УДК 338.27

АНАЛИТИЧЕСКИЕ ВОЗМОЖНОСТИ МАШИННОГО ОБУЧЕНИЯ В ЭКОНОМИЧЕСКОМ ПРОГНОЗИРОВАНИИ

Мирончук В.А.,
Кубанский государственный аграрный университет имени И.Т. Трубилина, Краснодар,
email: mironchuk.v@edu.kubsau.ru

Золкин А.Л.,
Поволжский государственный университет телекоммуникаций и информатики (ПГУТИ), Самара,
email: alzolkin@list.ru

Артамонова К.А.,
Финансовый университет при Правительстве Российской Федерации, Москва,
email: kartamonova@fa.ru

Подолько П.М.,
Российский университет дружбы народов имени Патриса Лумумбы, Москва,
email: podolich@yandex.ru

Аннотация. *Научное сообщество сохраняет повышенный интерес к исследованию взаимодействия современных технологий машинного обучения и прогнозирования экономических показателей. В данной статье осуществлен обзор принципов работы и потенциала применения алгоритмов машинного обучения, включая Random forest, линейную регрессию, Decision Stump и метод опорных векторов, а также техники бустинга для оптимизации процессов анализа данных, снижения затрат и уменьшения вероятности ошибок в экономических расчётах. В статье акцентировано внимание на эффективности нейронных сетей в задачах прогнозирования динамики стоимости акций на Московской фондовой бирже. Анализ сравнительной динамики и реальных результатов ценных бумаг показывает, что, несмотря на сложность точного прогнозирования из-за непредвиденных рыночных факторов, нейронные сети демонстрируют высокую вероятность корректного предсказания трендов, что свидетельствует о значительных перспективах использования данных технологий в финансовом анализе.*

Ключевые слова: машинное обучение, экономические показатели, алгоритмы прогнозирования, Random forest, линейная регрессия, метод опорных векторов, бустинг.

ANALYTICAL CAPABILITIES OF MACHINE LEARNING IN ECONOMIC FORECASTING

Mironchuk V.A.,
Kuban State Agrarian University named after I.T. Trubilin, Krasnodar,
email: mironchuk.v@edu.kubsau.ru

Zolkin A.L.,
Povolzhskiy State University of Telecommunications and Informatics (PGUTY), Samara,
email: alzolkin@list.ru

Artamonova K.A.,
Financial University under the Government of the Russian Federation, Moscow,
email: kartamonova@fa.ru

Podolko P.M.,
Peoples' Friendship University of Russia, Moscow, email: podolich@yandex.ru

Abstract. *The scientific community retains an increased interest in studying the interaction of modern machine learning technologies and forecasting economic indicators. This article provides an overview of the principles of operation and potential applications of machine learning algorithms, including Random forest, linear regression, Decision Stump and the support vector machine, as well as boosting techniques to optimize data analysis processes, reduce costs and reduce the likelihood of errors in economic calculations. The article focuses on the effectiveness of neural networks in predicting the dynamics of stock prices on the Moscow Stock Exchange. An analysis of the comparative dynamics and real results of securities shows that, despite the difficulty of accurate forecasting due to unforeseen market factors, neural networks demonstrate a high probability of correctly predicting trends, which indicates significant prospects for using these technologies in financial analysis.*

Keywords: machine learning, economic indicators, forecasting algorithms, Random forest, linear regression, support vector machine, boosting.

Научный интерес к вопросу применения современных технологий машинного обучения в прогнозировании экономических показателей и улучшения процессов принятия экономических решений остается высоким. Возможности машинного обучения, особенно алгоритмов таких как Random forest, линейная регрессия, Decision Stump и метод опорных векторов, а также техники бустинга, могут существенно ускорить обработку данных, снизить затраты и минимизировать риски ошибок в экономических расчетах. Данная статья представляет собой обзор основных методов технического анализа, применяемых для предсказания экономических индикаторов.

Основной целью исследования является анализ эффективности использования нейронных сетей для прогнозирования динамики стоимости акций голубых фишек на Московской фондовой бирже. В рамках работы проведен сравнительный анализ прогнозов, полученных с помощью нейросетей, и реальной динамики стоимости ценных бумаг на фондовом рынке.

Следует отметить, что абсолютно точный прогноз рыночной динамики невозможен из-за множества непредвиденных и случайных факторов, влияющих на рынок. Тем не менее, способность нейронной сети предсказывать тенденции изменения стоимости акций с высокой вероятностью уже является значимым достижением и может считаться успехом в области применения нейросетевых технологий в финансовом анализе.

Обзор литературы

В процессе подготовки данной статьи авторы опирались на обширный анализ научной литературы. Так, в работе К. А. Макеева, К. В. Шаповаленко и А. В. Гречневой проведено исследование алгоритмов машинного обучения с целью их интеграции в качестве модуля для системного прогнозирования [7]. Д. Ю. Саханевич в своем исследовании акцентирует внимание на классификации и систематизации аспектов машинного обучения, подчеркивая необходимость внедрения алгоритмов искусственного интеллекта для оптимизации управления процессами [8]. Е. Шуляк разрабатывает индексы, на основе которых осуществляется прогнозирование макроэкономических показателей с использованием различных методик машинного обу-

чения [13]. В статье Н. Д. Тимакова рассматривается фазовый анализ и разработка прогнозной модели на базе линейно-клеточного автомата [9]. Наконец, исследование Э. А. Бекировой и М. Э. Бекрирова посвящено прогнозированию на основе методов машинного обучения, где анализируются преимущества и недостатки применения данных технологий [3].

В исследовательской работе В. А. Филиппенко и И. В. Решетниковой представлен процесс обучения нейронной сети, а также проведен анализ предсказаний, сгенерированных данным инструментом [10]. В статье А. С. Р. Филиппова и О. Л. Ксенофонтова осуществляется систематический обзор методологий машинного обучения, при этом авторы детализируют механизмы функционирования каждого из методов, выделяя их преимущества и ограничения [11]. А. Н. Давыдов в своём исследовании фокусируется на анализе вклада машинного обучения в экономическую науку и прогнозирует дальнейшее воздействие этой технологии на экономические процессы [5]. П. А. Чагина оценивает применимость различных типов нейронных сетей в алгоритмах, анализирует их потенциал для прогнозирования и обсуждает перспективы машинного обучения [12]. Наконец, в работе А. С. Балыкова, Е. А. Калединой и С. В. Володина проводится анализ экспериментальных данных, полученных с применением современных техник машинного обучения с учителем, используемых для решения задач регрессии, прогнозирования и экстракции знаний [2].

В научном исследовании, авторами которого являются А. С. Колосова, А. С. Каменева, Г. В. Чуков и А. Ю. Никифоров, подвергаются анализу алгоритмы, методики машинного обучения и их потенциал для оценки качественных аспектов [4]. В работе представлены кейс-стади из практики, демонстрирующие успешное применение этих алгоритмов для повышения значений показателей. В работе И. А. Астраханцевой, А. С. Герасимова и Р. Г. Астраханцева производится разбор прогнозирования инфляционных процессов с применением методов машинного обучения, в результате чего создается предиктивная модель [1]. Анализируется данная модель как одно из средств машинного обучения, обосновывается её пригодность для использования, рассматриваются основные этапы работы и результаты анализа.

В исследовании О. Д. Казакова и Н. Ю. Азаренко представлены данные о метриках качества алгоритмов машинного обучения, включая алгоритм «Random Forest» [6]. Особое внимание уделено формулировке точных правил классификации в управленческих методиках, а также рассмотрены размеченные данные по указанным методам.

Таким образом, в исследованиях активно анализируют и обсуждают влияние машинного обучения на различные стороны экономической и прогностической деятельности, подкрепляя свои исследования обширным обзором научной литературы. Работы охватывают широкий спектр тем, начиная от интеграции алгоритмов машинного обучения в системы прогнозирования до разработки прогностических моделей для макроэкономических показателей. Особенное внимание уделяется не только техническим аспектам, таким как классификация и систематизация машинного обучения, но и практическому применению этих технологий для оптимизации процессов управления.

Авторы подчеркивают важность машинного обучения для повышения точности и эффективности прогностических моделей, подтверждая как теоретическими

исследованиями, так и практическими кейс-стади. Например, использование нейронных сетей и методов обучения с учителем демонстрирует значительные преимущества в задачах регрессии, прогнозирования и экстракции знаний.

Кроме того, акцентируется внимание на проблемах и ограничениях, связанных с машинным обучением, включая необходимость дальнейших исследований в повышении качества и точности прогнозов. Выводы полученные из работ подтверждают, что машинное обучение является мощным инструментом, способным существенно влиять на экономическую науку и практику, требующие дополнительной адаптации и оптимизации для решения специфических задач в различных условиях.

Введение в алгоритмы машинного обучения

Анализ случайных процессов представляет собой одну из наиболее динамичных и развивающихся областей современной науки, призванную улучшить понимание и прогнозирование как технических так и экономических процессов. В технических науках случайные процессы анализируются с целью предсказания и планирования использования ресурсов, при этом осознавая, что такие процессы по своей природе не могут быть предсказаны абсолютно точно, а только с определённой вероятностью [2]. В экономических науках, хотя случайности встречаются реже, активно используются методы прогнозирования для оценки таких параметров, как валютные курсы и другие экономические показатели, что говорит о возросшем интересе к этой области в последние годы. Оба направления демонстрируют значительный интерес к разработке методов, позволяющих с большей точностью оценивать и предсказывать исходы в условиях неопределенности [6].

Прогнозирование является ключевой задачей во многих областях науки и практики, при этом можно выделить два основных подхода к его реализации: использование методов, основанных на теории вероятностей и математической статистике, а также методов, основанных на машинном обучении. Методы первой группы, такие как стохастические дифференциальные уравнения, основаны на моделировании случайных процессов с различными характеристиками изменчивости траекторий. Второй подход заключается в создании и обучении моделей на основе данных, что позволяет машине самостоятельно извлекать закономерности и использовать их для прогнозов. Оба подхода имеют свои преимущества и могут быть использованы для решения прогностических задач в зависимости от специфики и требований к точности и устойчивости результатов [1].

Прогнозирование, используя методы машинного обучения, представляет собой мощный инструмент для определения контрольных показателей, анализа влияния нововведений, планирования бюджетов и прогнозирования спроса. Благодаря способности обрабатывать большие объемы данных, машинное обучение обеспечивает возможность выявления неочевидных закономерностей, что значительно повышает точность прогностических моделей [7, 8].

Существующие алгоритмы машинного обучения для прогнозирования

Использование алгоритмов машинного обучения считается одной из наиболее эффективных методик для прогнозирования. Реализация данного подхода требует выполнения ряда конкретных шагов. Начальным этапом является подготовка данных, которая включает очистку от выбросов и ошибок, заполнение пропусков в данных, а также их трансформацию в формат, пригодный для анализа. После под-

готовки данных следует выбор алгоритма машинного обучения, который наилучшим образом соответствует поставленным задачам и условиям данных. Среди популярных алгоритмов можно выделить линейную регрессию, метод случайных лесов, градиентный бустинг и нейронные сети. Каждый из этих алгоритмов имеет уникальные характеристики, которые выбираются для определенных типов задач прогнозирования (табл. 1).

Таблица 1

Методы машинного обучения

Наименование метода	Краткое описание	Этапы реализации
Алгоритм Decision Stump	Простой алгоритм машинного обучения, который представляет собой одноуровневое дерево решений, состоящее из одного корневого узла и двух листьев. Он используется в качестве базового алгоритма в ансамблевых методах, таких как бустинг. Принцип работы алгоритма Decision Stump заключается в поиске оптимального разделения данных по одному признаку (или их комбинации) таким образом, чтобы минимизировать ошибку классификации или регрессии	1.Выбрать признак и порог 2.Вычислить ошибку 3.Выбрать оптимальный порог 4.Оценить качество
Линейная регрессия	Метод статистического моделирования, который используется для анализа отношений между зависимыми и независимыми переменными. Основная идея линейной регрессии заключается в том, что зависимая переменная может быть линейной комбинацией одной или нескольких независимых переменных с некоторыми коэффициентами	1.Подготовить данные 2.Выбрать модель 3.Обучить модель 4.Оценить модель 5.Применить модель 6.Мониторить и обновлять
Метод опорных векторов	Мощный алгоритм машинного обучения, который используется как для задач классификации, так и для регрессии. Основная идея метода опорных векторов заключается в поиске оптимальной разделяющей гиперплоскости, которая максимально разделяет классы в обучающем наборе данных	1.Подготовить данные 2.Выбрать ядро 3.Обучить модель 4.Оптимизировать гиперпараметры 5.Оценить модель 6.Провести валидацию модели 7.Применить модель
Бустинг	Ансамблевый метод машинного обучения, который комбинирует несколько слабых моделей для создания сильной модели. Основная идея бустинга заключается в последовательном обучении моделей, где каждая следующая модель исправляет ошибки предыдущей модели	1.Выбрать базовый алгоритм 2.Обучить слабые места 3.Выбрать функции потерь 4.Построить ансамбль 5.Провести итерации 6.Вынести итоговое предсказание

Выбор наиболее подходящего алгоритма машинного обучения определяется спецификой задачи и характером доступных данных. Для успешного использования методов машинного обучения необходимы не только технические компетенции, но и понимание особенностей соответствующей отрасли, позволяя адаптировать технологические решения к конкретным потребностям.

Метод машинного обучения Random Forest

В представленной работе осуществляется анализ экономических показателей ценных бумаг ключевых российских компаний финансового сектора. Исследуемые активы включают ценные бумаги, которые являются частью списка системно значимых кредитных организаций, утверждённых Центральным банком РФ на 17 октября 2023 года, включая ПАО РОСБАНК (ROSB), ПАО Сбербанк (SBER), и АФК «Система» (AFKS). В данной работе применен алгоритм машинного обучения Random Forest для построения прогнозов по данным ценным бумагам, что позволило оценить их будущее поведение на финансовом рынке.

Алгоритм случайного леса выбран в связи с рядом преимуществ, которые делают его высокоэффективным инструментом в сценариях анализа данных. Основные достоинства включают:

1. Высокая точность. Структура алгоритма, включающая множество деревьев решений, каждое из которых обучено на своем подмножестве данных, позволяет агрегировать прогнозы для повышения точности результатов.

2. Непараметрический характер. Алгоритм не осуществляет предположений о распределении данных, что делает его адаптивным к разнообразным наборам данных и способным выявлять сложные закономерности.

3. Обработка отсутствующих данных. Random Forest естественным образом справляется с пропущенными значениями и выбросами, поскольку каждое дерево обучается на случайной выборке входных данных, что снижает влияние аномальных значений на общий результат.

4. Универсальность применения. Алгоритм эффективно работает как в задачах регрессии, так и классификации, предоставляя также возможность оценки важности входных переменных для принятия решений.

Таким образом, благодаря ключевым характеристикам, алгоритм случайного леса является мощным и гибким инструментом для анализа данных, подходящим для широкого спектра приложений.

Алгоритм случайный лес также является удобным, потому что гиперпараметры по умолчанию, которые он использует, часто дают хороший результат прогнозирования. Разобраться с гиперпараметрами довольно просто, и к тому же их не так уж много. [13]

Одна из самых больших проблем машинного обучения – это переоснащение, но в большинстве случаев этого не произойдёт благодаря классификатору случайных лесов. Если в лесу достаточно деревьев, классификатор не будет перегружать модель. [13]

На основе изложенных выше аргументов и учитывая их значимость в контексте данного исследования, выбор в пользу представленного алгоритма оказался обоснованным. Для анализа и последующего прогнозирования цен на ценные бумаги были использованы данные за период с 1 января 2020 года по 1 марта 2024 года. Прогнозируемые изменения стоимости ценных бумаг компании AFKS представлены на рисунке 1.

Среднеквадратичная ошибка в данном прогнозе равна 0.44, что является достойным показателем расстояния между фактическими и прогнозируемыми значениями.

Прогноз данных ценных бумаг ROSB представлен на рисунке 2.

Среднеквадратичная ошибка в данном прогнозе равна 0.51, что также является отличным показателем расстояния между фактическими и прогнозируемыми значениями.

Прогноз данных ценных бумаг SBER представлен на рисунке 3.

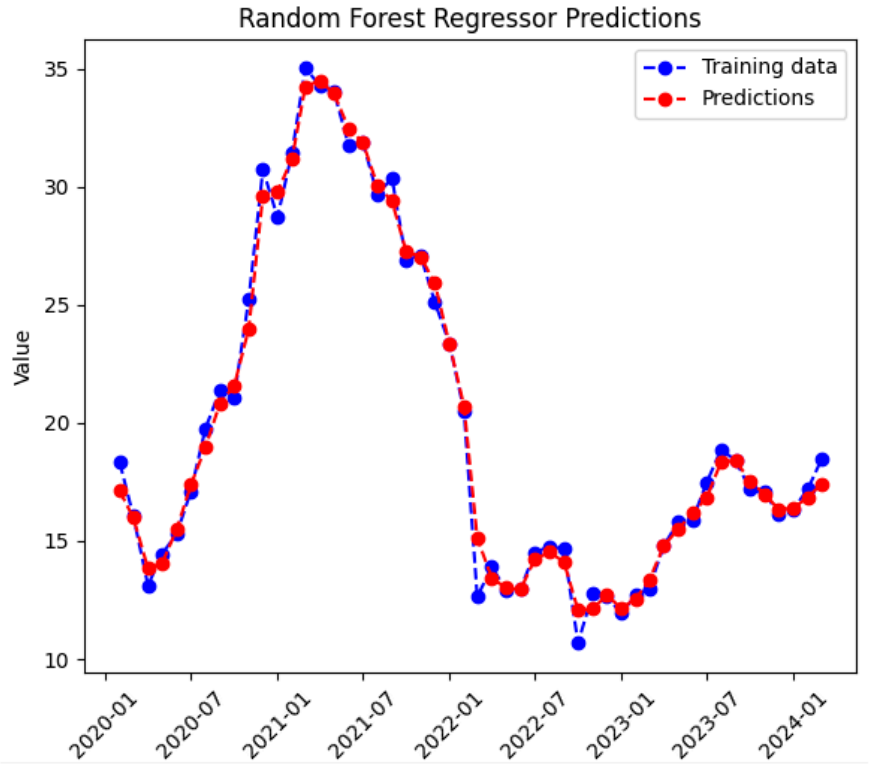

Рис. 1. *Прогноз AFKS*

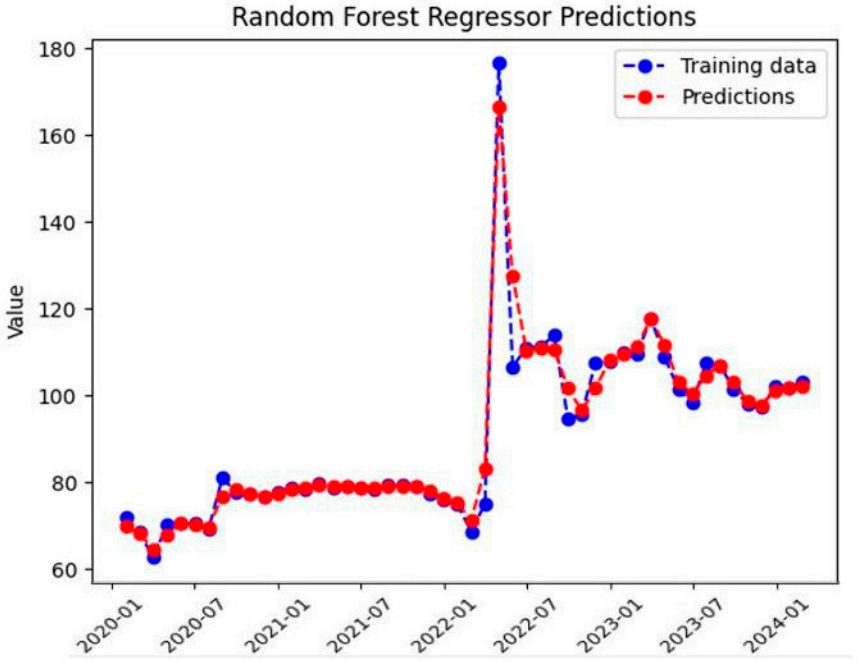

Рис. 2. *Прогноз ROSB*

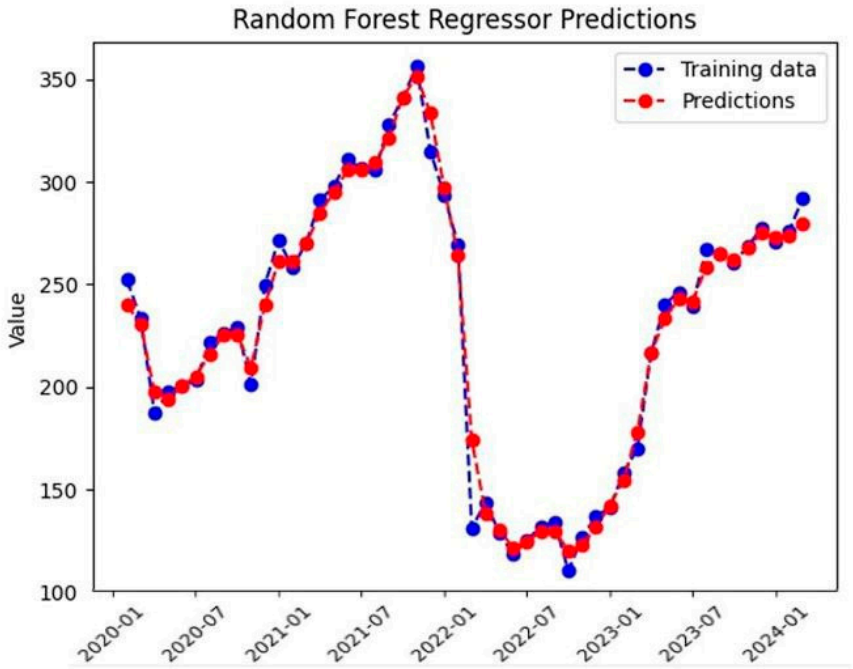

Рис. 3. *Прогноз SBER*

Среднеквадратичная ошибка в данном прогнозе равна 0.47, которая, как и два прошлых результата является хорошим показателем разности.

В рамках исследования применения алгоритмов машинного обучения для анализа рынка ценных бумаг особое внимание было уделено методу Random Forest. Настройка данного алгоритма была проведена с учетом следующих параметров: максимальная глубина каждого дерева в ансамбле составила 10 уровней, а общее количество деревьев достигло 10. Анализ результатов работы алгоритма, включая визуализацию данных и расчет среднеквадратичной ошибки, позволяет утверждать о высокой эффективности данной модели в задачах прогнозирования движения цен на ценные бумаги. Данные выводы подтверждаются как непосредственными показателями точности модели, так и графическими иллюстрациями результатов её работы.

Выводы

В рамках проведённого анализа методологий и инструментов, применяемых в области машинного обучения, можно сформулировать следующий вывод: использование статистических и математических методов является основой для проведения экспериментальных исследований и пилотных испытаний в процессе создания алгоритмов обучения машин.

Особое значение приобретает возможность применения одного алгоритма к различным задачам в экономической сфере. Так, алгоритм Random Forest заслуживает внимания в ситуациях, когда необходимы особо точные результаты или при наличии значительного объема данных для обработки, что предполагает использование мощного инструмента, способного эффективно справляться с поставленными задачами.

Литература

1. Астраханцева И.А., Герасимов А.С., Астраханцев Р.Г. Прогнозирование региональной инфляции с помощью алгоритмов машинного обучения // Известия высших учебных заведений. Серия: Экономика, финансы и управление производством. 2022. № 4 (54). С. 6-13.

2. Балыков А.С., Каледина Е.А., Володин С.В. Прогнозирование прочности при сжатии и проектирование составов конструкционных легких бетонов с применением методов машинного обучения // Нанотехнологии в строительстве: научный интернет-журнал. 2023. Т. 15, № 2. С. 171-186.

3. Бекирова Э.А., Бекирова М.Э. Экономический анализ данных и прогнозирование с использованием методов машинного обучения // Ученые записки Крымского инженерно-педагогического университета. 2023. № 2 (80). С. 49-53.

4. Колосова А.С., Каменева А.С., Чуков Г.В., Никифоров А.Ю. Возможность применения алгоритмов машинного обучения для прогнозирования качества ЭКБ и РЭА // Безопасность информационных технологий. 2023. Т. 30, № 1. С. 123-129.

5. Давыдов А.Н. использование машинного обучения в экономике // Современные проблемы лингвистики и методики преподавания русского языка в ВУЗе и школе. 2022. № 38. С. 845-848.

6. Казаков О.Д., Азаренко Н.Ю. Модель управления человеческим капиталом в системах принятия решений // Вестник Воронежского государственного университета инженерных технологий. 2019. Т. 81, № 3 (81). С. 290-298.

7. Макеев К.А., Шаповаленко К.В., Котов Я.С. Исследование алгоритмов машинного обучения в задачах прогнозирования урожайности сельскохозяйственных культур // Современные инструментальные системы, информационные технологии и инновации: Сборник научных трудов XVII Международной научно-практической конференции, Курск, 17–18 марта 2022 года / Редколлегия: Разумов М.С. (отв. ред.). Курск: Юго-Западный государственный университет, 2022. С. 264-266.

8. Саханевич Д.Ю. Исследование подходов и методов применения искусственного интеллекта и машинного обучения в социально-экономических процессах // Вестник Омского университета. Серия: Экономика. 2020. Т. 18, № 2. С. 65-79.

9. Тимакова Н.Д., Кумратова А.М., Курносова Н.С. Использование линейного клеточного автомата при прогнозировании цен на акции ММК // Информационное общество: современное состояние и перспективы развития: Сборник материалов XIV международного форума, Краснодар, 12–17 июля 2021 года. Краснодар: Кубанский государственный аграрный университет имени И.Т. Трубилина, 2021. С. 409-413.

10. Филиппенко В.А., Решетникова И.В. Использование машинного обучении для прогнозирования курса биткойна к курсу доллара // Будущее науки-2020: Сборник научных статей 8-й Международной молодежной научной конференции. В 5-ти томах, Курск, 21–22 апреля 2020 года. Том 3. Курск: Юго-Западный государственный университет, 2020. С. 345-349.

11. Филиппова А.С.Р., Ксенофонтова О.Л. Актуальность применения методов машинного обучения для прогнозирования продаж // Сборник научных трудов вузов России «Проблемы экономики, финансов и управления производством». 2023. № 53. С. 265-267.

12. Чагина П.А. Анализ возможности применения машинного обучения и нейронных сетей для проектирования и оптимизации торговых роботов // Альманах научных работ молодых ученых Университета ИТМО, Санкт-Петербург, 02–05 февраля 2022 года. Том 2. СПб.: федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО», 2022. С. 388-390.

13. Шуляк Е. Макроэкономическое прогнозирование с использованием данных социальных сетей // Деньги и кредит. 2022. Т. 81, № 4. С. 86-112.