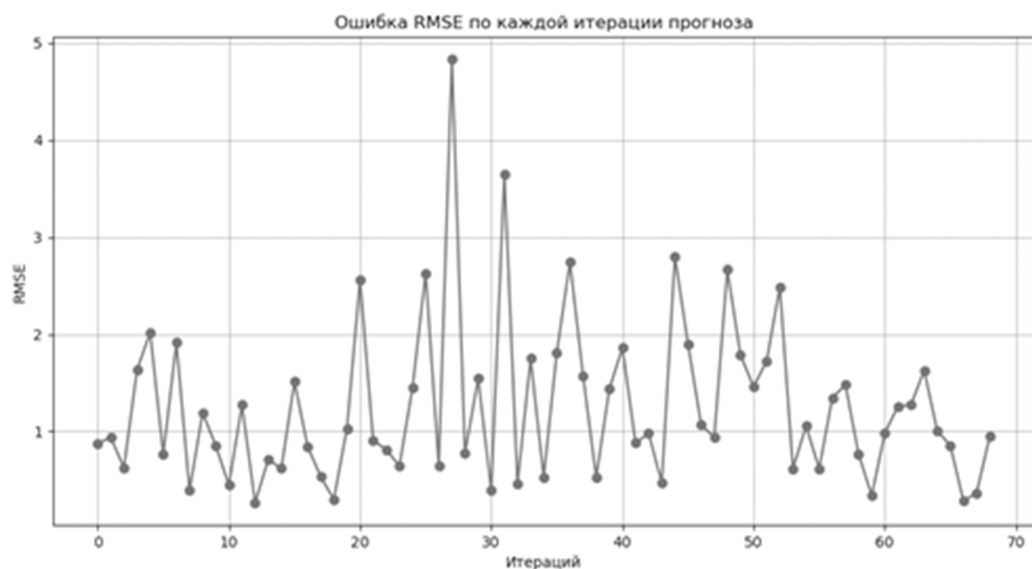


Рис. 6. RMSE для модели LSTM с экзогенной переменной (источник: сост. авторами)

Fig. 6. RMSE for the LSTM Model with an Exogenous Variable

моделей машинного обучения. Результаты согласуются с выводами других исследований, в частности работы Х. Ли и соавторов (Li et al., 2019), где показано, что использование моделей, основанных на методах машинного обучения, таких как Random Forest, также обеспечивает более высокую точность прогнозирования по сравнению с классическими статистическими моделями.

5. Выводы

В результате исследования обе гипотезы были подтверждены. Использование машинного обучения помогает лучше предсказывать цену не нефть, поскольку модели с машинным обучением демонстрирует более высокую предсказательную способность по сравнению с авторегрессионными моделями. Самые точные результаты прогноза (RMSE: 1,5) среди всех построенных моделей показала модель LSTM с использованием исторических данных цены нефти WTI и индекса Dow

Jones. Это свидетельствует о том, что у моделей с машинным обучением есть высокий потенциал в прогнозировании. Полученные результаты могут быть улучшены за счет применения более сложных структур модели. Также улучшению результатов может способствовать уменьшение горизонта прогнозирования. К таким же выводам приходят исследователи К. Джанг и М. Хонг (Zhang & Hong, 2022). Другим способом улучшения прогноза может быть использование дополнительных модулей обработки данных, например, для включения качественных данных, таких как новости о рынке нефти. Третьим способом повысить точность прогнозирования является построение другой модели на основе машинного обучения. Таким образом, применение машинного обучения в прогнозировании цены на нефть имеет большой потенциал, поскольку данный вид моделей может быть модернизирован, что отличает его от авторегрессионных моделей.

Другим важным выводом исследования является то, что включение экзогенных переменных в модель приводит к увеличению точности прогноза. Это гипотеза подтверждается как для авторегрессионных моделей, так и для моделей, использующих машинное обучение. Модель SARIMAX (RMSE: 1,51) показывает лучшие результаты прогнозирования, нежели модель ARIMA (RMSE: 1,668). Такие результаты позволяют достигать использование следующих экзогенных переменных: индекса доллара DXY, индекс Dow Jones и индекс Wilshire 5000, а также показатели спреда WTI-Brent и запасы нефти в США. Модель LSTM с экзогенными переменными также демонстрирует несколько более точные результаты прогноза (RMSE: 1,5), нежели её модификация только с прогнозируемой переменной (RMSE: 1,51). Для этого в модель должен быть включен индекс Dow Jones. Тем не менее для LSTM улучшение прогноза является незначительным, что не дает сделать однозначный вывод о превосходстве модели с экзогенными переменными над моделью без них. Стоит отметить, что в других исследованиях (Li et al., 2019) применение моделей с машинным обучением в связке с экзогенными переменными позволяет значительно увеличить точность прогноза относительно авторегрессионных моделей или моделей с машинным обучением, но без экзогенных переменных. Однако важно сказать, что в упомянутых работах использовались другие модели машинного обучения, например, Random Forest.

Модели, построенные в работе, наиболее актуальны для трейдеров, поскольку для них важны краткосрочные прогнозы цены на нефть. Для компаний и государств актуален долгосрочный прогноз, который не может быть построен только на основе исторических данных о цене на нефть и экзогенных переменных. Построение такого прогноза требует комплексного анализа рынка нефти (спроса и предложения). Однако надо упомянуть и том, что показатель RMSE, который используется в исследовании для оценки точности модели, отражает отклонение прогнозируемой цены от реальной, что означает, что самая точная модель, построенная в работе, в среднем «ошибается» на 1,5 доллара. Для использования на практике, например, в трейдинге такое отклонение является довольно большим. В связи с этим одним из логичных вариантов развития данных моделей является уменьшение горизонта прогнозирования.

6. Заключение

Область прогнозирования цены на нефть является важной частью экономических изысканий, поскольку имеет практический смысл как для государства, так

и для бизнеса. Рынок нефти является одним из наиболее непредсказуемых и волатильных рынков. Поэтому прогнозирование — крайне актуальная задача.

Внедрение машинного обучения — это новый этап в области прогнозирования. Методы, которые появляются и развиваются в последние годы, требуют больших данных и вычислительных мощностей, однако позволяют выявлять более сложные закономерности. Тестирование и настройка одной из таких моделей — LSTM — одна из основных задач, поставленных в данном исследовании. Данная модель была построена в разных вариантах: как с историческими данными о цене на нефть, так и с добавлением дополнительных переменных. Она была сравнена с традиционными моделями ARIMA и SARIMAX. Результаты, полученные в этой работе, говорят о том, что машинное обучение действительно позволяет добиться более точных результатов.

Учет различных экономических факторов в прогнозе — ещё одно из направлений, которое может помочь в улучшении качества прогноза, поскольку нефть — это ресурс, цена которого зависит от большого числа факторов. В данном исследовании модели SARIMAX и LSTM с экзогенными переменными действительно демонстрирует лучшие результаты, чем модели без учета факторов. Делается вывод, что для модели SARIMAX индексы Wilshire, Dow Jones, индекс доллара, запасы нефти и спред между ценой нефти марки WTI и ценой нефти марки Brent при добавлении в модель улучшают результаты её прогноза. А для модели LSTM фактором, улучшающим качество прогноза, является индекс Dow Jones.

Данная работа охватывает лишь небольшую часть существующих моделей и экзогенных переменных. В дальнейшем исследователи могут улучшить результаты использованных моделей путем обучения на других временных промежутках, добавления новых экзогенных переменных, а также тестирования других моделей машинного обучения (например, модели Random Forest) или доработки моделей.

Важно отметить, что прогнозирование — крайне сложная задача. В мире, где геополитические события, природные катастрофы или пандемии способны за день резко изменить долгосрочный тренд цены на нефть, появляется огромный фактор неопределенности, который не сможет предсказать никакая, даже самая точная модель. Однако те данные, которое человечество получает регулярно и умеет обрабатывать: финансовая информация, показатели добычи и запасов, метрики, отражающие спрос и предложение, — должны быть наиболее эффективно использованы при прогнозировании.

Список источников / References

- Agrawal, A., Kadam, S., Kapoor, P. A., & Rashid, M. (2024). Predicting crude oil prices using SARIMA-X method: An empirical study. *International Journal of Financial Engineering*, 12(01), 2450007. <https://doi.org/10.1142/S2424786324500075>
- Ariyanti, V., & Yusnitasari, T. (2023). Comparison of ARIMA and SARIMA for Forecasting Crude Oil Prices. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 7(2), 405–413. <https://doi.org/10.29207/resti.v7i2.4895>
- Fattouh, B., Kilian, L., & Mahadeva, L. (2012). The Role of Speculation in Oil Markets: What Have We Learned So Far? *The Energy Journal*, 34(3). <https://doi.org/10.5547/01956574.34.3.2>
- Foroutan, P., & Lahmiri, S. (2024). Deep learning systems for forecasting the prices of crude oil and precious metals. *Financial Innovation*, 10(1), 111. <https://doi.org/10.1186/s40854-024-00637-z>

- Gross, W. F., Hayden, C., & Butz, C. (2012). About the impact of rising oil price on logistics networks and transportation greenhouse gas emission. *Logistics Research*, 4(3–4), 147–156. <https://doi.org/10.1007/s12159-012-0072-2>
- Guan, K., & Gong, X. (2023). A new hybrid deep learning model for monthly oil prices forecasting. *Energy Economics*, 128, 107136. <https://doi.org/10.1016/j.eneco.2023.107136>
- Gupta, N., & Nigam, S. (2020). Crude Oil Price Prediction using Artificial Neural Network. *Procedia Computer Science*, 170, 642–647. <https://doi.org/10.1016/j.procs.2020.03.136>
- Kristjanpoller, W., & Minutolo, M. (2016). Forecasting volatility of oil price using an artificial neural network-GARCH model. *Expert Systems with Applications*, 65, 233–241. <https://doi.org/10.1016/j.eswa.2016.08.045>
- Li, X., Shang, W., & Wang, S. (2019). Text-based crude oil price forecasting: A deep learning approach. *International Journal of Forecasting*, 35(4), 1548–1560. <https://doi.org/10.1016/j.ijforecast.2018.07.006>
- Li, Y., Jiang, S., Li, X., & Wang, S. (2021). The role of news sentiment in oil futures returns and volatility forecasting: Data-decomposition based deep learning approach. *Energy Economics*, 95, 105140. <https://doi.org/10.1016/j.eneco.2021.105140>
- Mohammadi, H., & Su, L. (2010). International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Economics*, 32(5), 1001–1008. <https://doi.org/10.1016/j.eneco.2010.04.009>
- Natarajan, G., & Ashok, A. (2018). Multivariate Forecasting of Crude Oil Spot Prices using Neural Networks. *arXiv*. <https://doi.org/10.48550/arXiv.1811.08963>
- Pham, T. B., & Sala, H. (2019). *The macroeconomic effects of oil price and risk-premium shocks on Vietnam: Evidence from an over-identifying SVAR analysis*.
- Ratti, R. A., & Vespignani, J. (2014). Oil Prices and the Economy: A Global perspective. *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.2440986>
- Santos, M. V., Morgado-Dias, F., & Silva, T. (2023). Oil Sector and Sentiment Analysis — A Review. *Energies*, 16(12), 4824. <https://doi.org/10.3390/en16124824>
- Sek, S. K., Teo, X. Q., & Wong, Y. N. (2015). A comparative study on the effects of oil price changes on inflation. *Procedia Economics and Finance*, 26, 630–636. [https://doi.org/10.1016/s2212-5671\(15\)00800-x](https://doi.org/10.1016/s2212-5671(15)00800-x)
- Tang, W., Wu, L., & Zhang, Z. (2010). Oil price shocks and their short-and long-term effects on the Chinese economy. *Energy Economics*, 32, S3–S14. <https://doi.org/10.1016/j.eneco.2010.01.002>
- Tian, G., Peng, Y., & Meng, Y. (2023). Forecasting crude oil prices in the COVID-19 era: Can machine learn better? *Energy Economics*, 125, 106788. <https://doi.org/10.1016/j.eneco.2023.106788>
- Vo, A. H., Nguyen, T., & Le, T. (2020). Brent Oil Price Prediction Using Bi-LSTM Network. *Intelligent Automation & Soft Computing*, 26(6), 1307–1317. <https://doi.org/10.32604/iasc.2020.013189>
- Wang, Y. (2023). Oil Price Forecasting based on Improved SARIMA Model. *Highlights in Science, Engineering and Technology*, 39, 238–245. <https://doi.org/10.54097/hset.v39i.6535>
- Wu, Y.-X., Wu, Q.-B., & Zhu, J.-Q. (2019). Improved EEMD-based crude oil price forecasting using LSTM networks. *Physica A: Statistical Mechanics and its Applications*, 516, 114–124. <https://doi.org/10.1016/j.physa.2018.09.120>
- Zhang, K., & Hong, M. (2022). Forecasting crude oil price using LSTM neural networks. *Data Science in Finance and Economics*, 2(3), 163–180. <http://dx.doi.org/10.3934/DSFE.2022008>

Информация об авторах

Назарова Варвара Вадимовна — кандидат экономических наук, доцент департамента финансов СПбШЭиМ, Национальный исследовательский университет «Высшая школа экономики»; <https://orcid.org/0000-0002-9127-1644> (Российская Федерация, 190008, г. Санкт-Петербург, ул. Кантемировская, д. 3А; e-mail: nvarvara@list.ru).

Лодягин Борис Алексеевич — преподаватель департамента финансов СПбШЭиМ, Национальный исследовательский университет «Высшая школа экономики»; <https://orcid.org/0009-0001-0434-7381> (Российская Федерация, 190008, г. Санкт-Петербург, ул. Кантемировская, д. 3А; e-mail: blodyagin@hse.ru).

Круглов Фёдор Аркадьевич — магистрант, СПбШЭиМ, Национальный исследовательский университет «Высшая школа экономики» (Российская Федерация, 190008, г. Санкт-Петербург, ул. Кантемировская, д. 3А; e-mail: fakruglov@edu.hse.ru).

Круглов Арсений Вадимович — магистрант, СПбШЭиМ, Национальный исследовательский университет «Высшая школа экономики» (Российская Федерация, 190008, г. Санкт-Петербург, ул. Кантемировская, д. 3А; e-mail: avkruglov@edu.hse.ru).

About the authors

Varvara V. Nazarova — Cand. Sci. (Econ.), Associate Professor, Department of Finance, St. Petersburg School of Economics and Management, National Research University “Higher School of Economics”; <https://orcid.org/0000-0002-9127-1644> (3A, Kantemirovskaya St., St. Petersburg, 190008, Russian Federation; e-mail: nvarvara@list.ru).

Boris A. Lodiagin — Lecturer, Doctoral Student, Department of Finance, St. Petersburg School of Economics and Management, National Research University “Higher School of Economics”; <https://orcid.org/0009-0001-0434-7381> (3A, Kantemirovskaya St., St. Petersburg, 190008, Russian Federation; e-mail: blodyagin@hse.ru).

Fedor A. Kruglov — Master’s degree student, St. Petersburg School of Economics and Management, National Research University “Higher School of Economics” (3A, Kantemirovskaya St., St. Petersburg, 190008, Russian Federation; e-mail: fakruglov@edu.hse.ru).

Arseniy V. Kruglov — Master’s degree student, St. Petersburg School of Economics and Management, National Research University “Higher School of Economics” (3A, Kantemirovskaya St., St. Petersburg, 190008, Russian Federation; e-mail: avkruglov@edu.hse.ru).

Дата поступления рукописи: 30.06.2025.

Прошла рецензирование: 28.07.2025.

Принято решение о публикации: 03.09.2025.

Received: 30 Jun 2025.

Reviewed: 28 Jul 2025.

Accepted: 03 Sep 2025.