

© 2023

Роман Кузнецов

аспирант

(e-mail: romak2921@mail.ru)

Татьяна Тумарова

кандидат экономических наук, профессор, директор института магистратуры

(e-mail: tumarova@finec.ru)

(ФГБОУ ВО «Санкт-Петербургский государственный экономический университет»)
Санкт-Петербург, Российская Федерация

ПРОГНОЗИРОВАНИЕ КОТИРОВОК АКЦИЙ ПАО НОВАТЭК С ИСПОЛЬЗОВАНИЕМ МОДЕЛЕЙ ДЕРЕВА РЕШЕНИЙ И ЛИНЕЙНОЙ РЕГРЕССИИ

Развитие технологий и появление различных моделей машинного обучения оказывает влияние на многие сферы деятельности общества, в том числе и на анализ и прогнозирование фондового рынка. Умение грамотно подбирать и использовать модели машинного обучения в прогнозировании котировок акций является одним из ключевых в настоящее время конкурентным преимуществом, позволяющим увеличивать свою прибыль от деятельности на рынке как крупным инвестиционным компаниям, так и частным лицам. В исследовании отражена эффективность использования моделей «дерева решений» и линейной регрессии при прогнозировании дневных котировок ПАО НОВАТЭК.

Ключевые слова: фондовая биржа, дерево решений, линейная регрессия, машинное обучение.

DOI: 10.31857/S020736760026686-3

Введение

Интерес обычного человека к фондовому рынку за последние несколько десятилетий пережил экспоненциальный рост, что, в свою очередь, влияет на увеличение оборотов активов на финансовых биржах. Цель участников фондового рынка — это максимизация собственной прибыли, а достигается данная цель разными способами и инструментами, в том числе прогнозированием котировок акций на разных временных отрезках. Возможность достаточно точно прогнозировать поведение рынка дает возможность как институциональному, так и частному инвестору получать более высокую доходность с поправкой на риск. Поэтому для повышения точности прогноза в настоящее время активно используются различные модели машинного обучения.

Значительные изменения фондового рынка оказывают сильное влияние на экономику. Например, крах фондового рынка в 1929 году стал одним из основных триггеров, запустивших Великую депрессию 1930-х годов.

В противоположной ситуации, когда цены на акции растут, происходит стимулирование экономики, и многие компании стремятся провести публичное

размещение акций (IPO), чтобы привлечь капитал и развивать бизнес. Процессы слияний и поглощений также происходят, по большей части во время растущего рынка. Такой рост инвестиций, в свою очередь, ведет к росту экономики.

Машинное обучение — это раздел информационных технологий, в рамках которого создаются и тестируются алгоритмы, предназначенные для разных целей, включая прогнозирование. Как аналитики и брокеры, так и компании, создающие автоматизированные системы торговли на фондовом рынке прибегают к машинному обучению для прогнозирования котировок акций. В такой ситуации участники рынка зачастую нацелены не на долгосрочные ожидания, а на краткосрочные движения котировок вверх, за которыми зачастую следуют сильные распродажи акций и падения котировок. Прогнозирование котировок акций представляет собой задачу, требующую большого количества обработанных и структурированных данных о фондовом рынке.

В рамках исследования авторами используются и сравниваются модель «дерева решений» (Decision Tree) и линейная регрессия. «Дерево решений» — это контролируемый алгоритм обучения, ориентированный на вывод значений целевых переменных в соответствии с порядком, подготовленным на основе обучающих данных. «Дерево решений» считается удобным инструментом для прогнозирования, позволяющим пользователям без труда интерпретировать данные. В машинном обучении линейная регрессия выступает базовым инструментом, с помощью которого можно выявить линейный тренд.

Настоящая работа дополняет существующую литературу по использованию моделей «дерева решений» и линейной регрессии при прогнозировании котировок акций и финансовых деривативов. С момента появления фондового рынка множество исследователей пытаются предсказать стоимость акций, используя различные алгоритмы машинного обучения. Во многих работах наблюдается разнообразие вариантов использования нейронных сетей с разным набором параметров. Сингх С., Кумар В. и коллеги в своем исследовании на тему прогнозирования акций используют в качестве набора данных дневные цены закрытия акций компании Tesla [14], а М. Агравал, П. Шукла и коллеги в своей научно-исследовательской работе используют не только котировки акций, но и различные показатели технических индикаторов, что увеличивает объем параметров, которые модель учитывает при составлении прогноза.

В изученной литературе различаются не только используемые при обучении нейронных сетей параметры, но и прогнозируемые акции. Б. Егуз и Т. Кайа в своем исследовании прогнозируют стоимость акций турецких банков [2]. Работа М. Кумара, К. Чандры и коллег посвящена изучению данных о цене нефтяных акций IND в режиме реального времени и прогнозированию их изменения с учетом влияния нефтяных индексов [6]. В научно-исследовательской работе Серрано В. проведен анализ эффективности использования нейронных сетей при прогнозировании рынка недвижимости, акций и финтеха [11]:

- цены на недвижимость в Великобритании;
- цены индексов фондового рынка;
- цены криптовалют.

Прогнозирование цен на фондовом рынке вызывает все большую озабоченность в секторах с дивидендной доходностью. Исследование С. Раджесвара и коллег связано с применением алгоритмов обучения Левенберга-Марквардта, линейной регрессии и K -ближайших соседей для прогнозирования временных рядов. Точность прогноза при исследовании составила 82,58% [9].

М. Гани, М. Аваис совместно с коллегами в своей работе освещают вопрос важности анализа данных и их обработки в машинном обучении. Согласно авторам, анализ данных представляет собой процесс выделения сигналов для принятия решений со статистическими колебаниями результатов [3]. В работе также отмечается, что целостность и точность — две основные составляющие анализа статистических данных. Исследуемые наборы данных собраны авторами при использовании API таких сервисов, как Yahoo Finance, Quandle и Kaggle.

Линейная регрессия и «дерево решений»

Статистические методы используются для интерпретации данных множество лет. Линейная регрессия важна для оценки данных и установления определенной взаимосвязи между двумя и более переменными. Авторами было принято решение использовать модель линейной регрессии в качестве базового и фундаментального метода, послужившего основой для разработки более «нагруженных» моделей, таких как «дерево решений» и других.

Один из методов прогнозирования котировок заключается в построении прямой линии вида $y = mx + c$ на графике таким образом, чтобы линия проходила через максимальное количество точек имеющегося набора данных. Говоря математическим языком, при нанесении значений набора данных на график, необходимо провести прямую линию через точки так, чтобы квадрат расстояния между каждой точкой и линией был минимальным. Эта линия используется для прогнозирования значения y для любого заданного значения x . Данный метод прогнозирования называется линейной регрессией, в рамках которой используется функция линейной регрессии (см. уравнение 1). Этот метод используется и в статистике, и как одна из основных концепций машинного обучения.

$$y = h_{\theta}(x) = \theta_0 + \theta_1 x, \quad (1)$$

где y — «ожидаемое» значение,

x — независимая переменная, являющаяся единичным значением массива данных, используемых при прогнозировании,

$h_{\theta}(x)$ — возможные значения h , когда x ограничен некоторым единственным значением,

θ_0 — параметр перехвата,

$\theta_1 x$ — параметр наклона.

Большим преимуществом модели линейной регрессии и ее разновидностей является ее линейность, которая делает процедуру оценки простой и понятной для интерпретации. Это одна из основных причин, почему линейная и подобные модели очень актуальны в академических областях — таких, как медицина, социология, психология и других. В научно-исследовательских работах по экономическим тематикам данные модели и их разновидности также часто используются для прогнозирования различных финансовых показателей компаний.

«Дерево решений» — это непараметрический метод контролируемого обучения модели, используемый для классификации и регрессии. Цель состоит в том, чтобы создать модель, которая предсказывает значение целевой переменной путем обучения простым правилам принятия решений, выведенным при оценке данных. «Дерево решений» представляет собой древовидный классификатор, в котором внутренние узлы представляют признаки набора данных, «ветви» представляют правила принятия решений, а каждый «лист» представляет результат. Преимущества использования модели:

- простота понимания и интерпретация, дерево решений можно визуализировать;
- не требуется особой подготовки данных, в то время как другие модели часто требуют нормализации данных, создания фиктивных переменных и удаления пустых значений. Необходимо также обращать внимание на то, что модель не поддерживает отсутствующих значений;
- стоимость использования «дерева решений» выражается в объеме затрачиваемых вычислительных мощностей и логарифмически зависит от количества точек данных, используемых для обучения «дерева»;
- модель способна обрабатывать как числовые, так и категориальные данные (другие методы обычно специализируются на анализе наборов данных, содержащих только один тип переменных);
- возможность проверки модели с помощью статистических методов.

К недостаткам использования «дерева решений» можно отнести:

- модель может создавать слишком сложное «дерево», которое плохо обобщает данные. Это называется избыточной подгонкой. С целью предотвращения данной проблемы необходимо использовать такие инструменты, как обрезка, установка минимального количества образцов, необходимых в узле «листа», или установка максимальной глубины «дерева»;
- «дерево решений» может быть нестабильным, поскольку небольшие изменения в данных могут привести к созданию совершенно другого «дерева»;
- прогнозы «дерева решений» не являются ни гладкими, ни непрерывными, а представляют собой кусочно-постоянные аппроксимации, поэтому они плохо поддаются экстраполяции;
- обучающее «дерево» создает смещенное «дерево», если некоторые классы доминируют, поэтому необходимо сбалансировать набор данных перед установкой «дерева решений».

С целью реализации процесса сбора, обработки, анализа и прогнозирования тренда движения котировок акций ПАО НОВАТЭК, авторами использованы следующие средства:

- язык программирования *python* версии 3.9.6 – удобный и популярный язык программирования, используемый для решения широкого спектра задач из различных сфер программирования. Подходит для работы на различных операционных системах: Mac, Windows и Linux;
- *python pandas* версии 1.5.2 – быстрая и мощная библиотека с открытым исходным кодом, предназначенная для анализа данных и манипулирования ими;
- *python scikit-learn* версии 1.2.1 – всемирно известная библиотека с открытым исходным кодом, используемая для обучения моделей машинного обучения как по встроенным методам, так и с учетом персональных разработок;
- *python numpy* версии 1.21.4 – библиотека с открытым исходным кодом, позволяющая применять широкий выбор математических функций, методов линейной алгебры, генераторов чисел и иных методов для работы с числовыми данными;
- *python requests* версии 2.26.0 – HTTP-библиотека, построенная на основе языка программирования *python*. Используется для сбора данных о котировках компании;
- *python matplotlib* версии 3.7.0 – библиотека языка *python*, используемая для визуализации данных.

Общий процесс сбора, обработки, анализа данных и их прогнозирования состоит из ряда шагов, отраженных на рисунке (рис. 1).

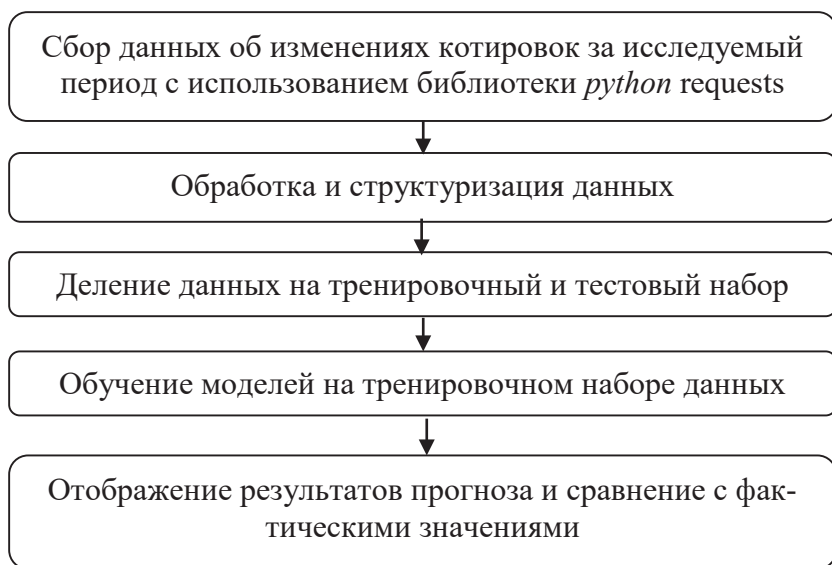


Рис. 1. Общая схема процесса прогнозирования котировок акций с использованием моделей дерева решений и линейной регрессии.

Источник: составлено авторами.

Для обучения моделей используются собранные с использованием средств языка программирования *python* данные по изменению котировок ПАО НОВАТЭК. Однодневные котировки отражены за период с 18.08.2006 по 06.04.2023, который включает в себя 4138 рабочих дней. Для обучения моделей используются значений котировок на момент закрытия (табл. 1).

Таблица 1

Пример набора данных с дневными котировками ПАО НОВАТЭК

datetime	ticker	exchange	interval	open	high	low	close
2006-08-18 06:00:00	Novatek	MOEX	1d	133,35	133,35	127,01	127,11
2006-08-21 06:00:00	Novatek	MOEX	1d	132	132,13	125,01	130
2006-08-22 06:00:00	Novatek	MOEX	1d	127,7	133	127,7	131,5
2006-08-23 06:00:00	Novatek	MOEX	1d	131,5	139	129,01	135
2006-08-24 06:00:00	Novatek	MOEX	1d	134,01	139	133,5	137,5
2006-08-25 06:00:00	Novatek	MOEX	1d	140	140	136,01	137
...	...	...	...	...	...	...	...
2023-03-29 07:00:00	Novatek	MOEX	1d	1160	1164	1150	1156,6
2023-03-30 07:00:00	Novatek	MOEX	1d	1156,4	1158,8	1151,2	1153,6
2023-03-31 07:00:00	Novatek	MOEX	1d	1154	1158	1135,8	1155,2
2023-04-03 07:00:00	Novatek	MOEX	1d	1165,6	1175	1152	1167,4
2023-04-04 07:00:00	Novatek	MOEX	1d	1171,8	1210,6	1170	1198,2
2023-04-05 07:00:00	Novatek	MOEX	1d	1198,4	1205	1181	1203,6
2023-04-06 07:00:00	Novatek	MOEX	1d	1204	1217,6	1202,4	1211

Источник: составлено авторами по: TradingView. URL: https://ru.tradingview.com/chart/jMH-Dra8v/?sy_mbol=MOEX%3NVTK.

Собираемый с помощью средств языка программирования *python* набор данных включает в себя следующие столбцы:

- datetime – содержит даты торговых дней, по которым были собраны значения котировок;
- ticker – название компании, по акциям которой собраны значения котировок;
- exchange – название биржи, где котируется анализируемая компания;
- interval – интервал свечей, по которым выгружаются значения;
- open – цена открытия интервала;
- high – наивысшая цена в интервале;
- low – низшая цена в интервале;
- close – цена закрытия интервала.

Для обучения моделей и осуществления прогноза используются значения котировок на момент закрытия, данные значения хранятся в столбце *close*. Прогноз будет осуществляться на 14 дней вперед, а для обучающей выборки будет отобрано несколько вариантов количества значений: 5000, 1000 и 300. Обучение моделей происходит средствами библиотеки языка *python* — *scikit-learn* (*sklearn.tree* и *sklearn.linear_model*). Тестирование вариантов с разным количеством значений, направляемых в модели, обусловлено необходимостью проверить, на каком объеме данных и за какой срок модели более эффективны. Графически результаты прогнозирования 14 последующих значений котировок моделью «дерева решений» представлены ниже (см. рис. 2, 3, 4).

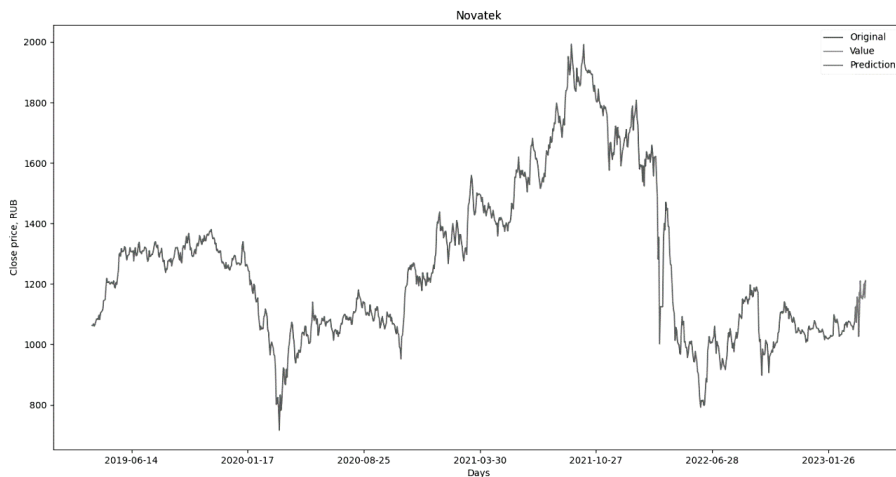


Рис. 2. Результаты прогнозирования значений котировок моделью «дерева решений» при входных параметрах в 5000 значений.

Источник: составлено авторами.

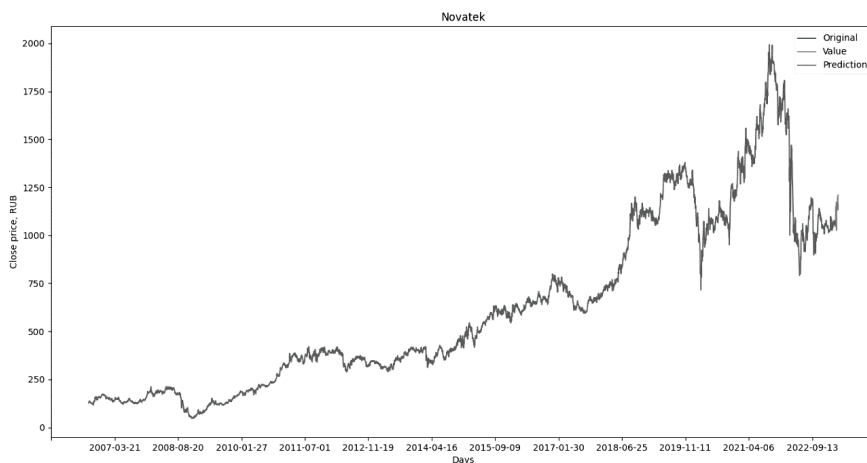


Рис. 3. Результаты прогнозирования значений котировок моделью «дерева решений» при входных параметрах в 1000 значений.

Источник: составлено авторами.

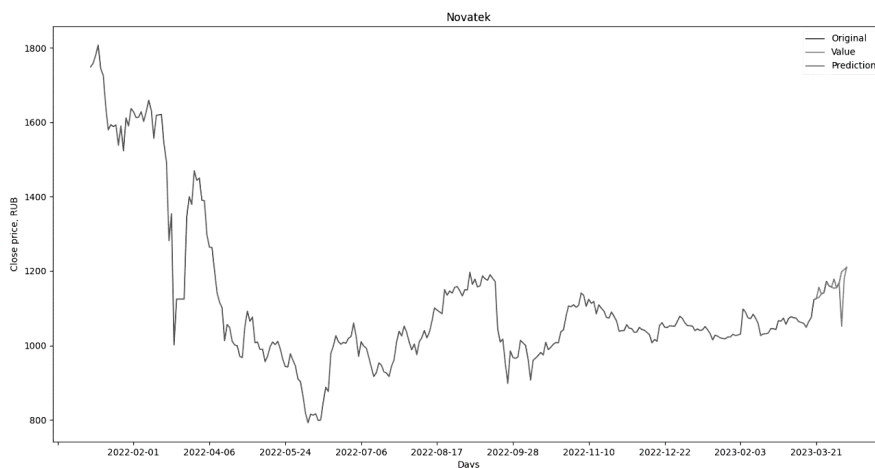


Рис. 4. Результаты прогнозирования значений котировок моделью «дерева» решений при входных параметрах в 300 значений.

Источник: составлено авторами.

При визуальном анализе отмечается, что модель «дерева решений» на разных объемах анализируемых данных проводит прогноз котировок разного качества. Стоит также отметить сложность прогнозирования котировок акций ПАО НОВАТЭК и в целом российских акций в связи с рядом экономических потрясений, произошедших в Российской Федерации и мире в 2014, 2021 и 2022 годах, которые модель не учитывает. Для более подробной оценки эффективности использования модели необходимо проанализировать взаимосвязь между фактическими значениями и прогнозными. Фактические значения содержатся в столбце close, прогнозные значения в столбце predictions и отклонение в столбце deviation (табл. 2, 3, 4).

Таблица 2

Фактические и прогнозные значения модели «дерева решений» при входных параметрах в 5000 значений

datetime	close	predictions	deviation
2023-03-20 07:00:00	1122,20	1122,20	0,00
2023-03-21 07:00:00	1126,20	1034,60	91,60
2023-03-22 07:00:00	1128,80	1156,60	–27,80
2023-03-23 07:00:00	1138,40	989,60	148,80
2023-03-24 07:00:00	1141,80	1025,80	116,00
2023-03-27 07:00:00	1172,00	1126,00	46,00
2023-03-28 07:00:00	1160,00	1134,60	25,40
2023-03-29 07:00:00	1156,60	1156,60	0,00
2023-03-30 07:00:00	1153,60	1139,80	13,80

2023-03-31 07:00:00	1155,20	1134,53	20,67
2023-04-03 07:00:00	1167,40	1137,20	30,20
2023-04-04 07:00:00	1198,20	1198,20	0,00
2023-04-05 07:00:00	1203,60	1139,80	63,80
2023-04-06 07:00:00	1210,00	1134,60	75,40

Источник: составлено авторами.

Таблица 3

Фактические и прогнозные значения модели «дерева решений» при входных параметрах в 1000 значений

datetime	close	predictions	deviation
2023-03-20 07:00:00	1122,20	1122,20	0,00
2023-03-21 07:00:00	1126,20	1045,90	80,30
2023-03-22 07:00:00	1128,80	1156,60	–27,80
2023-03-23 07:00:00	1138,40	989,60	148,80
2023-03-24 07:00:00	1141,80	1083,80	58,00
2023-03-27 07:00:00	1172,00	1092,67	79,33
2023-03-28 07:00:00	1160,00	1160,00	0,00
2023-03-29 07:00:00	1156,60	1156,60	0,00
2023-03-30 07:00:00	1153,60	1178,60	–25,00
2023-03-31 07:00:00	1155,20	1150,10	5,10
2023-04-03 07:00:00	1167,40	1167,40	0,00
2023-04-04 07:00:00	1198,20	1198,20	0,00
2023-04-05 07:00:00	1203,60	1178,60	25,00
2023-04-06 07:00:00	1210,00	1160,00	50,00

Источник: составлено авторами.

Таблица 4

Фактические и прогнозные значения модели «дерева решений» при входных параметрах в 300 значений

datetime	close	predictions	deviation
2023-03-20 07:00:00	1122,20	1122,20	0,00
2023-03-21 07:00:00	1126,20	1072,67	53,53
2023-03-22 07:00:00	1128,80	1142,70	–13,90
2023-03-23 07:00:00	1138,40	1138,40	0,00
2023-03-24 07:00:00	1141,80	1141,80	0,00
2023-03-27 07:00:00	1172,00	1172,00	0,00

2023-03-28 07:00:00	1160,00	1160,00	0,00
2023-03-29 07:00:00	1156,60	1142,70	13,90
2023-03-30 07:00:00	1153,60	1178,60	–25,00
2023-03-31 07:00:00	1155,20	1155,20	0,00
2023-04-03 07:00:00	1167,40	1167,40	0,00
2023-04-04 07:00:00	1198,20	1051,20	147,00
2023-04-05 07:00:00	1203,60	1178,60	25,00
2023-04-06 07:00:00	1210,00	1210,00	0,00

Источник: составлено авторами.

Далее проведем обучение и прогноз 14 следующих значений с использованием модели линейной регрессии. Линейная модель требует идентичной подготовки данных, как и для «дерева решений», что позволяет технически облегчить задачу прогнозирования, а также экономить вычислительные мощности при использовании обеих моделей. Данная ситуация особенно удобна для аналитических компаний, осуществляющих непрерывно большое количество прогнозов по разным котировкам и таймфреймам в условиях ограниченных техническими характеристиками вычислительных мощностей и баз данных. Также открывается возможность оптимизации программной архитектуры аналитического и прогнозного модулей программы. Графические результаты прогнозов с использованием модели линейной регрессии при разном объеме входных данных отражены ниже (рис. 5, 6, 7).

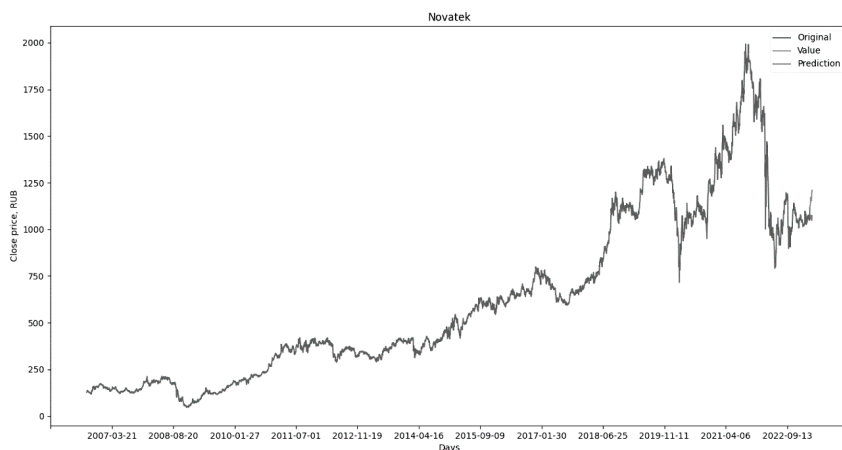


Рис. 5. Результаты прогнозирования значений котировок моделью линейной регрессии при входных параметрах в 5000 значений.

Источник: составлено авторами.

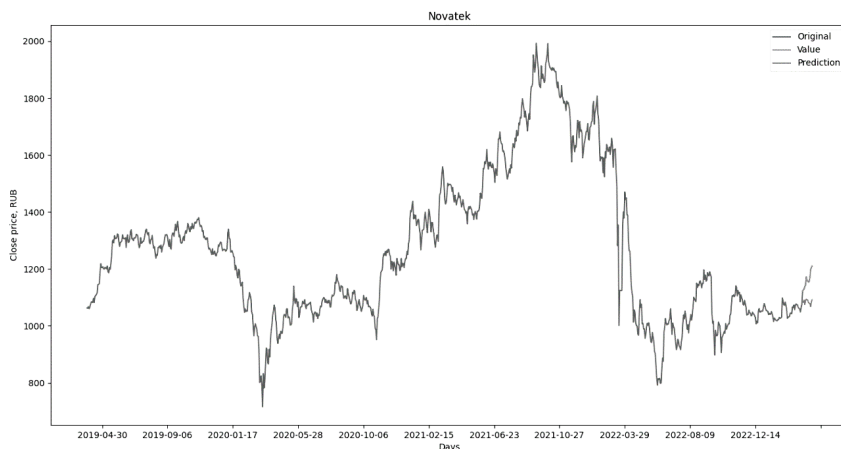


Рис. 6. Результаты прогнозирования значений котировок моделью линейной регрессии при входных параметрах в 1000 значений.

Источник: составлено авторами.

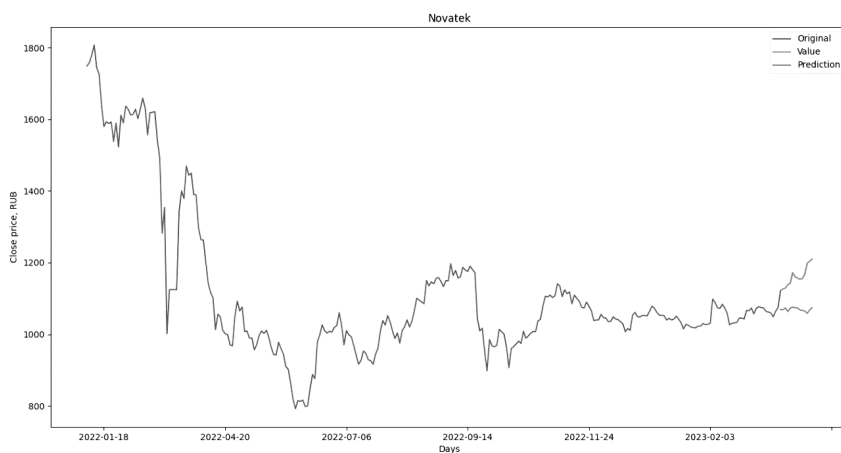


Рис. 7. Результаты прогнозирования значений котировок моделью линейной регрессии при входных параметрах в 300 значений.

Источник: составлено авторами.

Необходимо отметить, что уже при визуальной оценке качества прогноза котировок с использованием модели линейной регрессии отмечается низкое качество прогноза и высокие значения отклонений. В то же время более сложная модель «дерева решений» показывает себя лучше, прогнозируя приближенные к фактическим значения. В связи с этим стоит отметить лучшую эффективность у модели «дерева решений», чем у модели линейной регрессии.

Кроме проведения прогнозов необходимо иметь возможность оценить их результативность. Для оценки эффективности прогнозов используются такие метрики, как MAPE и WAPE. MAPE (mean absolute percentage error) — один из наиболее распространенных методов определения точности прогноза, измеряющий процентную ошибку прогноза по отношению к фактическим значениям. WAPE (weighted absolute percent error) представляет собой разновидность MAPE и чаще используется на практике в связи с возможностью учесть вес каждой из частей уравнения. Для оценки точности прогнозирования будем использовать показатель взвешенной абсолютной процентной ошибки (WAPE) (см. уравнение 2).

$$\text{Ошибка (WAPE)} = \frac{\sum |\text{Факт} - \text{Прогноз}|}{\sum \text{Факт}} * 100 \quad (2)$$

Точность прогнозирования рассчитывается как вычитание из 100 процентов процента ошибки. Показатели точности прогнозирования в разрезе модели и количества входных данных отражены ниже (см. табл. 5).

Таблица 5

Показатели точности прогнозирования используемых моделей

Модель	Количество входных данных	WAPE	Точность прогнозирования
«Дерево решений»	5000	3,7	96,3
	1000	2,4	97,6
	300	1,2	98,8
Линейная регрессия	5000	7,9	92,1
	1000	6,2	93,8
	300	8,8	91,2

Источник: составлено авторами.

Анализируя данные, отметим большую точность прогнозирования при использовании модели «дерева решений». Ориентируясь на данную модель, можно определять с высокой точностью как будущие значения котировок, так и направление движений котировок. При использовании данной модели в промышленных масштабах необходимо проведение более широких тестов с разными сочетаниями количества данных, котировок и таймфреймов. Также стоит отметить, что в модели «дерева решений» точность прогноза увеличивается с сокращением количества входных данных. Данный факт необходимо учитывать при разработке и использовании данной прогнозной модели в деятельности компаний. Необходимо проводить ряд тестовых мероприятий,

направленных на выведение оптимального количества входных данных при обучении модели.

С целью дополнительной оценки эффективности моделей, рассчитаем потенциальную экономическую выгоду от торговых сделок, осуществляемых на основе проведенных прогнозов. Будем считать, что торговая позиция открывается по цене закрытия последнего дня перед прогнозом и закрывается по цене последнего предсказанного дня. Объем каждой позиции составляет 1000 акций. Результаты расчетов приведены ниже (табл. 6).

Таблица 6

Результаты эффективности моделей при торговле на фондовом рынке

Модель	Количество входных данных	Цена от-кры-тия	Цена за-крытия, факт	Цена за-крытия, прогноз	При-быль, факт	При-быль, про-гноз	Разница (факт—прогноз)
Дерево реше-ний	5000	1074,6	1210	1134,6	135400	60000	75400
	1000	1074,6	1210	1160,0	135400	85400	50000
	300	1074,6	1210	1210,0	135400	135400	0
Линей-ная ре-грессия	5000	1074,6	1210	1075,8	135400	1200	134200
	1000	1074,6	1210	1095,3	135400	20699	114701
	300	1074,6	1210	1067,5	135400	–7100	142500

Источник: составлено авторами.

Анализируя данные, отметим, что модель «дерева решений» значительно эффективней модели линейной регрессии и даже на объеме в триста торговых дней смогла спрогнозировать точное значение цены закрытия. Как и было отмечено при расчете показателя WAPE, наилучший результат прогноза показала модель «дерева решений» при объеме входных данных в 300 торговых дней, а модель линейной регрессии пока показала наилучший результат при объеме входных данных в 1000 торговых дней. Таким образом, в случае организации автоматической торговли на основе прогноза нейронных сетей, более эффективной была бы торговля с использованием «дерева решений».

Использование нейронных сетей при прогнозировании изменений котировок на фондовом рынке стало важным инструментом поддержания и усиления конкурентоспособности многих компаний, связанных с фондовым рынком. Данная ситуация, в свою очередь, подталкивает компании и научно-исследовательские центры к разработке собственных моделей нейронных сетей на основе новых подходов или на основе доработки уже существующих решений. Во время создания эффективного механизма прогнозирования компании тестируют различные модели нейронных сетей, включая «дерево решений», линейную регрессию, их доработанные аналоги и иные виды моделей.

Использование базовых моделей нейронных сетей позволяет понять и изучить фундаментальные основы, на которых строятся более сложные и высоконагруженные модели нейронных сетей. Проведение прогнозов с использованием основных моделей нейронных сетей с целью определения их эффективности также обусловлено ограниченностью вычислительных мощностей и их дороговизной. Необходимо также отметить, что при прогнозировании котировок на фондовом рынке необходимо тестировать модели на разном объеме входных данных и выбирать оптимальные варианты.

Фондовый рынок очень изменчив, однако он дает инвесторам огромные возможности по увеличению собственных капиталов. Для прогнозирования котировок с целью планирования торговой стратегии инвесторы в настоящее время обращаются не только к техническому и фундаментальному анализу, но и к различным моделям машинного обучения. Модели могут быстро обрабатывать исторические данные, линии тренда, графики и иные показатели, предлагая в результате прогноз будущих котировок. Машинное обучение в настоящее время является революционной технологией и выступает в качестве одного из важных инструментов повышения конкурентоспособности компаний на рынке. На выбор модели или набора моделей, используемых в компании, влияет ряд факторов: таймфреймы, с которыми работает компания, количество и качество входных данных для моделей, технические характеристики вычислительных мощностей, используемых в компании.

Литература

1. *Agrawal M., Shukla P., Nair R., Nayyar A.* Stock prediction based on technical indicators using deep learning model // Tech Science Press. 2022. No. 1. Pp. 287–304.
2. *Eguz B., Kaya T.* Stock price prediction of Turkish banks using machine learning methods // International conference on intelligent and fuzzy systems. 2021. No. 308. Pp. 1–17.
3. *Ghani M., Awais M., Muzammul M.* Stock market prediction using machine learning (ML) algorithms // ADCAIJ: advances in distributed computing and artificial intelligence journal. 2019. No. 4. Pp. 97–116.
4. *Gururaj V., Ashwini K.* Stock market prediction using linear regression and support vector machines // International journal of applied engineering research. 2019. No. 14. Pp. 1931–1934.
5. *Hota J., Chakravarty S., Paikaray B., Bhoyar H.* Stock market prediction using machine learning techniques // CEUR. 2022. No. 3283. Pp. 163–171.
6. *Kumar M., Chandra K., Gupta K.* Stock analysis and prediction of Indian oil trading using big data analytics // International journal of mechanical engineering. 2022. No. 1. Pp. 6734–6738.
7. *Ly P., Wu Q., Xu J.* Stock index prediction based on time series decomposition and hybrid model // Entropy. 2022. No. 24. Pp. 1–18.
8. *Majumder A., Rahman M., Biswas A.* Stock market prediction: a time series analysis // Smart systems: innovations in computing. 2021. No. 235. Pp. 389–401.
9. *Rajeswar S., Ramalingam P., Sudalai T.* Comparative analysis of stock market price behaviors using machine learning techniques // International conference on advances in materials, computing and communication technologies. 2020. No. 2385. Pp. 25–37.

10. *Raubitzek S., Neubauer T.* An exploratory study on the complexity and machine learning predictability of stock market data // *Entropy*. 2022. No. 332. Pp. 1–34.
11. *Serrano W.* The random neural network in price predictions // *Neural computing and applications*. 2022. No. 34. Pp. 855–873.
12. *Shah J., Vaidya D., Shah M.* A comprehensive review on multiple hybrid deep learning approaches for stock prediction // *ELSEVIER*. 2022. No. 16. Pp. 1–14.
13. *Singh G.* Machine learning models in stock market prediction // *International Journal of innovative technology and exploring engineering (IJITEE)*. 2022. No. 11. Pp. 18–28.
14. *Singh S., Rehan S., Kumar V.* Stock price prediction using linear regression, LSTM and decision tree // *EasyChair Preprint*. 2022. No. 7805. Pp. 1–5.
15. *Zhao A., Cheng T.* Stock return prediction: stacking a variety of models // *ELSEVIER*. 2022. No. 67. Pp. 1–12.

Roman Kuznetsov (e-mail: romak2921@mail.ru)

graduate student, St. Petersburg State University of Economics,
(St. Petersburg, Russian Federation)

Tatiana Tumarova (e-mail: tumarova@finec.ru)

Ph.D. in Economics, Professor, Director of the Graduate Institute of Saint-Petersburg
State University of Economics (Saint-Petersburg, Russian Federation)

PREDICTING NOVATEK SHARE PRICES USING THE MODELS OF DECISION TREE AND LINEAR REGRESSION

The development of technology and the emergence of various machine learning models influence social life in many ways, including the analysis and forecasting of the stock market. The ability to competently select and use machine learning models in predicting stock quotes is one of the key competitive advantages that allow large investment companies and individuals to increase their profits from the market activity. The study reveals the effectiveness of using decision tree and linear regression models in predicting daily quotes of NOVATEK.

Keywords: stock exchange, decision tree, linear regression, machine learning.

DOI: 10.31857/S020736760026686-3