

ОЦЕНКА ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНОЙ НЕЙРОННОЙ СЕТИ С УЧЕТОМ ИСТОРИЧЕСКИХ ДАННЫХ И ОЦЕНКОЙ ТОНАЛЬНОСТИ ТЕКСТА ДЛЯ АНАЛИЗА АКЦИЙ НА ФОНДОВОМ РЫНКЕ

ASSESSMENT OF THE EFFECTIVENESS OF APPLYING GENERATIVE ADVERSARY NEURAL NETWORK TAKEN INTO ACCOUNT OF HISTORICAL DATA AND ASSESSMENT OF TEXT TONE FOR ANALYSIS OF STOCKS IN THE STOCK MARKET

A. Tereshenko

Summary. This publication is part of a study in the field of systems analysis of the stock market and focuses on assessing the effectiveness of the use of generative adversarial neural networks (GAN) taking into account historical data and sentiment analysis of the text. The author considers the possibility of using this technology to analyze the dynamics of shares on the stock market. The results of the study are of particular interest because the author suggests that predicting stock prices is impossible. This aspect becomes especially relevant in the context of using neural networks to test this statement.

Keywords: neural networks, dataset, system analysis, discriminator, sentiment assessment.

Терешенко Андрей Алексеевич

Аспирант, ФГАОУ ВО «Северный (Арктический) Федеральний Университет имени М.В. Ломоносова», г. Архангельск
andrey.tereshenko2017@mail.ru

Аннотация. Данная публикация является частью исследования в области системного анализа фондового рынка и сосредотачивается на оценке эффективности применения генеративно-сопоставительных нейронных сетей (GAN) с учетом исторических данных и анализа тональности текста. Автором рассматривается возможность использовать эту технологию для анализа динамики акций на фондовом рынке. Результаты исследования представляют особый интерес, поскольку автор предполагает, что предсказание цен акций является невозможным. Этот аспект становится особенно актуальным в контексте использования нейронных сетей для проверки данного утверждения.

Ключевые слова: нейронные сети, датасет, системный анализ, дискриминатор, оценка тональности.

В современном мире фондовый рынок играет ключевую роль в экономическом развитии страны, обеспечивая финансирование корпоративных проектов, стимулируя инновации и способствуя росту национального богатства. Для России, как для страны с развивающейся экономикой и амбициозными планами по модернизации, участие простых граждан в инвестировании на фондовом рынке приобретает особенное значение. На данный момент спрос на инвестирование в нашей стране невелик, но за счёт системного анализа мы можем постепенно повысить его. Системный анализ, в силу своего интегративного подхода, представляет собой мощный инструмент для решения сложных проблем, включая слабый интерес к инвестированию на фондовом рынке. Рассматривая фондовый рынок как сложную систему, системный анализ позволяет анализировать множество взаимосвязанных компонентов, таких как инвесторы, компании, регулирующие органы, и макроэкономические факторы. Моделирование и анализ, используемые в рамках системного анализа, помогают изучить поведение системы в различных ус-

ловиях и оценить эффективность различных сценариев инвестирования. Учет многофакторности позволяет системно анализировать влияние экономических, политических, и технологических факторов на рыночные процессы. Кроме того, системный подход способствует оптимизации распределения инвестиционных ресурсов и управлению рисками, путем выявления наиболее перспективных секторов и компаний для инвестирования, а также разработки стратегий диверсификации портфеля и применения инструментов хеджирования. Таким образом, системный анализ представляет собой комплексный подход к решению проблемы, позволяя учитывать многообразие факторов, влияющих на рыночные процессы, и разрабатывать эффективные стратегии для их улучшения.

Согласно Тарасенко Ф. П., в контексте прикладного системного анализа выявление проблемы является важнейшим этапом, так как позволяет сконцентрироваться на решении ключевого аспекта, оказывающего значительное воздействие на общую работу системы [1, 14с].

Определение главной проблемы обеспечивает направленность аналитических усилий и помогает эффективно распределять ресурсы для создания стратегий и решений, направленных на её устранение [2]. Этот подход способствует более глубокому и систематическому пониманию причин и механизмов функционирования системы, что, в свою очередь, повышает шансы на успешное решение проблемы. Глобальной проблемой в моём исследовании является недостаточный интерес людей в нашей стране к инвестированию на фондовом рынке. Мои исследования, опросы различных возрастных групп, проведенные в рамках написания диссертации, показали, что данная проблема состоит из ряда малых, но важных аспектов [3]. Одним из них является то, что существует распространенное убеждение среди людей о возможности прогнозирования цен акций, преимущественно на основе анализа новостных событий и их влияния на финансовые рынки. На первый взгляд сама идея предсказания цен акций является абсурдной и неосуществимой в реальном мире. Это мнение подкреплено не только эмпирическими наблюдениями, но и теоретическими размышлениями о природе финансовых рынков. Однако, для обоснования этого утверждения требуется подтверждение.

В современной эпохе нейронные сети и методы машинного обучения приобретают все большую популярность. Их применение в контексте финансовых рынков представляет собой перспективный инструмент для решения различных задач. Нейронные сети в настоящее время активно применяются на фондовом рынке для прогнозирования динамики ценных бумаг и определения оптимальных стратегий инвестирования. Они используются для анализа больших объемов данных, включая исторические данные о ценах, объемах торгов, финансовых показателях компаний, новостях и социальных медиа. Нейронные сети способны выявлять сложные закономерности и нелинейные зависимости в этих данных, что позволяет анализировать возможные будущие тенденции. Кроме того, они могут использоваться для создания алгоритмов торговли, которые автоматически адаптируются к изменяющимся условиям рынка и могут обеспечить оптимальное исполнение сделок. Эти методы находят применение как в высокочастотной торговле, так и в инвестиционном управлении. Информационные технологии, особенно нейронные сети, тесно связаны с системным анализом из-за их способности обрабатывать большие объемы данных и выявлять скрытые закономерности в сложных системах. Нейронные сети, моделирующие работу человеческого мозга, способны адаптироваться к разнообразным задачам, осуществлять классификацию, прогнозирование и оптимизацию, что делает их мощным инструментом для анализа сложных систем. Системный анализ, в свою очередь, предполагает изучение взаимосвязей и взаимодействий между элементами системы, а также выявление

её структуры и функционирования. В совокупности информационные технологии и нейронные сети предоставляют возможности для анализа данных, извлечения значимых паттернов и обнаружения скрытых зависимостей в системах, что делает их неотъемлемой частью современного системного анализа. Цель же данной публикации заключается в оценке эффективности стратегии инвестирования в акции на российском фондовом рынке, основанной на анализе новостной информации и настроений людей. В рамках исследования будет проведена проверка данной методики инвестирования с использованием генеративно-сопоставительных сетей (GAN) для оценки ее эффективности. Используя терминологию, GAN — это тип нейронных сетей, в котором две нейронные сети, генератор и дискриминатор, соревнуются друг с другом в игре с нулевой суммой. Генератор создает данные, постепенно улучшая их качество, в то время как дискриминатор пытается различить реальные данные от сгенерированных. Цель GAN заключается в создании реалистичных данных, которые неотличимы от реальных, и обучении модели на основе имеющихся данных без прямого участия человека [4]. Применение этих сетей для изучения акций на фондовой бирже с учётом общественного мнения и новостных сообщений, различных постов представляет собой новаторский подход, основанный на объединении реальных финансовых данных. GAN можно обучить на исторических временных рядах стоимости акций, новостных заголовках и публикациях в социальных сетях, чтобы анализировать будущие тенденции. Кроме того, модуль для определения настроения, использующий анализ тональности текста, может быть интегрирован для оценки общей позиции инвесторов. Это позволяет принимать во внимание человеческий фактор и новостные события для анализа динамики рынка. Таким образом, подход, основанный на GAN, обеспечивает более глубокое изучение фондового рынка с учётом множества факторов, включая общественное мнение и новостные события.

Сбор данных играет ключевую роль в научных исследованиях и разработке технологий по многим причинам. Первоначально, данные позволяют исследователям выявлять паттерны, тенденции и взаимосвязи в информации, что существенно помогает понять влияние различных факторов на исследуемые явления или определить важные параметры для разработки технологий [5]. Кроме того, сбор обширного и разнообразного набора данных необходим для обучения моделей машинного обучения, что в свою очередь позволяет строить более точные и обобщающие модели. Данные также необходимы для оценки производительности алгоритмов и моделей в реальных условиях, а также для проверки полученных результатов на статистическую значимость и устойчивость. Наконец, собранные данные используются для разработки моделей прогнозирования, аналитических инструментов и систем поддержки принятия решений,

что важно для планирования действий и управления рисками. В целом, сбор данных является основой для научных исследований, инноваций и развития технологий в различных областях науки и промышленности. Инициатива по сбору данных началась с создания обширного набора информации, включающего новости и публикации о крупной компании — Мобильные ТелеСистемы (далее — МТС). Этот набор включает в себя широкий спектр данных, в том числе общественные отзывы, комментарии и аналитические материалы. Собранный датасет предоставляет ценную основу для анализа реакции общественности на деятельность компании МТС, а также для выявления трендов и образования прогностических моделей в области рыночной динамики и общественного мнения. Таким образом, был собран набор данных, включающий в себя четыре категории информации: дату, начиная с июня 2023 года и заканчивая апрелем 2024 года, текстовую информацию: различные посты, новости, идентификатор (тикер) и название компании.

При оценке тональности, использовалась нейросеть, разработанная для оценки полярности настроений в тексте. Методика основана на словаре, который содержит оценки тональности для слов, а также правила для обработки устойчивых конструкций и эмоциональных интенсификаторов, также особенностью является то, что учитывается особенности выражения эмоций в онлайн-коммуникациях, таких как социальные медиа. Рассмотрим подробнее. Нейросеть Dostoevsky, созданная специально для анализа тональности русскоязычных текстов, применяет методы NLP и глубокого обучения, такие как рекуррентные нейросети (RNN) и трансформаторные модели, для обработки последовательных данных, подобных текстам [6,7]. Набор данных RuSentiment включает размеченные по тональности тексты на русском языке, относящиеся к различным темам, таким как отзывы о продуктах, фильмах и книгах. Обучение модели происходит на основе этого набора данных с использованием методов обучения с учителем, что позволяет выявлять в текстах закономерности и паттерны, связанные с определёнными эмоциональными оттенками, такими как положительный, отрицательный или нейтральный, благодаря чему модель успешно анализирует и классифицирует тональность текстов [8]. Сеть Vader, применяется для анализа настроений в текстовых данных. Она часто используется в сфере обработки естественного языка для определения тональности текста и выявления позитивных, негативных или нейтральных эмоциональных состояний [9]. В своей работе она использует словарь, включающий слова с их оценками настроения, и правила для обработки сочетаний слов. Алгоритм Vader рассчитывает общую эмоциональную направленность текста, принимая во внимание его контекст и языковые особенности, что позволяет определить окраску текста с высокой точностью. В результате анализа настроений были получены средние оценки полярности

всех ранее собранных постов и новостей о компании. В таблице 1, можно увидеть итоговую датасета с обработанными данными.

Таблица 1.
Часть обработанных результатов оценки тональности

Дата поста	Оценка настроений
2024-04-22	0.8677
2024-04-22	0.643
2024-04-22	0.4871
2024-04-22	0.844

Далее был собран второй датасет для построения графика акции. Были использованы библиотеки Pandas, Matplotlib, Seaborn и др [10]. Из собранного датасета были загружены данные о ценах акций за каждый день начиная с июня 2023 года и заканчивая апрелем 2024 года, включая цену открытия и закрытия, максимальную и минимальную цены за день, а также объем торгов. Данный временной промежуток соответствует датам в первом датасете. После загрузки данных они были обработаны, включая преобразование формата даты и времени, и подготовлены для построения графика. Затем с помощью Matplotlib был создан график, отображающий изменение цен акций в течение выбранного периода времени. Таким образом, используя Seaborn и Matplotlib, был создан график акций, представленный на рисунке 1, что соответствует реальному графику, а значит построение было выполнено верно.

При создании архитектуры модели GAN использовались следующие параметры: генератор с 5 блоками LSTM, дискриминатор имеет 5 сверточных и 3 плотных слоя с сигмовидной функцией активации. На рисунке 2 показан результат работы. На этом рисунке можно увидеть пунктирную линию, она соответствует прогнозируемой цене, именно так это видит нейросеть, сам график представлен зеленым цветом. Данный участок соответствует графику представленному на рисунке 1.

В ходе проведённого исследования был сделан вывод о том, что модель генеративно-состязательных сетей (GAN) проявляет значительную эффективность в определенных задачах, но наблюдаемое слишком высокое отклонение еще раз подтверждает основной посыл исследования: невозможно предсказать цену акции, даже используя нейронные сети. Предсказание цен акций — это сложная и на данный момент нереализуемая задача. Она требует учета множества переменных, таких как экономические, политические и социальные факторы. Несмотря на применение передовых технологий, точность прогнозов остается невысокой из-за нескольких ключевых причин. Во-первых, финансовые рынки сложны и динамичны, подверженные влиянию множества факторов, что делает моделирование их поведения

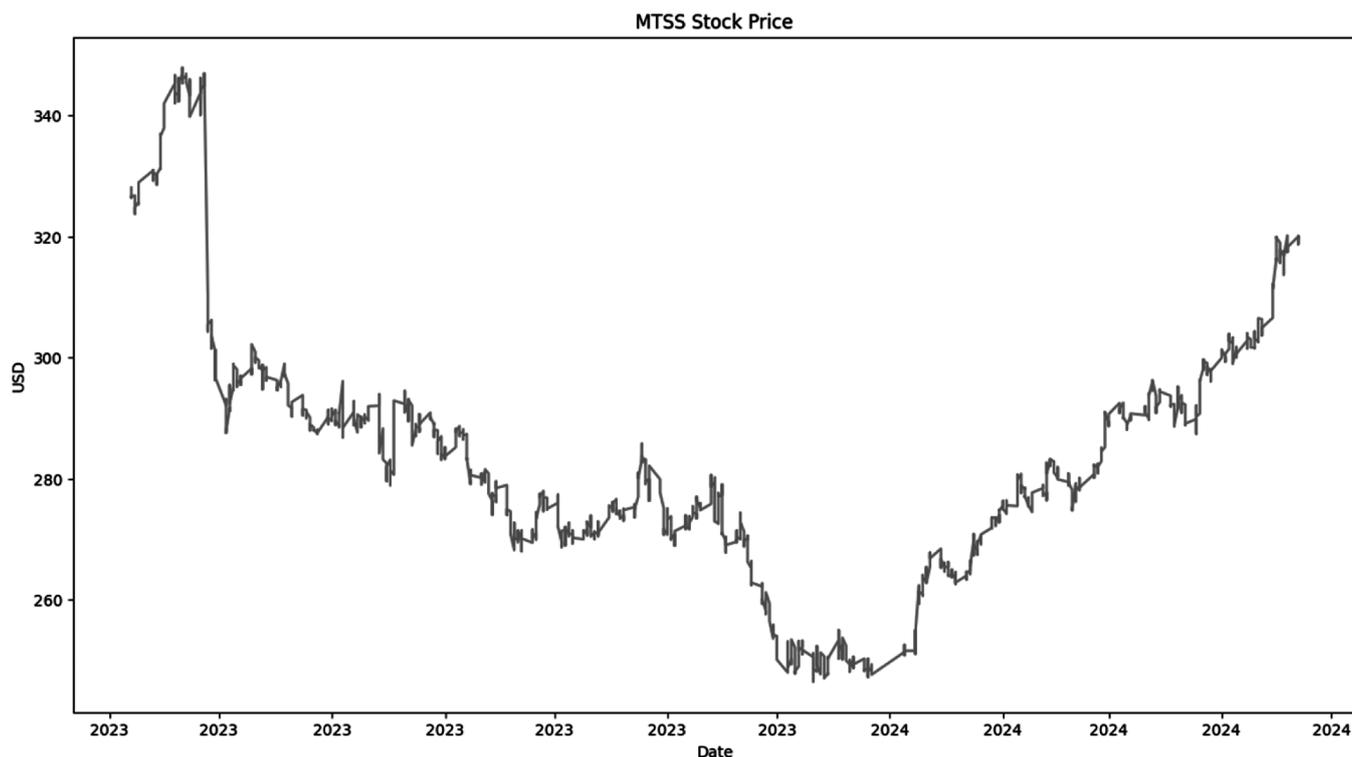


Рис. 1. График компании



Рис. 2. Результат работы нейронной сети

сложной задачей. Во-вторых, рынки характеризуются высокой степенью неопределенности и случайными колебаниями, что усложняет прогнозирование. Кроме того, ограничения данных и изменение рыночных условий со временем также вносят свои сложности. И даже с применением передовых методов исследования пока-

зывают, что точность прогнозов остается невысокой. Что показало и данное исследование. В свете этих ограничений данную модель могут рассматривать профильные компании, но лишь как вспомогательный инструмент в сфере анализа, который сможет помочь инвесторам принимать более информированные решения.

ЛИТЕРАТУРА

1. Тарасенко Ф.П. Прикладной системный анализ: учебное пособие / Ф.П. Тарасенко. — М.: КИИ. — 2010. — 224 с.
2. Качала В.В. Теория систем и системный анализ: учебник для студ. учреждений высшего профессионального образования / В.В. Качала. — М.: Издательский центр «Академия», 2013. — 272 с.
3. Анализ активности частных инвесторов на фондовом рынке в 2024 году [Электронный ресурс]: — URL: <https://alfabank.ru/make-money/investments/learn/t/dolya-chastnikh-investorov-v-ovorotakh-mosbirzhi-padaet/>, свободный (дата обращения: 01.05.2024)
4. Основы генеративно-состязательных сетей / Хабр [Электронный ресурс]: — URL: <https://habr.com/ru/articles/726254/>, свободный (дата обращения: 01.05.2024)
5. О важности датасета и о том, как сделать его лучше — опыт компании / Хабр [Электронный ресурс]: — URL: <https://habr.com/ru/articles/678808/>, свободный (дата обращения: 01.05.2024)
6. Dostoevsky — анализ тональности в Python за 5 минут. [Электронный ресурс]: — URL: <https://egorovegor.ru/analiz-tonalnosti-s-python-i-dostoevsky/>, свободный (дата обращения: 01.05.2024)
7. Dostoevsky — нейросеть для анализа текста. [Электронный ресурс]: — URL: <https://github.com/bureaucratic-labs/dostoevsky>, свободный (дата обращения: 01.05.2024)
8. RuSentiment — набор данных для оценки анализа текста на русском языке. Источник: <https://egorovegor.ru/analiz-tonalnosti-s-python-i-dostoevsky> [Электронный ресурс]: — URL: <https://github.com/text-machine-lab/rusentiment>, свободный (дата обращения: 01.05.2024)
9. Vader: A Comprehensive Guide to Sentiment Analysis [Электронный ресурс]: — URL: <https://medium.com/@rslavanyageetha/vader-a-comprehensive-guide-to-sentiment-analysis-in-python-c4f1868b0d2e>, свободный (дата обращения: 01.05.2024)
10. Топ 6 библиотек Python для визуализации: какую и когда лучше использовать? / Хабр [Электронный ресурс]: — URL: <https://habr.com/ru/companies/otus/articles/558478/>, свободный (дата обращения: 01.05.2024)

© Терешенко Андрей Алексеевич (andrey.tereshenko2017@mail.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»