

УДК 004.02 + 004.855.5
DOI: 10.15827/0236-235X.139.396-412

Дата подачи статьи: 28.02.22, после доработки: 06.06.22
2022. Т. 35. № 3. С. 396–412

Прогнозирование технологических тенденций на основе анализа разнородных данных

Нгуен Тхань Вьет^{1,2}, аспирант, vietqn1987@gmail.com
А.Г. Кравец^{1,3}, д.т.н., профессор, agk@gde.ru

¹ Волгоградский государственный технический университет,
г. Волгоград, 400005, Россия

² Фам Ван Донг университет, г. Куанг Нгай, Вьетнам

³ Государственный университет «Дубна», г. Дубна, 141980, Россия

Для достижения конкурентоспособности предприятиям необходимо разработать стратегию развития путем прогноза новых и перспективных технологий при использовании ограниченных ресурсов. Многочисленные предыдущие исследования показали, что неожиданные изменения в направлениях НИОКР и патентования результатов интеллектуальной деятельности связаны с большими изменениями рыночной стоимости фирмы. Существует сильная корреляция между изменчивостью рыночных долей, цен на акции на ранних стадиях развития высокотехнологичного предприятия и тем периодом, когда инновационная технология еще не определена. По этой причине можно предположить: если цены акций предприятия сохраняют восходящий тренд, то разработанные технологии, скорее всего, станут многообещающими инновациями в будущем.

В статье предложен метод прогнозирования технологических трендов путем анализа веб-новостей для выявления высокотехнологичных предприятий, прогнозирования тенденций изменения стоимости акций для выбранных предприятий и анализа кластеризации патентных заявок. В отличие от других исследований предлагаемый метод продвигает идею прогнозирования технологических трендов путем прогнозирования (восходящей/нисходящей) тенденции стоимости акций с использованием одномерных и многомерных подходов к подготовке данных и с помощью байесовской оптимизации для нахождения лучших гиперпараметров моделей машинного и глубокого обучения.

Исследование выполнено с помощью разработанной программной системы для анализа обнаружения взрыва частоты слов и прогнозирования цен акций. В частности, предлагаемый метод рассмотрен для прогнозирования динамики цен акций компаний Tesla и Samsung в качестве примера.

Ключевые слова: интеллектуальный анализ данных, технологический прогноз, цена акций, Tesla, байесовская оптимизация, машинное обучение, глубокое обучение, патентный анализ, кластерный анализ, VOSviewer.

Выявление как можно раньше будущих тенденций изменения новых технологий имеет решающее значение для стратегического планирования государственных и корпоративных исследований и разработок (НИОКР), чтобы получить преимущество первопроходца в конкурентной бизнес-среде. Многие руководители организаций осознают важность понимания будущих тенденций развития новых технологий для устойчивого развития [1].

С другой стороны, благодаря промышленным революциям технологии привели к быстрым изменениям в обществе, причем финансы вкладываются туда, где есть полезность, экономическая выгода. Технологии ускоряют производство, улучшают качество жизни, делают многие аспекты деятельности более эффективными. Одной из областей, в которых произошла революция благодаря технологическим достижениям, являются финансовые рынки.

Есть немало примеров, как технологии изменили и сформировали текущее состояние финансовых рынков, а также будущее направление их развития.

Например, в качестве объектов исследования [2] были выбраны 73 зарегистрированные на бирже компании, опубликовавшие объявления о применении блокчейна. Метод анализа краткосрочных событий использован для количественного анализа влияния применения технологии блокчейна на рыночные показатели этих компаний. Результаты показали, что после объявления о применении блокчейна в собственном бизнесе исследуемые компании получили более высокий доход. Таким образом, рынок капитала имеет существенный положительный отклик на применение блокчейна. В то же время применение анализа событий позволяет количественно оценить влияние появления или применения новых технологий на цену

акций компании. Например, некоторые ученые [3] эмпирически проверяют влияние на доходность акций внедрения 3D-печати, облачных вычислений и политики интеллектуальной логистики.

Если цены на акции отражают прогноз будущих прибылей, то следует ожидать взаимосвязи инноваций, которые в случае успеха могут оказать положительное влияние на рост фирмы и цен на акции. На самом деле существует сильная корреляция между изменчивостью рыночных долей, цен на акции на ранних стадиях развития высокотехнологичного предприятия и тем периодом, когда инновационная технология еще не определена. Другими словами, изменчивость цен на акции отражает динамику риска и неопределенности, связанную с инновациями [4].

В ряде литературных источников подчеркивается связь уровня цен на акции с инновациями, а также рыночной стоимости с патентами. Это объясняется тем, что если патентная статистика отражает изменения технологических возможностей, то ее следует соотносить с текущими изменениями рыночной стоимости, поскольку рыночная стоимость определяется ожиданиями относительно будущего роста. Некоторые авторы исследовали взаимосвязь количества успешных патентных заявок фирм, показателя инвестиций фирмы в изобретательскую деятельность и показателя ее изобретательской продукции. Результат показал, что неожиданные изменения в патентах и НИОКР действительно связаны с большими изменениями рыночной стоимости фирмы.

В частности, в [5] исследована взаимосвязь патентной активности высокотехнологичных компаний и динамики движения цен их акций. Кроме того, продемонстрированы взаимосвязь доходности акций с поправкой на рынок и количества патентных заявок, а также разнообразие соответствующих категорий патентов. Применяв метод скользящего окна, авторы создали приспособленные модели с использованием различных терминов запаздывания характеристик патентной активности. Результаты подтверждают значительное влияние патентной активности на движение акций и на их основные статистические характеристики изменчивости. Более того, в [6] заявлено, что обратные ссылки на патенты, патентные заявки и выданные патенты в некоторой степени воздействуют на объем торгов, так как каждый из этих параметров может играть роль переменной поглощающей способности по мере связи

нового информационного потока с экономическими инновациями.

Точно так же эмпирические данные о цене акций показывают, что фирмы, успешно инвестирующие в радикальные инновации, получают более высокую доходность акций. В работе [7] представлена модель, выявившая взаимосвязь уровня инноваций фирмы и доходности акций, динамика которых отражает ожидания в отношении прибыльности и роста фирмы. Результаты показывают, что предположение о неприятии неопределенности имеет решающее значение для определения более высокой отдачи при наличии радикальных инноваций и конкретизация ожидаемой полезности устанавливает степень отдачи.

В работе [8] показано, что цена акций компании, использующей определенную технологию, будет меняться в зависимости от жизненного цикла технологии. В частности, автор отфильтровал компании, которые в основном занимаются дополненной реальностью и котируются на корейском рынке KOSDAQ. Затем сгруппировал эти компании на основе конкретных технологий дополненной реальности. Автор использовал метод исследования событий для расчета доходности акций по отношению к контрольному показателю. В результате на этапе «Пик завышенных ожиданий» портфели всех компаний, использующих дополненную реальность, в целом показали более высокую доходность, чем контрольный показатель.

В статье [9] исследуется связь между инновационной активностью крупнейших корпоративных инвесторов в НИОКР по всему миру и их рыночной оценкой. В анализе используется выборка из более чем 1250 зарегистрированных на бирже мультинациональных корпораций и их прав на интеллектуальную собственность – патенты и товарные знаки, зарегистрированные в период с 2005 по 2012 гг. Кроме того, исследование позволяет дифференцировать эффекты увеличения рыночной стоимости, получаемой от дополнительных прав на интеллектуальную собственность.

Исследования [10, 11] посвящены обзору степени разработанности анализа и прогнозирования технологических тенденций. К тому же в них утверждается, что веб-новости являются важным ресурсом данных для изучения общественной осведомленности о новых технологиях. Веб-новости, которые относятся к недавним изменениям, обновляются в реальном времени и размещают появляющиеся в Интернете обзоры, что делает их ценными для

изучения будущих тенденций развития технологий. Алгоритм обнаружения взрывов частоты слов и кластеризация обнаруженных терминов позволили выявить значимые предприятия и основные технологические тенденции, наблюдаемые с января 2016 г. по январь 2020 г.

Таким образом, на основании приведенных исследований авторы выдвинули следующую гипотезу: если предсказательная модель показывает восходящий тренд акций некоторых высокотехнологичных предприятий, наиболее вероятно, что разработанные ими технологии станут многообещающими инновациями в будущем. Целью является разработка метода прогнозирования технологических тенденций на основе анализа разнородных данных. Метод включает следующие шаги:

- интеллектуальный анализ веб-новостей для извлечения значимых предприятий;
- прогнозирование тенденций стоимости акций выбранных предприятий;
- патентный анализ предприятий с ценой акций в восходящем тренде.

Общая процедура метода показана на рисунке 1.

Обзор исследований по прогнозированию цен акций

Прогнозирование фондового рынка – классическая, но достаточно сложная задача, привлекающая внимание экономистов и специалистов по информатике. В частности, если цель состоит в том, чтобы предсказать конкретное значение цены, то определяем ее как задачу регрессии, а если предсказать направление движения цены – как задачу классификации. В большинстве исследований рассматривается ежедневный прогноз и лишь в некоторых внутрисуточных, например, 5-минутный или часовой. Для построения эффективной модели прогнозирования в течение нескольких последних десятилетий изучались инструменты как эконометрического моделирования, так и машинного обучения.

В последнее время модели глубокого обучения были представлены как новые рубежи для решения задач анализа и прогноза временных рядов с быстрым развитием [12, 13]. Благодаря высокой эффективности работы с большими данными и изучения нелинейной связи между

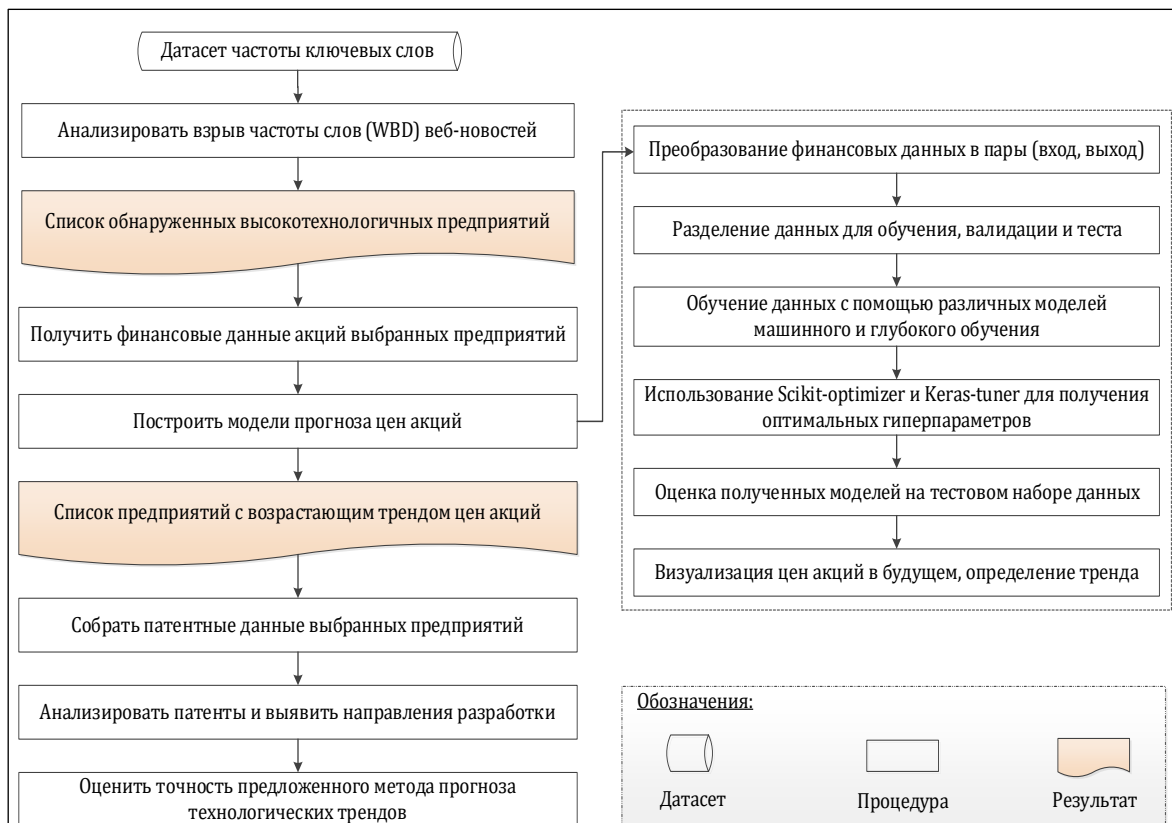


Рис. 1. Общая процедура предлагаемого метода прогнозирования тренда технологий

Fig. 1. General procedure of the proposed method for technology trend forecasting

входными и целевыми (выходными) переменными прогнозирования, модели глубокого обучения показали лучшую производительность, чем модели линейного и машинного обучения, в задачах прогнозирования фондового рынка [14, 15]. Производительность нейросети с архитектурой LSTM была выше, чем у традиционных моделей машинного обучения и моделей без временных рядов. В [16] описана модель глубоко ответственного инвестиционного портфеля (Deep Responsible Investment Portfolio) и сравнивается влияние архитектур LSTM, двунаправленной (Bi-LSTM) и GRU (Gated Recurrent Unit) на прогноз цен акций. Авторы обнаружили, что модель Bi-LSTM считывает данные на один шаг назад, в результате чего повышается точность прогнозов, особенно при прогнозировании последовательных данных, таких как финансовые временные ряды.

В исследованиях [17, 18] делается обзор работ по моделям глубокого обучения и современным методам для прогнозирования фондового рынка. Авторы не только анализируют современные методы, классифицируют различные источники данных, структуры нейронных сетей и часто используемые метрики оценки, но и оценивают их реализацию и воспроизводимость. При этом комбинация нескольких моделей обычно имеет лучшую производительность, чем одна модель, и гибридизация становится основным направлением в прогнозировании цен акций [14, 17, 19].

В [20] предложена система прогнозирования цен на акции, полученная путем объединения SVR и ансамблевой адаптивной нейро-нечеткой системы вывода ENANFIS. Экспериментальные результаты показали, что модель

SVR-ENANFIS имеет более высокую производительность прогнозирования, чем ENANFIS, SVR-Linear, SVR-SVR и SVR-ANN (искусственные нейронные сети). В [21] описана разработанная финансовая экспертная система, которая включает исторические цены акций, восемь видов технических индикаторов, подсчеты и оценки настроений в опубликованных новостных статьях, тенденции в поиске Google и информацию из Википедии для прогнозирования краткосрочных цен на акции. Авторы [22] использовали 44 технических индикатора в своих моделях гибридных программных вычислений для предсказания цен акций на 1, 2, 3, 5, 7 и 10 дней вперед. В [23] описана разработанная обучающая архитектура LR2GBDT для прогнозирования и торговли фондовыми индексами, в качестве исходных переменных добавлены 12 технических индикаторов.

Таким образом, эконометрические модели и модели машинного обучения являются наиболее широко используемыми методами прогнозирования цен акций. Однако глубокое обучение – это быстро развивающаяся технология, которая может эффективно обрабатывать данные временных рядов, в том числе за несколько периодов.

Предлагаемый метод и эмпирические результаты прогнозирования цен акций

В ходе исследования была разработана программная система для анализа обнаружения взрыва частоты слов и прогнозирования временного ряда стоимости акций, информационно-функциональная схема которой представлена на рисунке 2.

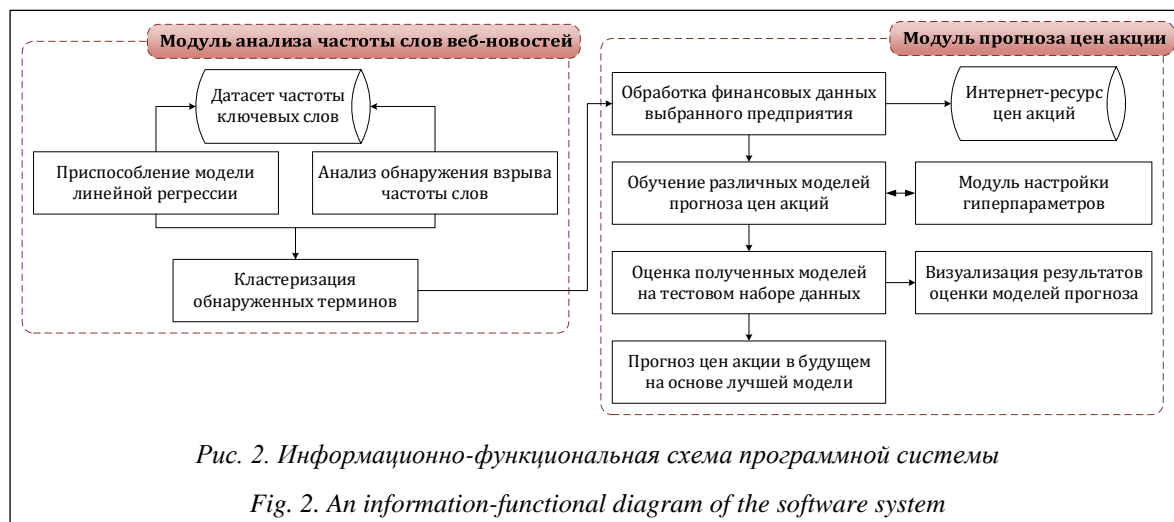


Рис. 2. Информационно-функциональная схема программной системы

Fig. 2. An information-functional diagram of the software system

Программная система состоит из двух модулей: анализа частоты и кластеризации слов и прогноза цен на акции.

Программный код написан на языке Python с поддержкой популярных библиотек машинного/глубокого обучения, таких как Scikit-learn, Keras, Pandas, Matplotlib, Seaborn, Numpy и др. Все эмпирические запуски проводились в интерактивной среде Google Colab. Более подробно модуль анализа частоты и кластеризации слов уже описан в статье [10]. В данной работе описываются функции модуля прогноза цен на акции.

В исследовании [10] показано, что веб-новости являются важным ресурсом данных для изучения общественной осведомленности о новых технологиях. Поэтому после применения алгоритма обнаружения взрыва частоты слов к данным веб-новостей и кластеризации обнаруженных ключевых слов были выявлены некоторые значимые предприятия и основные технологические тенденции, которые произошли с января 2016 г. по январь 2020 г. Среди терминов взрыва появились следующие технологические предприятия: Samsung, Huawei, Intel, Uber, Tesla, Apple и Facebook. В качестве примера рассмотрим применение предлагаемого метода для прогнозирования динамики цен акций компаний Tesla и Samsung.

Исходные данные. Доступные данные о ценах акции Tesla (код NASDAQ: TSLA) за период с января 2011 г. по декабрь 2021 г. загружены с Yahoo Finance (<https://finance.yahoo.com/quote/TSLA?p=TSLA>) для дальнейшего построения модели прогнозирования. Конечной целью данного раздела является прогнози-

рование тренда цен акции (восходящего или нисходящего) Tesla в 2022 г. с использованием многочисленных моделей машинного обучения и глубокого обучения. Исходные данные включают в себя все торговые действия, которые происходят на фондовом рынке, то есть цены открытия/цены закрытия/скорректированные цены закрытия/максимальные/минимальные цены и объем торгов. Загруженные данные о цене акции Tesla представлены в таблице 1.

Предобработка данных. Временное окно горизонта прогнозирования используется для обозначения временного отрезка в будущем, которое будет предсказано моделью. Большинство предыдущих исследований сосредоточено на краткосрочном горизонте прогнозирования, например, на один день или пять минут, за некоторыми исключениями для более длительного горизонта, такого как пять или десять дней.

Существуют два подхода к подготовке данных для прогнозирования цены акций: одномерные и многомерные входные данные. В данном исследовании для одномерных входных данных авторы установили временное окно горизонта равным 1 и временной лаг равным 10, как в работе [24]. Это означает, что на основе данных скорректированной цены закрытия за предыдущие 10 дней модель должна предсказать скорректированную цену закрытия на 11-й день. При одномерном подходе к подготовке данных используется только столбец «Скорректированная цена закрытия», а все остальные столбцы (цены открытия/закрытия/максимум/ минимум, объем торгов) исклю-

Таблица 1

Исходные данные о ценах (в долларах США) акции Tesla с 2011 по 2021 гг.

Table 1

Original data of Tesla stock price (USD) from 2011 to 2021

№	Дата	Цена открытия	Максимальная цена	Минимальная цена	Цена закрытия	Скорректированная цена закрытия	Объем торгов
1	03/01/2011	5.368	5.4	5.18	5.324	5.324	6415000
2	04/01/2011	5.332	5.39	5.204	5.334	5.334	5937000
3	05/01/2011	5.296	5.38	5.238	5.366	5.366	7233500
4	06/01/2011	5.366	5.6	5.362	5.576	5.576	10306000
5	07/01/2011	5.6	5.716	5.58	5.648	5.648	11239500
...
2765	27/12/2021	1073.67	1117	1070.72	1093.94	1093.94	23715300
2766	28/12/2021	1109.49	1119	1078.42	1088.47	1088.47	20108000
2767	29/12/2021	1098.64	1104	1064.14	1086.19	1086.19	18718000
2768	30/12/2021	1061.33	1095.55	1053.15	1070.34	1070.34	15680300
2769	31/12/2021	1073.44	1082	1054.59	1056.78	1056.78	13577900

чаются. Этот столбец используется в качестве входных переменных (то есть исторических цен в ретроспективном окне), так и целевых (выходных) переменных прогноза (то есть скорректированной цены закрытия следующего дня).

В то же время для подхода многомерных данных устанавливаем временное окно горизонта равным 250, так как это среднее количество рабочих дней в году (фондовый рынок закрыт в выходные и праздничные дни). Для этого входными являются финансовые данные в течение дня: цены открытия/скорректированные закрытия/максимум/минимум, объем торгов, модель должна предсказать скорректированную цену закрытия на следующий, 250-й день. Разница между двумя подходами к подготовке данных показана на рисунке 3.

Следует отметить, что полученный датасет разделяется на обучающий набор данных с 80 % наблюдений, которые можно использовать для обучения модели (обучающий набор также содержит 20 % данных для валидации), и оставшиеся 20 % – для тестирования модели.

При наличии входных данных с разным масштабом нормализация данных используется, чтобы гарантировать эффективную работу моделей машинного/глубокого обучения, а также повысить скорость и производительность обучения моделей. Таким образом, в ра-

боте нормализация данных относится к процессу изменения масштаба входных данных, чтобы все значения находились в диапазоне от 0 до 1.

Модели машинного/глубокого обучения для прогнозирования цен акции. Данные модели обычно включают два типа не предназначенных для обучения параметров, включая переменные проектирования модели и гиперпараметры. Переменные проектирования модели включают тип оптимизатора, функцию потерь, функцию активации и прочие. Гиперпараметр – это параметр, значение которого используется для управления процессом обучения. Например, он может включать количество слоев, количество нейронов, размер ядра, процент исключения, скорость обучения и т.д. Таким образом, для выбора эффективных гиперпараметров моделей используется байесовская оптимизация с помощью библиотек Scikit-optimize (<https://scikit-optimize.github.io/stable>) и Keras-tuner (https://keras.io/keras_tuner). Целью байесовской оптимизации [25] являются поиск и настройка гиперпараметров данной модели для эффективной работы с валидационным набором данных. Считается, что этот метод оптимизации превосходит такие методы, как случайный поиск или поиск по сетке. Чтобы использовать байесовскую оптимизацию, необходим эффективный метод модели-

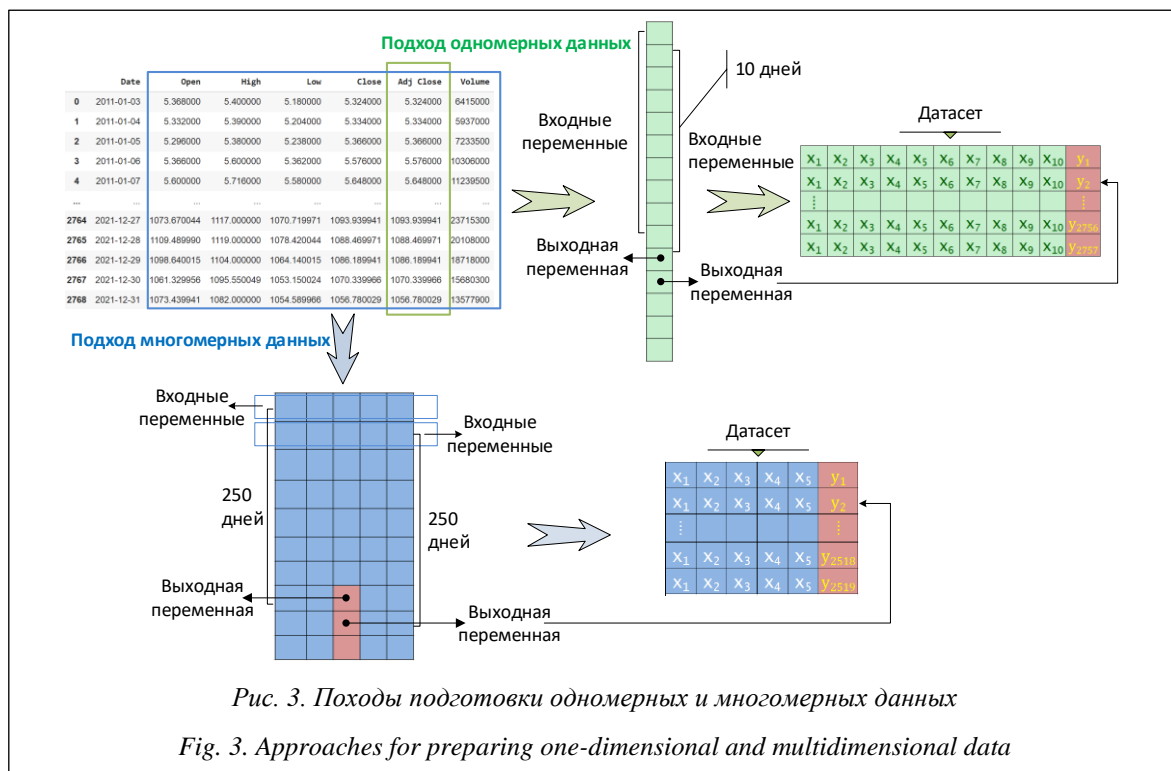


Рис. 3. Подходы подготовки одномерных и многомерных данных

Fig. 3. Approaches for preparing one-dimensional and multidimensional data

рования распределения целевой функции. Подход использует гауссовские процессы для создания вероятностной модели целевой функции. В процессе выборки выявляются более многообещающие гиперпараметры для оценки реальной целевой функции, а предварительные знания используются для аппроксимации апостериорного распределения неизвестной целевой функции для корректировки гиперпараметров. Этот метод может определить лучшие настройки гиперпараметров, чем случайный поиск или поиск по сетке, за меньшее количество итераций.

В данной статье исследованы популярные модели машинного обучения для регрессии, например XGB, GBR, SVR и KNN. Кроме того, довольно хорошие результаты оценки обеспечивают и перечисленные далее модели.

1. *Линейная регрессия с регуляризацией Elastic Net.* Линейная регрессия – это классическая линейная модель, которая пытается согласовать взаимосвязь прогнозируемой цели и входных переменных с помощью линейной модели, в которой параметры могут быть обучены методом наименьших квадратов. При подборе моделей линейной или логистической регрессии Elastic Net (https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.ElasticNet.html) представляет собой регуляризованный регрессионный метод, который линейно комбинирует штрафы L1 и L2 методов Lasso и Ridge. Пространство поиска гиперпараметров модели следующее: $\alpha \in \{0.01, 0.02, \dots, 0.99, 1\}$, $l1_ratio$ – случайное действительное число в диапазоне $[0, 1]$.

2. *Нейронная сеть прямого распространения FFNN.* Это простейший тип искусственной нейронной сети, в которой связи между узлами не образуют цикл. В этой сети информация движется только в одном прямом направлении от входных узлов, через скрытые узлы и к выходным узлам (https://keras.io/guides/sequential_model). Пространство поиска гиперпараметров следующее: количество слоев $\in \{1, 2, 3\}$, количество нейронов $\in \{50, 70, 90\}$ для подхода одномерных данных и $\in \{200, 250, 300, 350, 400\}$ для подхода многомерных данных, процент исключения $\in \{0.01, 0.1\}$.

3. *Сверточная нейронная сеть CNN.* Сеть предназначена для обработки двумерных изображений. Каждая группа нейронов, которую также называют фильтром, выполняет операцию свертки к разным областям входного изображения, нейроны имеют одинаковые веса, что уменьшает количество параметров по сравне-

нию с полносвязной сетью FFNN. Операции объединения, например, максимальное объединение, используются для уменьшения исходного размера и могут применяться несколько раз, пока конечный выход не будет объединен в полносвязный слой. Для сведения операций свертки и объединения к одному временному измерению 1D CNN (https://keras.io/api/layers/convolution_layers/convolution1d) часто предлагается для классификации и прогнозирования временных рядов. Пространство поиска гиперпараметров 1D CNN следующее: количество фильтров $\in \{50, 100, 150, 200, 250\}$ (одномерные данные) и $\in \{300, 350, \dots, 750, 800\}$ (многомерные данные); размер ядра $\in \{2, 3, 4, 5\}$ (одномерные данные) и $\in \{2, 3, 4\}$ (многомерные данные); процент исключения $\in \{0.01, 0.1\}$.

4. *Рекуррентная нейронная сеть RNN.* По сравнению с нейронной сетью FFNN рекуррентная нейронная сеть представляет собой искусственную нейронную сеть, в которой соединения между узлами образуют цикл вдоль временной последовательности, что помогает демонстрировать временное динамическое поведение. Однако на практике обычным сетям RNN мешает проблема исчезающего или взрывного градиента, когда градиенты некоторых весов начинают уменьшаться или увеличиваться, если сеть разворачивается слишком много раз. Сети с долгой краткосрочной памятью LSTM (https://keras.io/api/layers/recurrent_layers/lstm) представляют собой RNN, которые решают проблему исчезающего и взрывного градиента, где скрытый слой заменяется повторяющимися вентилями, называемыми вентилями забывания. В отличие от традиционных RNN сеть LSTM хорошо приспособлена к обучению на задачах классификации, обработки и прогнозирования временных рядов в случаях, когда важные события разделены временными лагами с неопределенной продолжительностью и границами. Пространство поиска гиперпараметров LSTM следующее: для одномерных данных количество единиц в первом и втором слоях $\in \{40, 50, 60\}$, для многомерных данных число единиц $\in \{200, 250, \dots, 500\}$, процент исключения $\in \{0.01, 0.1\}$.

5. *Гибридная модель CNN-LSTM.* CNN имеет свойство, позволяющее извлекать наиболее очевидные характеристики, поэтому она широко используется в разработке признаков. LSTM имеет свойство расширяться в зависимости от временной последовательности и широко используется во временных рядах. По ха-

рактическим CNN и LSTM также тестируется гибридная модель CNN-LSTM для прогнозирования цен акции. Пространство поиска гиперпараметров гибридной модели представляет собой комбинацию пространств поиска в моделях CNN и LSTM.

Метрики оценки. В работе использованы *MAE*, *RMSE* и R^2 в качестве показателей для оценки эффективности моделей прогнозирования. *MAE* и *RMSE* – это меры близости, которые оценивают точность прогнозируемого значения по отношению к фактической цене акции. R^2 измеряет линейную корреляцию между двумя переменными и устраняет влияние размерности на различные проблемы регрессии. В основном необходимо найти модель, которая дает самые низкие значения *MAE*, *RMSE* и наиболее высокий R^2 . Эти три показателя определяются следующими формулами:

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_{true}^{(i)} - y_{predict}^{(i)}|,$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_{true}^{(i)} - y_{predict}^{(i)})^2},$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_{true}^{(i)} - y_{predict}^{(i)})^2}{\sum_{i=1}^m (y_{true}^{(i)} - \bar{y})^2},$$

где $y_{true}^{(i)}$ – истинное значение целевой переменной выборки (sample) i ; $y_{predict}^{(i)}$ – прогнозируемое значение целевой переменной выборки i ; \bar{y} – среднее значение истинного значения целевой переменной для всех выборок; m – количество выборок.

Далее рассмотрим результаты, полученные с помощью различных подходов к подготовке

данных и моделей машинного/глубокого обучения.

Эмпирические результаты на тестовом наборе данных

Подход одномерных данных. С помощью такого подхода, основываясь на данных скорректированной цены закрытия за 10 предыдущих дней (в качестве примера используется временной лаг 10), исследуем модель, которая лучше всего предсказывает скорректированную цену закрытия на 11-й день. На рисунке 4 показаны оценки рассматриваемых моделей. Для большей наглядности все значения R^2 , которые меньше -200 , обозначаются как -200 , кроме того, другие положительные значения R^2 также умножаются на 100.

Из рисунка 4 можно определить, что модели глубокого обучения, то есть FFNN, CNN, LSTM, CNN~LSTM, обеспечивают довольно хорошее качество прогноза (низкие *RMSE*, *MAE* и высокие R^2), как показано на рисунке 5а.

Основное назначение точечных диаграмм (рис. 5а) заключается в наблюдении и отображении взаимосвязей двух числовых переменных (то есть истинной и скорректированной прогнозируемой ценой акции). Точки на диаграмме отражают не только отдельные значения данных, но и закономерности, когда данные взяты в целом. Чем ближе точки расположены к диагональной (пунктирной) линии, тем точнее прогноз цен. Среди них модель CNN достигает лучших показателей, то есть $RMSE = 28.47$, $MAE = 18.52$, $R^2 = 0.992$. Настроенные гиперпараметры: количество фильтров = 250, размер ядра = 3, процент исключения = 0.01. Структура наилучшей модели CNN представлена на рисунке 5б.

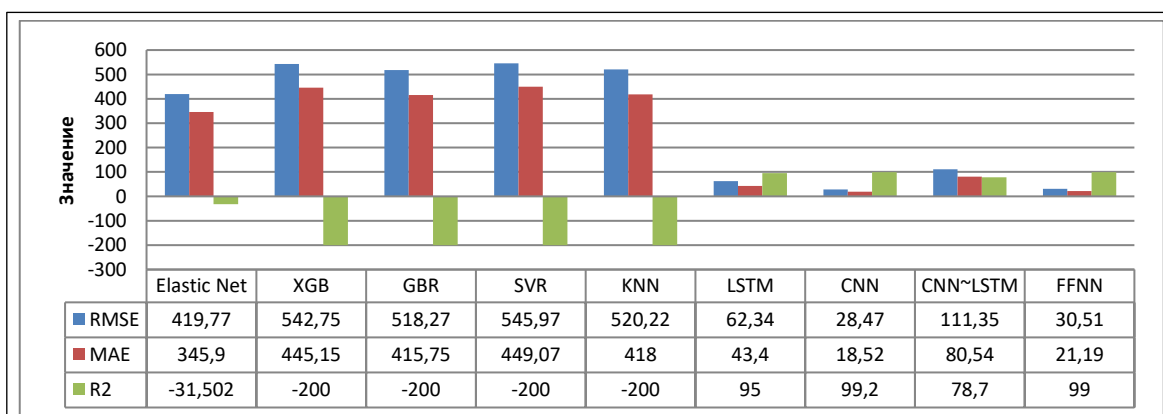
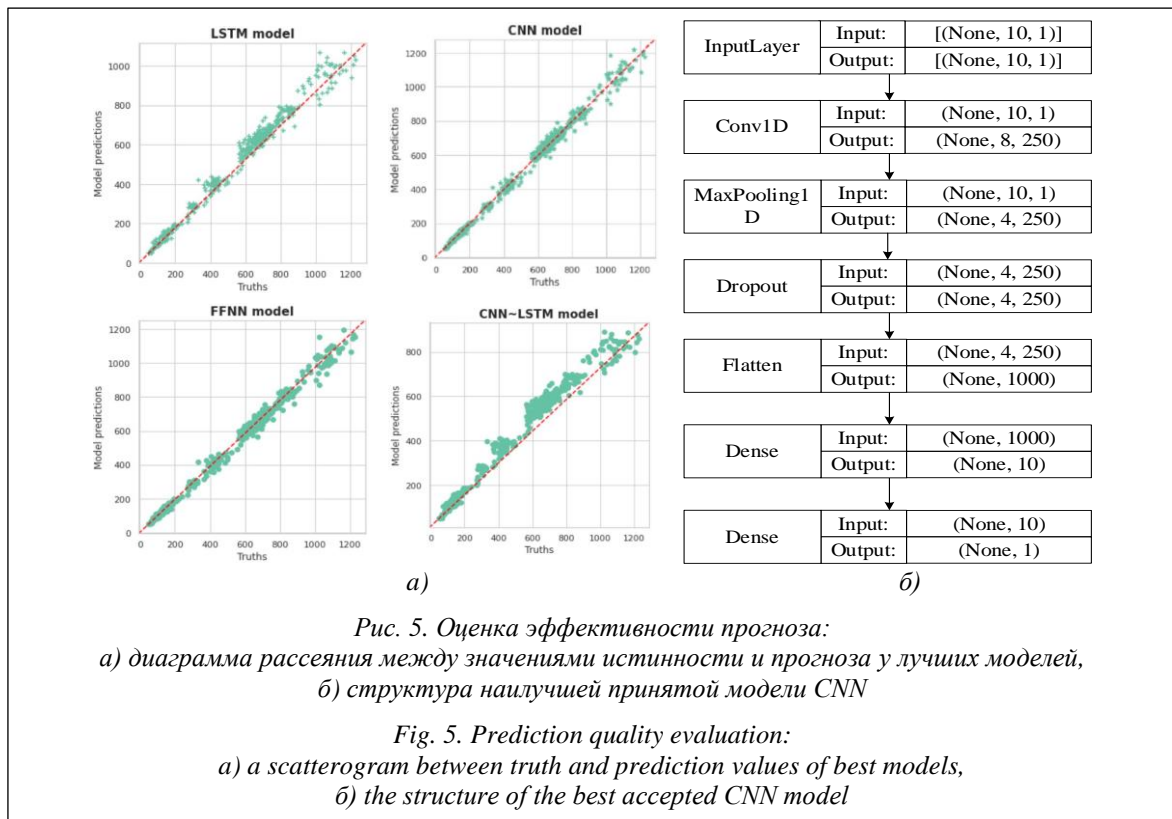


Рис. 4. Оценки рассматриваемых моделей (одномерные данные)

Fig. 4. Estimates of the considered models (univariate data)



Подход многомерных данных. При использовании этого подхода устанавливаем временное окно равным 250 (среднее количество рабочих дней в году), поэтому входными данными являются финансовые данные в течение одного дня (открытые/скорректированные цены закрытия/максимальные/минимальные цены, объем торгов). Необходимо выбрать модель, которая лучше всего предсказывает скорректированную цену закрытия в следую-

щий, 250-й день. В таблице 2 показаны оценки рассматриваемых моделей и настройки их гиперпараметров.

Из таблицы 2 можно определить лучшую модель как LSTM (ее структура представлена на рисунке б), которая обеспечивает наиболее оптимальные метрики оценки прогноза по отношению к другим моделям, то есть $RMSE = 365.8$, $MAE = 304.4$, $R^2 = -0.1$. В этом случае результат теста не так хорош, как при

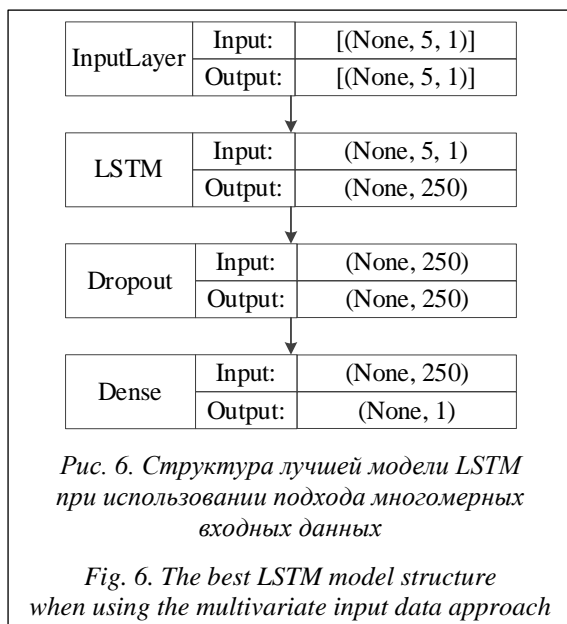
Таблица 2

Оценки рассматриваемых моделей (многомерные данные)

Table 2

Estimates of the considered models (multivariate data)

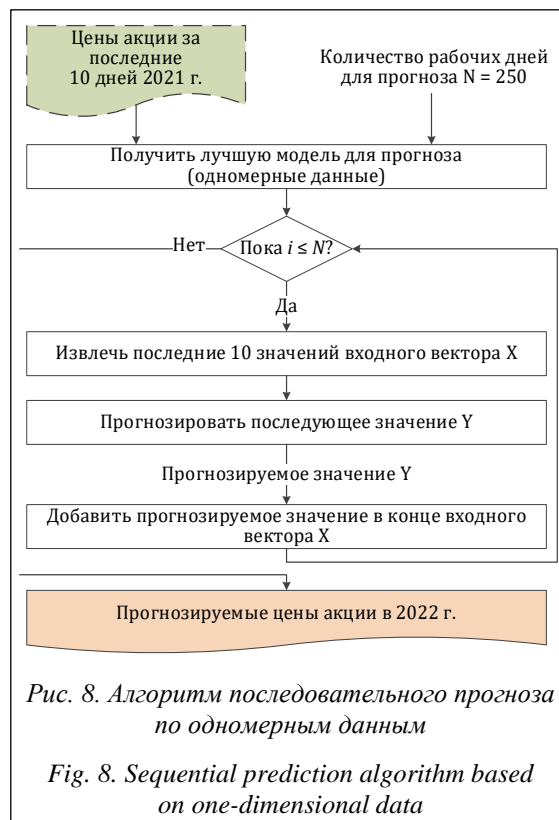
№ п/п	Модель	RMSE	MAE	R ²	Лучшие гиперпараметры
1	LSTM	365.8	304.4	-0.107	units: 250, dropout: 0.1
2	CNN	443.3	388.4	-11.615	filters: 700, kernels: 3, dropout: 0.01
3	FFNN	443.6	387.9	-12.418	num_layers: 3, units_1: 400, dropout: 0.01, units_2: 400, units_3: 400
4	Elastic Net	485.5	423.6	-35.18	alpha=0.01, l1_ratio=0.86
5	SVR	521.2	451.4	-129.745	C=10.00, gamma=auto, degree=1, kernel=poly, epsilon=0.3
6	KNN	560.5	479	-8046.701	n_neighbors = 38
7	GBR	559.3	477.9	-9580.948	max_depth=1, learning_rate=0.33, min_samples_split=100, min_samples_leaf=1, n_estimators=100
8	XGB	568.7	488.2	-16495.61	max_depth=50, n_estimators=1, min_child_weight=1, learning_rate=0.84, gamma=3.82
9	CNN~LSTM	560	477.6	-25373.58	filters: 350, kernels: 2, units_1: 60, units_2: 60, dropout: 0.1



подходе одномерных данных, поскольку модель должна предсказать на 250 дней вперед с финансовыми данными на один день, в то время как с подходом одномерных данных модель должна прогнозировать только цену акций на следующий день, основываясь на значениях 10 предыдущих дней. Кроме того, если уменьшить временное окно, то можно получить лучшие меры оценки, например, когда временное окно горизонта равно 50, R^2 увеличится до 0.752, а $RMSE = 187.87$, $MAE = 140.36$. Однако, как уже упоминалось, конечной целью данного исследования является прогнозирование восходящего/нисходящего тренда цен на акции, а не точных значений стоимости на определенную дату. Таким образом, при использовании модели LSTM восходящий тренд предсказывается довольно точно, как показано на рисунке 7.

Прогноз цен акций Tesla на 2022 год. Выбраны лучшие модели при использовании под-

ходов одномерных и многомерных данных. Соответственно, для подхода одномерных данных модель CNN получает последние 10 скорректированных цен закрытия в 2021 г. и прогнозирует цену на следующий день. Это полученное значение добавляется в конец входного вектора, который формирует входные данные для следующего прогноза. По этой логике данная итерация повторяется 250 раз. Для большего понимания алгоритм прогнозирования по одномерным данным показан на рисунке 8.



В случае использования подхода многомерных данных лучшая модель LSTM получает последние 250 значений цен акции в наборе

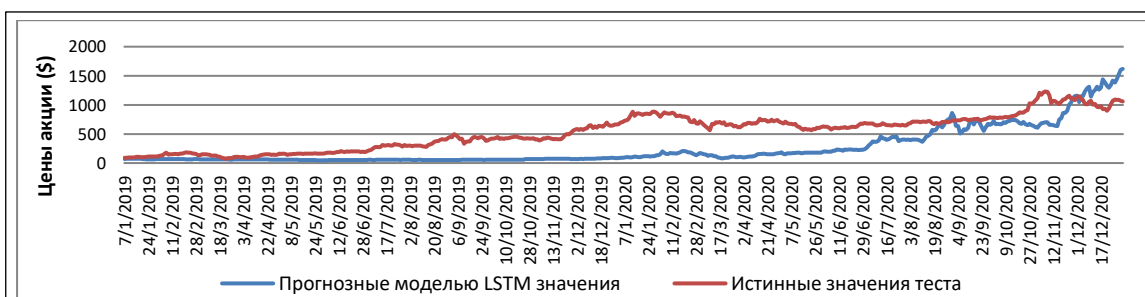


Рис. 7. Сравнение истинных и прогнозируемых значений по лучшей модели LSTM для тестового набора

Fig. 7. Comparing true and predicted values according to the best LSTM model for a test set

данных, каждая выборка в этом наборе обеспечивает прогнозируемое значение на следующей, 250-й день. Таким образом, при входных данных о ценах акции за 250 дней с 06/01/2021 по 31/12/2021 выходные прогнозные значения будут находиться в диапазоне данных с 03/01/2022 по 30/12/2022. Прогнозные значения, полученные из двух лучших моделей, приведены на рисунке 9.

Более того, для проверки трендов, созданных двумя лучшими моделями, авторы пытаются вычислить другие значения прогноза: комбинацию этих двух линий тренда, где каждое новое значение прогноза равно линейной комбинации значений от прогноза подходами одномерных и многомерных данных, с линейными коэффициентами -0.278 и 1.526 . Эти числа извлечены после приспособления модели линейной регрессии к истинным ценам акции за первые 20 дней 2022 г. Формула прогнозируемых будущих значений (y_{future}), основанная на прогнозируемых значениях с помощью многомерного ($y_{multivariate}$) и одномерного ($y_{univariate}$) подходов, описана следующим образом: $y_{future} = -0.278 y_{multivariate} + 1.526 y_{univariate}$.

Расчетные значения прогноза y_{future} также представлены на рисунке 9в. Наблюдая за графиками, можно сделать вывод, что в целом цены акции Tesla сохраняют восходящий тренд – высокие цены относительно 2021 г. – в 2022 г. Кроме того, для проверки точности метода прогноза авторами также приведена статистика о доходах компании Tesla. После сбора данных отчета о доходах за период 2009–2021 гг. (<https://www.macrotrends.net/stocks/charts/TSLA/tesla/income-statement>) имеем значения Доход, Валовая прибыль и Операционный доход (рис. 10). Доход может быть определен как сумма денег, которую компания получает от своих клиентов в обмен на товары или услуги. Валовая прибыль может быть определена как прибыль, которую компания получает после вычета переменных затрат, непосредственно связанных с производством и продажей ее продукции или предоставлением услуг. Операционный доход может быть определен как доход после вычета операционных расходов и до вычета процентных платежей и налогов. Следует отметить, что доход Tesla за квартал, закончившийся 31 марта 2022 г., составил 18.756 млрд

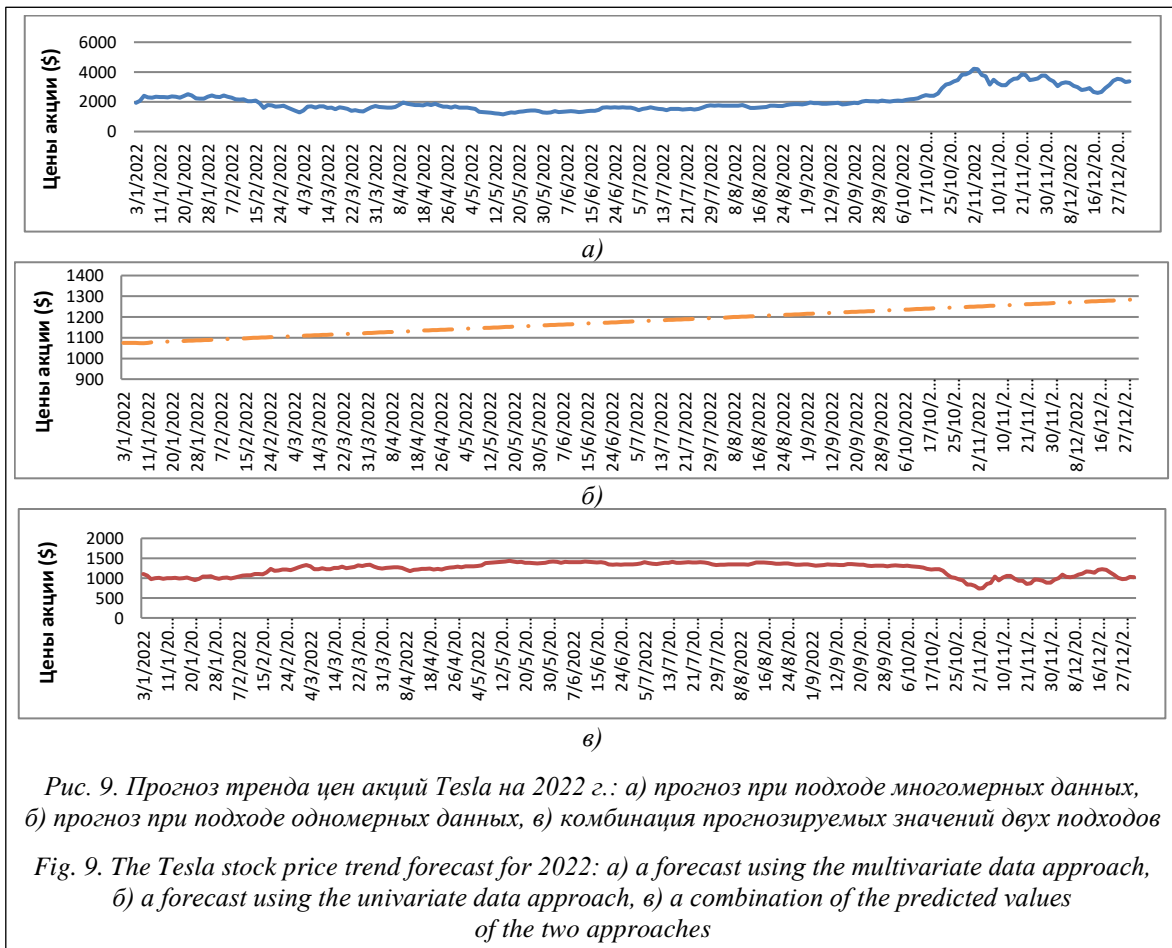


Рис. 9. Прогноз тренда цен акций Tesla на 2022 г.: а) прогноз при подходе многомерных данных, б) прогноз при подходе одномерных данных, в) комбинация прогнозируемых значений двух подходов

Fig. 9. The Tesla stock price trend forecast for 2022: а) a forecast using the multivariate data approach, б) a forecast using the univariate data approach, в) a combination of the predicted values of the two approaches

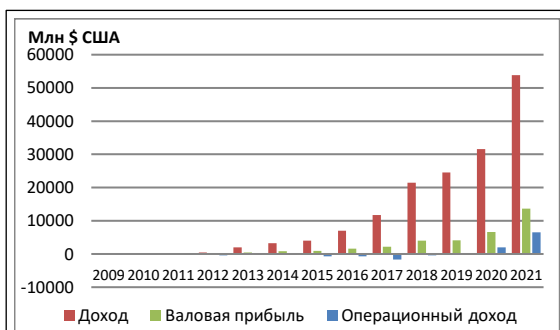


Рис. 10. Отчет о разных доходах компании Tesla за период 2009–2021 гг.

Fig. 10. Miscellaneous income report of Tesla in period 2009–2021

долларов, что на 80.54 % больше, чем в предыдущем году. Доход Tesla за 12 месяцев, закончившихся 31 марта 2022 г., составил 62.190 млрд долларов, что на 73.04 % больше, чем в предыдущем году. Годовой доход Tesla за 2021 г. составил 53.823 миллиарда долларов, что на 70.67 % больше, чем в 2020 г.

Патентный анализ Tesla для определения технологических тенденций в 2022 г.

Авторами уже были приведены аргументы того, что если цены акции предприятия останутся в восходящем тренде, то технологии, разработанные предприятием, наиболее вероятно, станут многообещающими инновациями в будущем. После наблюдения за прогнозными значениями, полученными из предложенных моделей, с высокой уверенностью можно предположить, что в целом цена акции Tesla сохранит восходящий тренд в 2022 г. По этой причине рассмотрим патентные заявки Tesla, которые были поданы с 2019 по 2021 гг., чтобы предсказать тенденции развития технологий в 2022 г. Метаданные патентных заявок Tesla загружаются с известной платформы The Lens (<https://about.lens.org>).

Запрос фильтра для получения патентных заявок Tesla: `earliestPriorityDate = (2019-01-01–2021-12-31)`, `Grouped by Simple Families`, `Document Type = (excl Design_right)`, `Owner Name Exact = (Tesla Inc, Tesla Motors Inc)`. Это означает, что самая ранняя дата приоритета (самая ранняя дата подачи в семействе патентных заявок) – с 01/01/2019 по 31/12/2021; патенты на право дизайна также были исключены. Запрос возвращает 20 патентных заявок; соответственно, можно синтезировать наиболее часто встречающиеся коды совместной па-

тентной классификации (Cooperative Patent Classification – CPC) патентов Tesla, которые представлены на рисунке 11.

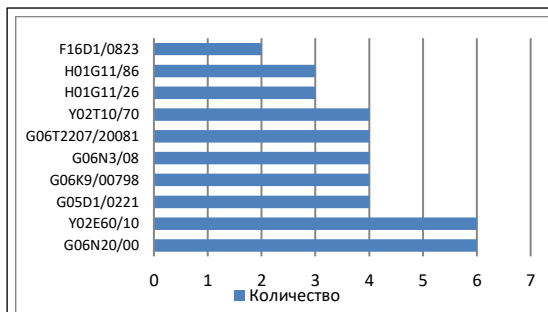


Рис. 11. Топ-10 кодов классификации CPC развивающихся технологий Tesla

Fig. 11. Top 10 CPC Classification Codes for technologies developed by Tesla

Среди них топ-5 классификационных кодов CPC.

G06N20/00: Машинное обучение.

Y02E60/10: Хранение энергии с помощью аккумуляторов.

G05D1/0221: Контроль положения, курса или высоты наземных, водных, воздушных или космических транспортных средств, например, автопилот (радионавигационные системы или аналогичные системы, использующие другие волны), включающий процесс обучения.

G06K9/00: Методы или устройства для распознавания шаблонов для чтения графика или преобразования шаблона механических параметров, например, силы или присутствия, в электрические сигналы; распознавание или понимание изображения или видео; распознавание речи.

G06N3/08: Вычислительные механизмы, основанные на биологических моделях (на основе физических объектов, управляемых смоделированным интеллектом, для воспроизведения разумных форм жизни, например, роботов, воспроизводящих домашних животных или людей по их внешнему виду или поведению).

С другой стороны, чтобы лучше понять технологии, разработанные Tesla, авторы сначала создали корпус текстов из разделов Заголовков, Аннотация и Основание 20 патентных заявок, а затем подали этот корпус в программу VOSviewer [26] для извлечения и кластеризации релевантных ключевых слов (параметры настройки: метод подсчета = Full counting, порог появления Occurrence threshold = 3).

На основании описания кодов классификации CPC и полученных с помощью VOSviewer

кластеров ключевых слов получено семь практических технологических отраслей, в которых Tesla ведет разработки. Рассматриваем эти отрасли как технологические тренды в 2022 г. Для проверки выдвинутой гипотезы и полученных семи трендов в разработках Tesla были изучены публикации на ведущих научно-технологических порталах. Эксперты предметных областей подтвердили эти тренды опубликованными статьями, которые относятся к соответствующим группам ключевых слов.

1. Автономное вождение: autonomous driving, autonomous vehicle applications; machine learning, deep learning system, time series elements, image data (<https://www.gtlaw.com/en/insights/2022/1/published-articles/5-trends-to-watch-in-2022-autonomous-vehicles>).

2. Литий-ионные аккумуляторы (LIB): elastic polymer binder, dry electrode film, electrode film fabrication, extracting lithium from clay minerals, lithium ion battery, energy storage system (device). LIB стали одним из основных решений для хранения энергии в современном обществе. Области применения и доля рынка LIB быстро увеличивались и продолжают демонстрировать устойчивую тенденцию к росту ([27] и <https://www.startus-insights.com/innovators-guide/top-10-renewable-energy-trends-2022>).

3. Электродвигатель, электромобиль: stator, electric motor, electric vehicle (<https://www.electronicdesign.com/resources/industry-insights/article/21213883/intelligent-product-solutions-top-5-technology-trends-for-2022>).

4. Переработанные материалы для автомобиля: environmentally-friendly (recycled) materials for vehicle (<https://abcnews.go.com/Business/car-made-recycled-plastic-future/story?id=81634993>, <https://www.businessinsider.com/sc/10-technology-trends-for-2022-beyond-2022-1>).

5. Пользовательские интерфейсы для управления автомобилем: user interfaces for vehicle operation (<https://grapeup.com/blog/what-trends-will-set-the-course-of-change-in-the-automotive-industry-for-2022-year>).

6. Упаковка ИС: packaged integrated circuits, IC packaging technologies (<https://www.startus-insights.com/innovators-guide/electronics-manufacturing-trends>).

7. Фотоэлектрическая структура: solar cells, photovoltaic structure, photovoltaic roof tile.

Шесть из семи выявленных технологических кластеров были подтверждены публикациями в экспертных сообществах, поэтому оценка точности прогноза может быть рассчитана как 6/7 (~85.7 %).

Аналогично были проведены эксперименты для анализа цены акций компании Samsung (код 005930.KS) за период 2000–2021 гг. После наблюдения за прогнозными значениями, полученными из предложенных моделей, с высокой уверенностью можно предположить, что в целом цены акций Samsung тоже сохраняют восходящий тренд в 2022 г. По этой причине рассмотрены патентные заявки Samsung, которые были поданы с 2020 по 2021 гг., чтобы предсказать тенденции развития технологий в 2022 г. Запрос фильтра для получения патентных заявок Samsung из платформы The Lens: *earliestPriorityDate = (2020-01-01–2021-12-31), Grouped by Simple Families, Document Type = (excl Design_right), Owner Name Exact = (Samsung Electronics Co Ltd)*. Далее после применения кластеризации получены 16 практических технологических отраслей, в которых Samsung ведет разработки. Рассматриваем эти отрасли как технологические тренды в 2022–2023 гг. Аналогично для проверки выдвинутой гипотезы и полученных 16 трендов в разработках Samsung были изучены публикации на ведущих научно-технологических порталах. Эксперты предметных областей подтвердили эти тренды опубликованными статьями, которые относятся к соответствующим группам ключевых слов. Этими трендами являются: Полупроводниковая промышленность, интеграция хранения и вычислений, упаковка ИС (IC packaging); Беспроводная технология, сеть 5G; Автоматизация электронного проектирования; Автономное вождение; Технология Bio-Sensing-PPG (фотоплетизмография); Metaverse – отслеживание взгляда и 3D-сенсоры; Датчик расстояния; Технология нейроморфных вычислений; Литий-ионные батареи; Биометрическая промышленность; Металловоздушные аккумуляторы; Обработка естественного языка; Программное обеспечение для анализа временных рядов; Видеотрансляция; Технология водоочистки и очистки сточной воды; Плазменная технология. Первые 15 из 16 технологических кластеров были одобрены, следовательно, оценка точности прогноза (precision) может быть рассчитана как 15/16 (~93.8 %).

Заключение

Технологическое развитие оказывает существенное влияние на принятие стратегических решений. Заблаговременное осознание возможных предстоящих или возникающих технологических тенденций может привести к

укреплению конкурентоспособности и рыночных позиций предприятий. Однако, если компании, ориентированные на инновации, игнорируют появляющиеся технологические разработки, они могут не использовать весь потенциал своих собственных продуктов или технологий.

В работе была выдвинута новая гипотеза: если стоимость акции предприятия сохраняет восходящий тренд, то технологии, разработанные предприятием, скорее всего, станут перспективными инновациями в будущем. При этом предложен и программно реализован новый метод прогнозирования технологических трендов путем объединения следующих шагов: анализ веб-новостей для извлечения значимых предприятий, прогнозирование тенденций стоимости акций и анализ кластеризации патентных заявок. В отличие от других исследований предложенный метод продвигает идею прогнозирования технологических трендов путем прогнозирования тренда цен акций с использованием байесовской оптимизации для поиска лучших гиперпараметров моделей машинного и глубокого обучения. Кроме того, авторы объединили использование подходов к подготовке одномерных и многомерных данных для прогнозирования тренда цен акций. В частности, предлагаемый метод был рассмотрен для прогнозирования динамики цен акций компаний Tesla и

Samsung в качестве примера. Для этого были исследованы различные популярные модели машинного и глубокого обучения. После наблюдения за прогнозными значениями, полученными из изложенных моделей, с высокой уверенностью можно предположить, что в целом цены акций Tesla и Samsung сохраняют восходящий тренд в 2022 г. После сбора патентных заявок Tesla и Samsung метод кластеризации текстового корпуса был использован для изучения технологических отраслей, в которых Tesla и Samsung ведут разработки. Эти технологические тенденции недавно были подтверждены экспертами в соответствующих опубликованных статьях. В итоге точность полученного прогноза составляет около 85.7 % (для патентов Tesla) и 93.8 % (для патентов Samsung). Таким образом, это свидетельствует о положительной надежности предложенного метода.

Дальнейшая работа будет направлена на повышение точности метрики оценки R^2 с целью улучшения моделей для прогноза стоимости акций при использовании подхода многомерных данных и учета конъюнктуры рынка таких данных, как новости, мировая экономика, общественные настроения и другая конфиденциальная финансовая информация. Кроме того, также планируется исследовать и разработать методы извлечения значимых ключевых слов из текстового корпуса.

Исследование выполнено при поддержке РФФИ в рамках научного проекта № 20-37-90092.

Литература

1. Кравец А.Г., Сальникова Н.А. Предсказательное моделирование трендов технологического развития // Изв. СПбГТИ (ТУ). 2020. № 55. С. 103–108. DOI: 10.36807/1998-9849-2020-55-81-103-108.
2. Xu M. The Impact of blockchain technology on stock price: An empirical study. SHS Web of Conf., 2021, vol. 96, art. 04008. DOI: 10.1051/SHSCONF/20219604008.
3. Lam H.K.S., Ding L., Cheng T. et al. The impact of 3D printing implementation on stock returns. Int. J. of Operations and Production Management, 2019, vol. 39, no. 6/7/8, pp. 935–961. DOI: 10.1108/ijopm-01-2019-0075.
4. Ballestra L.V., Fontana S., Scuotto V., Solimene S. A multidisciplinary approach for assessing open innovation model impact on stock return dynamics. Management Decision, 2018, vol. 56, no. 6, pp. 1430–1444. DOI: 10.1108/MD-05-2017-0481.
5. Vitt C.A., Xiong H. The impact of patent activities on stock dynamics in the high-tech sector. Proc. IEEE Int. Conf. on Data Mining, 2015, pp. 399–408. DOI: 10.1109/ICDM.2015.95.
6. Senarathne C.W., Wei J. The impact of patent citation information flow regarding economic innovation on common stock returns: Volume vs. patent citations. Int. J. of Innovation Studies, 2018, vol. 2, no. 4, pp. 137–152. DOI: 10.1016/J.IJIS.2019.02.001.
7. Grieco D. Innovation and stock market performance: A model with ambiguity-averse agents. J. of Evolutionary Economics, 2018, vol. 28, no. 2, pp. 287–303. DOI: 10.1007/S00191-017-0537-1.
8. Lee B. The relationship between technology life cycle and Korean stock market performance. Int. J. of Financial Studies, 2018, vol. 6, no. 4, p. 88. DOI: 10.3390/IJFS6040088.
9. Dosso M., Vezzani A. Firm market valuation and intellectual property assets. Industry and Innovation, 2020, vol. 27, no. 7, pp. 705–729. DOI: 10.1080/13662716.2019.1685374.

10. Viet N.T., Gneushev V. Analyzing and forecasting emerging technology trends by mining web news. *CCIS*, 2021, pp. 55–69. DOI: 10.1007/978-3-030-87034-8_5.
11. Viet N.T., Kravets A., Duong Q.H.T. Data mining methods for analysis and forecast of an emerging technology trend: A systematic mapping study from SCOPUS papers. *LNAI*, 2021, pp. 81–101. DOI: 10.1007/978-3-030-86855-0_7.
12. Тормозов В.С., Золкин А.Л., Менциев А.У. Интеллектуальный подход к задаче информирования и краткосрочного прогнозирования временного ряда в онлайн-режиме // Программные продукты и системы. 2021. Т. 34. № 3. С. 399–408. DOI: 10.15827/0236-235X.135.399-408.
13. Ji X., Wang J., Yan Z. A stock price prediction method based on deep learning technology. *Int. J. of Crowd Science*, 2021, vol. 5, no. 1, pp. 55–72. DOI: 10.1108/IJCS-05-2020-0012.
14. Thu H.G.T., Thanh T.N., Quy T. Le. A Neighborhood deep neural network model using sliding window for stock price prediction. *Proc. IEEE Int. Conf. on BigComp*, 2021, pp. 69–74. DOI: 10.1109/BigComp 51126.2021.00022.
15. Nguyen T.-T., Yoon S. A novel approach to short-term stock price movement prediction using transfer learning. *Applied Sciences*, 2019, vol. 9, no. 22, art. 4745. DOI: 10.3390/app9224745.
16. Vo N.N.Y., He X., Liu S., Xu G. Deep learning for decision making and the optimization of socially responsible investments and portfolio. *Decision Support Systems*, 2019, vol. 124, art. 113097. DOI: 10.1016/J.DSS.2019.113097.
17. Jiang W. Applications of deep learning in stock market prediction: Recent progress. *Expert Systems with Applications*, 2021, vol. 184, art. 115537. DOI: 10.1016/j.eswa.2021.115537.
18. Li A.W., Bastos G.S. Stock market forecasting using deep learning and technical analysis: a systematic review. *IEEE Access*, 2020, vol. 8, pp. 185232–185242. DOI: 10.1109/ACCESS.2020.3030226.
19. Yang C., Zhai J., Tao G. Deep learning for price movement prediction using convolutional neural network and long short-term memory. *Mathematical Problems in Engineering*, 2020, pp. 1–13. DOI: 10.1155/2020/2746845.
20. Zhang J., Li L., Chen W. Predicting stock price using two-stage machine learning techniques. *Computational Economics*, 2021, vol. 57, no. 4, pp. 1237–1261. DOI: 10.1007/s10614-020-10013-5.
21. Weng B., Lu L., Wang X. et al. Predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications*, 2018, vol. 112, pp. 258–273. DOI: 10.1016/j.eswa.2018.06.016.
22. Göçken M., Özçalıcı M., Boru A., Dosdoğru A.T. Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection. *Neural Computing and Applications*, 2019, vol. 31, no. 2, pp. 577–592. DOI: 10.1007/s00521-017-3089-2.
23. Zhou F., Zhang Q., Sornette D., Jiang L. Cascading logistic regression onto gradient boosted decision trees for forecasting and trading stock indices. *Applied Soft Computing*, 2019, vol. 84, art. 105747. DOI: 10.1016/j.asoc.2019.105747.
24. Lu W., Li J., Li Y. et al. A CNN-LSTM-Based model to forecast stock prices. *Complexity*, 2020, vol. 2020, pp. 1–10. DOI: 10.1155/2020/6622927.
25. Wu J., Chen X., Xiong L.-D. et al. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. of Electronic Science and Technology*, 2019, vol. 17, no. 1, pp. 26–40.
26. van Eck N.J., Waltman L. Visualizing bibliometric networks. *Measuring Scholarly Impact*, 2014, pp. 285–320. DOI: 10.1007/978-3-319-10377-8_13.
27. Liu Y., Zhang R., Wang J., Wang Y. Current and future lithium-ion battery manufacturing. *iScience*, 2021, vol. 24, no. 4, art. 102332. DOI: 10.1016/j.isci.2021.102332.

Forecasting technological trends based on the heterogeneous data analysis

Nguyen Thanh Viet^{1,2}, Postgraduate Student, vietqn1987@gmail.com
A.G. Kravets^{1,3}, Dr.Sc. (Engineering), Professor, agk@gde.ru

¹ Volgograd State Technical University, Volgograd, 400005, Russian Federation

² Pham Van Dong University, Quang Ngai, Viet Nam

³ Dubna State University, Dubna, 141980, Russian Federation

Abstract. To achieve competitiveness, enterprises need to exploit a development strategy by forecasting promising technologies using limited resources. Numerous previous studies have shown that unexpected changes in R&D and patenting of intellectual property are associated with large changes in the market value of an enterprise. In fact, there is a strong correlation between the volatility of market shares, stock prices in the early stages of high-tech enterprise development, and the period when the innovative technology is not yet defined. Thus, we propose that if company's share prices continue to trend upward, then the developed technologies are likely to become promising innovations in the future.

This paper proposes a method for forecasting technology trends by analyzing web news to identify high-tech enterprises, predicting stock price trends for selected enterprises and analyzing the clusters of patent applications. Unlike other studies, our method advances the idea of predicting technology trends by forecasting the stock price trend using univariate and multivariate data preparation approaches, and using Bayesian optimization to explore the best hyperparameters for machine and deep learning models.

The study uses the developed software system for analyzing word frequency burst detection and stock price prediction. In particular, the proposed method is adopted to predict the price dynamics of Tesla and Samsung shares as case studies.

Keywords: data mining, technology forecast, stock price, Tesla, Bayesian optimization, machine learning, deep learning, patent analysis, cluster analysis, VOSviewer.

Acknowledgements. *The reported study was funded by RFBR according to the research project no. 20-37-90092.*

References

1. Kravets A.G., Salnikova N.A. Predictive modeling of technological development trends. *Bull. of St PbSIT (TU)*, 2020, vol. 55, no. 81, pp. 103–108 (in Russ.)
2. Xu M. The Impact of blockchain technology on stock price: An empirical study. *SHS Web of Conf.*, 2021, vol. 96, art. 04008. DOI: 10.1051/SHSCONF/20219604008.
3. Lam H.K.S., Ding L., Cheng T. et al. The impact of 3D printing implementation on stock returns. *Int. J. of Operations and Production Management*, 2019, vol. 39, no. 6/7/8, pp. 935–961. DOI: 10.1108/ijopm-01-2019-0075.
4. Ballestra L.V., Fontana S., Scuotto V., Solimene S. A multidisciplinary approach for assessing open innovation model impact on stock return dynamics. *Management Decision*, 2018, vol. 56, no. 6, pp. 1430–1444. DOI: 10.1108/MD-05-2017-0481.
5. Vitt C.A., Xiong H. The impact of patent activities on stock dynamics in the high-tech sector. *Proc. IEEE Int. Conf. on Data Mining*, 2015, pp. 399–408. DOI: 10.1109/ICDM.2015.95.
6. Senarathne C.W., Wei J. The impact of patent citation information flow regarding economic innovation on common stock returns: Volume vs. patent citations. *Int. J. of Innovation Studies*, 2018, vol. 2, no. 4, pp. 137–152. DOI: 10.1016/J.IJIS.2019.02.001.
7. Grieco D. Innovation and stock market performance: A model with ambiguity-averse agents. *J. of Evolutionary Economics*, 2018, vol. 28, no. 2, pp. 287–303. DOI: 10.1007/S00191-017-0537-1.
8. Lee B. The relationship between technology life cycle and Korean stock market performance. *Int. J. of Financial Studies*, 2018, vol. 6, no. 4, p. 88. DOI: 10.3390/IJFS6040088.
9. Dosso M., Vezzani A. Firm market valuation and intellectual property assets. *Industry and Innovation*, 2020, vol. 27, no. 7, pp. 705–729. DOI: 10.1080/13662716.2019.1685374.
10. Viet N.T., Gneushev V. Analyzing and forecasting emerging technology trends by mining web news. *CCIS*, 2021, pp. 55–69. DOI: 10.1007/978-3-030-87034-8_5.
11. Viet N.T., Kravets A., Duong Q.H.T. Data mining methods for analysis and forecast of an emerging technology trend: A systematic mapping study from SCOPUS papers. *LNAI*, 2021, pp. 81–101. DOI: 10.1007/978-3-030-86855-0_7.
12. Tormozov V.S., Zolkin A.L., Mentsiev A.U. An intelligent approach to online reporting and short-term forecasting of time series. *Software & Systems*, 2021, vol. 13, pp. 399–408. DOI: 10.15827/0236-235X.135.399-408 (in Russ.).
13. Ji X., Wang J., Yan Z. A stock price prediction method based on deep learning technology. *Int. J. of Crowd Science*, 2021, vol. 5, no. 1, pp. 55–72. DOI: 10.1108/IJCS-05-2020-0012.
14. Thu H.G.T., Thanh T.N., Quy T. Le. A Neighborhood deep neural network model using sliding window for stock price prediction. *Proc. IEEE Int. Conf. on BigComp*, 2021, pp. 69–74. DOI: 10.1109/BigComp 51126.2021.00022.

15. Nguyen T.-T., Yoon S. A novel approach to short-term stock price movement prediction using transfer learning. *Applied Sciences*, 2019, vol. 9, no. 22, art. 4745. DOI: 10.3390/app9224745.
16. Vo N.N.Y., He X., Liu S., Xu G. Deep learning for decision making and the optimization of socially responsible investments and portfolio. *Decision Support Systems*, 2019, vol. 124, art. 113097. DOI: 10.1016/J.DSS.2019.113097.
17. Jiang W. Applications of deep learning in stock market prediction: Recent progress. *Expert Systems with Applications*, 2021, vol. 184, art. 115537. DOI: 10.1016/j.eswa.2021.115537.
18. Li A.W., Bastos G.S. Stock market forecasting using deep learning and technical analysis: a systematic review. *IEEE Access*, 2020, vol. 8, pp. 185232–185242. DOI: 10.1109/ACCESS.2020.3030226.
19. Yang C., Zhai J., Tao G. Deep learning for price movement prediction using convolutional neural network and long short-term memory. *Mathematical Problems in Engineering*, 2020, pp. 1–13. DOI: 10.1155/2020/2746845.
20. Zhang J., Li L., Chen W. Predicting stock price using two-stage machine learning techniques. *Computational Economics*, 2021, vol. 57, no. 4, pp. 1237–1261. DOI: 10.1007/s10614-020-10013-5.
21. Weng B., Lu L., Wang X. et al. Predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications*, 2018, vol. 112, pp. 258–273. DOI: 10.1016/j.eswa.2018.06.016.
22. Göçken M., Özçalıcı M., Boru A., Dosdoğru A.T. Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection. *Neural Computing and Applications*, 2019, vol. 31, no. 2, pp. 577–592. DOI: 10.1007/s00521-017-3089-2.
23. Zhou F., Zhang Q., Sornette D., Jiang L. Cascading logistic regression onto gradient boosted decision trees for forecasting and trading stock indices. *Applied Soft Computing*, 2019, vol. 84, art. 105747. DOI: 10.1016/j.asoc.2019.105747.
24. Lu W., Li J., Li Y. et al. A CNN-LSTM-Based model to forecast stock prices. *Complexity*, 2020, vol. 2020, pp. 1–10. DOI: 10.1155/2020/6622927.
25. Wu J., Chen X., Xiong L.-D. et al. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. of Electronic Science and Technology*, 2019, vol. 17, no. 1, pp. 26–40.
26. van Eck N.J., Waltman L. Visualizing bibliometric networks. *Measuring Scholarly Impact*, 2014, pp. 285–320. DOI: 10.1007/978-3-319-10377-8_13.
27. Liu Y., Zhang R., Wang J., Wang Y. Current and future lithium-ion battery manufacturing. *iScience*, 2021, vol. 24, no. 4, art. 102332. DOI: 10.1016/j.isci.2021.102332.

Для цитирования

Нгуен Тхань Вьет, Кравец А.Г. Прогнозирование технологических тенденций на основе анализа разнородных данных // Программные продукты и системы. 2022. Т. 35. № 3. С. 396–412. DOI: 10.15827/0236-235X.139.396-412.

For citation

Nguyen Thanh Viet, Kravets A.G. Forecasting technological trends based on the heterogeneous data analysis. *Software & Systems*, 2022, vol. 35, no. 3, pp. 396–412 (in Russ.). DOI: 10.15827/0236-235X.139.396-412.